NATURAL VARIATION OF LARGE PLASMIDS IN BACTERIAL POPULATIONS

by

LAURA E. WILLIAMS

(Under the Direction of Anne O. Summers)

ABSTRACT

Plasmids encode an extraordinary range of adaptive functions, including antibiotic resistances, virulence factors and degradative enzymes. Plasmids also mediate horizontal gene transfer, which impacts evolution of bacterial chromosomes. Despite their important ecological and evolutionary roles, we lack a comprehensive picture of the variation and evolution of plasmids themselves. In this dissertation, I investigated the natural variation of large plasmids on two different scales: 1. individual plasmid sequences from Gram-negative and -positive bacteria and 2. total plasmid content of *E. coli* and *Salmonella* reference collections.

For the sequence-based studies, I sequenced four large plasmids from *E. coli*, *Staphylococcus* and *Corynebacterium*. Whereas 98% of the *E. coli* plasmid was highly similar to other plasmid sequences, 40-60% of the three Gram-positive plasmids had no apparent similarity to known sequences. The *E. coli* plasmid's backbone, which encodes "core" functions such as replication, was almost identical to that of a *Salmonella enterica* sv Typhimurium plasmid. Subsequent analysis identified this backbone in five other enterobacterial plasmids. All seven plasmids differed in accessory gene content, leading to variation in plasmid-encoded phenotypes.

For the collection-based studies, I analyzed 228 *E. coli* and *Salmonella* strains from four reference collections. Plasmids, especially large plasmids, were abundant. Replicon typing identified two common families, IncF and IncI1, by the presence of characteristic backbone genes. At least 20% of strains with large plasmids in each collection were untypable, suggesting variation in plasmid backbones. Restriction fragment length polymorphism (RFLP) analysis showed high variation in large plasmid genomes. The majority of *E. coli* and *Salmonella* large plasmids (87.5% and 56%, respectively) had unique RFLP patterns. The lack of widespread RFLP patterns suggested that the prevalence of IncF and IncI1 is due to similar backbone genes on otherwise different plasmids. The only exception was the IncFII virulence plasmid pSLT, which was detected in multiple Typhimurium strains from different hosts and geographic locations, suggesting that virulence plasmids evolve differently from other plasmids.

These results show that plasmid evolution is a significant force shaping the horizontal gene pool. Rather than existing as static elements, plasmids are dynamic and flexible genetic scaffolds driving gene flux in host bacteria.

INDEX WORDS: plasmid evolution, horizontal gene transfer, mobile genetic element

NATURAL VARIATION OF LARGE PLASMIDS IN BACTERIAL POPULATIONS

by

LAURA E. WILLIAMS

B.S., Illinois Wesleyan University, 2001

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2009

NATURAL VARIATION OF LARGE PLASMIDS IN BACTERIAL POPULATIONS

by

LAURA E. WILLIAMS

Major Professor:	Anne O. Summers
Committee:	Jessica Kissinger
	Mark Schell
	Patricia Sobecky

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
May 2009

DEDICATION

Dedicated to Bram Tucker and Joan Williams. They made a hard journey easier.

ACKNOWLEDGEMENTS

Most importantly, I would like to thank my advisor Anne Summers for guiding me through my time here at UGA with constant enthusiasm. I admire how she fully integrates her students into the scientific community, which is one of the most important but intangible purposes of graduate school. From the beginning, she included me as a full participant in interactions with our collaborators and faculty colleagues at UGA and elsewhere. I felt very welcomed into the community of scientists, and I am grateful to her for that.

My committee members Jessica Kissinger, Mark Schell and Patricia Sobecky have given me valuable help and advice during this journey. They have been supportive and patient, and I am very glad to have benefited from their expertise throughout my graduate work.

I am indebted to a number of incredible mentors, including Beth Arthington-Skaggs, David Bollivar, Marcio Chedid, Jonathan Dey, Don LeBlanc, Jo Porter, Slava Titov, and Chris Zook. Each of them played an important role in getting me to this point, and I hope that their experiences were as rewarding as mine.

I would also like to thank my fellow graduate students at UGA, especially Caran Cagle, Adrienne Cottrell, Erinn Howard, Bijal Patel and Julie Shearer. I wish that Bijal was still here to see his own work through to the end, but I imagine that he has all the answers now.

Joy Wireman deserves special recognition. It is easy to forget how much work it takes to keep things running smoothly, and Joy is a true master. Beyond that, she is a thoughtful and accomplished researcher. If I ever have a lab of my own, I'll be lucky to find someone half as good as Joy.

Balance in my life has always kept me sane, and during my time in Athens I have been blessed with a truly remarkable group of friends. The GSYPF crew, with special mention to Christy and Paul, and the Killer Tomahtoes tennis team have been a constant and delightful source of hilarity and friendship. I know that I will miss them when I move on.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

Half a century ago, Esther Lederberg, Luigi Cavalli-Sforza and Joshua Lederberg published the first description of the fertility factor F (94). At the time, they described the F factor as an "infective agent" that could be inherited vertically and transmitted horizontally. Later that year, in the context of an ongoing debate regarding the roles of cytoplasmic elements in heredity, Joshua Lederberg proposed the term "plasmid" to encompass all extrachromosomal genetic elements (92, 93). Although Lederberg intended to include mitochondria, chloroplasts and viruses under this umbrella term, in subsequent years the broad definition of "plasmid" was narrowed to identify extrachromosomal genetic elements that replicate autonomously and, in the most popular operational definition, do not encode essential cellular functions (56).

The initial report regarding the F plasmid was quickly followed by a series of important discoveries. In Japan, researchers led by Tsutomu Watanabe documented the transfer of antibiotic resistances between bacteria by plasmids (165); Pierre Fredericq and his colleagues identified colicinogenic plasmids responsible for toxin production (55); and Richard Novick and Mark Richmond characterized plasmids encoding penicillin resistance in *Staphylococcus* (120). These studies fueled the growth of the new field of plasmid biology. Since the pioneering work of the 1950s and 1960s, plasmids have been reported in a wide range of hosts, including archaea and eukaryotic fungi (66, 171, 174). Surveys of disparate geographic locations have isolated plasmids from bacterial and

archaeal strains of polar regions, desert ecosystems and deep sea hydrothermal vents, indicating that plasmids are abundant globally (83, 111, 172, 174).

The ability of plasmids to confer "instant" antibiotic resistance on bacterial hosts was the hook that interested many researchers in plasmid biology, and this property continues to attract much attention today. However, recent advances in comparative genomics have refocused the spotlight on the broader roles of plasmids as mediators of horizontal gene transfer. The adaptive potential provided by plasmids is extensive, and it is not limited to antibiotic resistance. Plasmids are a flexible scaffold for acquiring new gene content in the form of transposons, integrons or other mobile genetic elements. They also mediate the transfer of DNA across species, genus and even domain boundaries, enabling integration of new or modified functions into the chromosome, which can dramatically impact host evolution (121). In light of plasmids' important ecological and evolutionary roles, it is essential to develop a comprehensive picture of the natural variation of plasmids found in bacterial communities. This aim is the focus of my dissertation research.

In this introductory chapter and literature review, I will present an overview of the genetic systems encoded by plasmids, including replication, maintenance and transfer systems, discuss some of the phenotypes associated with plasmids, such as antibiotic resistance and virulence, and review the literature related to plasmid ecology and evolution. Finally, in the context of this information, I will discuss the objectives of the dissertation.

GENETIC STRUCTURE OF PLASMIDS

In the early years of plasmid research, Watanabe and colleagues noted that antibiotic resistance plasmids could be separated into two components, termed the resistance transfer factor and the resistance determinant (166). Genetic analysis later showed that the resistance transfer factor contained essential functions involved in plasmid replication, stable maintenance and conjugative transfer, whereas the resistance determinant encoded antibiotic resistance genes, which are not essential to the plasmid but rather affect the survival of the host under certain circumstances. Such observations eventually led to the classification of plasmid genes as either "backbone" or "accessory", paralleling the concept of "core" and "accessory" genes arising from comparative genomics of bacterial chromosomes (56, 156).

In a plasmid genome, backbone and accessory genes are often organized into functional modules (156, 160) (Figure 1.1). To some degree, this modularity permits a reductionist view of plasmids as assemblages of individual modules producing a stable plasmid replicon (95). However, it is also important to consider the complete plasmid as a higher order structure whose modules may be subject to coordinate regulation (156). Backbone gene modules, which encode replication, maintenance and transfer functions, are often clustered together, with relatively few interruptions from insertions of accessory genes (for example, Figure 1.1). Accessory elements are frequently acquired at the same places on the plasmid genome, suggesting that random insertion of accessory genes may disrupt regulation of backbone functions and reduce the fitness of the plasmid (156).

BACKBONE SYSTEMS

Replication

Circular plasmids replicate by one of three general mechanisms: theta type, strand displacement or rolling circle replication (38) (Figure 1.2). A common feature of all three mechanisms is the requirement for a plasmid-encoded origin of replication (*oriV*). These origins are *cis*-acting DNA regions that support autonomous replication of the plasmid. They are often defined by identifying the smallest fragment of a plasmid that can support the replication of a mini-plasmid lacking its own origin. This approach identifies the minimal origin required for plasmid DNA replication, however, the plasmid may also possess additional sequences that are not strictly essential for replication but contribute to efficiency or copy number control. In addition to an origin of replication, most plasmids encode a replication initiation protein, though a few plasmids, such as ColE1, rely solely on host-encoded proteins. All characterized plasmids require some host-encoded elements, such as DNA polymerase, for replication.

Theta Type Replication

Theta type replication was named for the structure of replication intermediates, which resembled the Greek letter θ when visualized by electron microscopy. Initiation of theta type replication occurs when the plasmid-encoded initiator (Rep) protein binds to specific sequences within the origin of replication (for example, (64, 107)). For some plasmids, these sequences include tandem arrays of direct repeats called iterons, which are important for initiation and control of replication (116). Once Rep proteins are bound to the origin, they mediate interactions with host-encoded replication proteins, which assemble to form a replisome (18).

The host-encoded DnaA replication initiation protein, which functions in chromosomal replication, often plays a role in plasmid replication as well, binding to *dnaA* boxes in the plasmid origin. Formation of the open complex, in which the DNA strands are separated, occurs at an AT-rich region within the plasmid origin. Strand separation is promoted in part by DNA bending induced by both the plasmid- and host-encoded replication initiation proteins, as well as other host-encoded factors such as IHF (integration host factor) and Fis (factor for inverse stimulation) (40, 148). The host-encoded DnaB helicase also plays an important role in open complex formation.

In addition to DnaA and DnaB, theta type plasmid replication also requires host-encoded primase (DnaG) and DNA polymerase III (38). The primase catalyzes the synthesis of RNA primers for each strand. DNA polymerase III uses these primers to initiate leading- and lagging-strand synthesis. DNA synthesis proceeds around the plasmid, facilitated by the helicase, until reaching a termination sequence (*ter*). This sequence binds the host-encoded Tus protein, which is thought to interact with the helicase to catalyze termination of DNA synthesis (73, 143).

Although the above mechanism is generally applicable for most theta type replicating plasmids, there are some variations (38). For instance, plasmid ColE1 does not encode its own replication initiation protein. Instead, ColE1 replicates via the concerted action of a number of host-encoded proteins. Initiation occurs when RNA polymerase synthesizes a preprimer for leading-strand synthesis from a specific ColE1 origin sequence. Host RNases process the preprimer, and DNA polymerase I then extends the processed primer until a specific nucleotide sequence is exposed. This site allows assembly of the host-encoded primase/helicase/DNA polymerase III replisome complex.

The complex primes lagging-strand synthesis from the exposed site, and then continues DNA synthesis on both strands.

Strand Displacement Replication

Plasmids replicating by strand displacement encode three replication proteins: RepA (helicase), RepB (primase), and RepC (replication initiator) (137). These plasmids do not rely on host-encoded initiator (DnaA), helicase (DnaB) or primase (DnaG) for replication, therefore they are able to replicate in a broader range of hosts than many other plasmids. A typical plasmid origin for strand displacement replication consists of iterons, an adjacent AT-rich segment, and two single-stranded initiation (*ssi*) sites, one on each strand (74).

In a model proposed by Scherzinger and colleagues, strand displacement replication is initiated when the RepC protein binds to the iterons within the origin (138). RepC binding may directly promote open complex formation by inducing bending in the plasmid DNA. In addition, RepC is thought to recruit the RepA helicase to the AT-rich segment adjacent to the iterons, whereupon RepA separates the double-stranded DNA in a processive fashion, eventually exposing the *ssi* sites (176). Each site forms a stem-loop structure that promotes binding of RepB primase, which primes continuous DNA synthesis in opposite directions by host-encoded DNA polymerase III. Termination of synthesis yields two double-stranded plasmids. If replication initiates at only one of the *ssi* sites, the end products are a double-stranded plasmid and a displaced single strand. In this case, the *ssi* site on the single strand initiates synthesis of the complementary strand, generating the second copy of the plasmid.

Rolling Circle Replication

Some small plasmids (<10 kb) replicate using a rolling circle replication (RCR) mechanism. Originally identified in plasmids of Gram-positive bacteria (80), this mode of replication is also used by plasmids of Gram-negative bacteria (for example, (39)) and archaea (for example, (175)). There are three essential plasmid-encoded elements in RCR: a replication initiation protein, a double-strand origin and a single-strand origin (81). Replication initiates at the double-strand origin (*dso*), which frequently has sequences prone to forming secondary structure (114, 115). The replication initiation (Rep) protein binds to a site within the *dso* and stabilizes the formation of either a hairpin or cruciform structure (78, 115). This secondary structure exposes a *nic* site, where the Rep protein nicks the plasmid DNA (84).

The Rep protein also recruits host-encoded proteins to the nicked site to form the replisome (24, 147). A host-encoded helicase unwinds the plasmid DNA (3, 76), exposing single-stranded DNA. DNA polymerase uses the 3′ OH at the nick for leading-strand synthesis, displacing the original leading strand (1). The replication fork proceeds until it returns to the *dso*, at which point termination occurs via an undetermined mechanism (81). A series of cleavage events releases a complete, circularized copy of the plasmid as well as the displaced leading strand. To generate the second copy, host-encoded proteins initiate replication at the single-strand origin on the displaced strand and synthesize the lagging strand (12, 85). Host-encoded DNA gyrase supercoils both copies to complete the replication process.

Control of Plasmid Replication

Plasmid copy number control is an essential element of plasmid replication. If there are too many plasmid copies, the cell experiences a growth disadvantage due to the extra metabolic burden of DNA replication and gene expression. If there are too few copies, the plasmid is at risk of being lost from the population. Therefore, most plasmids encode mechanisms for controlling plasmid replication (25, 37). Although the regulatory strategies are diverse, the main theme in each strategy involves negative regulation of replication initiation (132). This regulation is achieved primarily through either antisense RNA-mediated inhibition or iteron binding of replication initiation proteins.

Antisense RNAs, or countertranscript RNAs, are synthesized from the complementary strand of a gene encoding an essential replication element, such as the replication initiation protein or, in the case of ColE1, the RNA primer (37). Because antisense RNAs are constitutively transcribed but unstable, changes in plasmid copy number will be reflected by changes in antisense RNA concentration (19). Annealing of antisense RNAs to the mRNA of the replication initiation protein causes a range of inhibitory effects in different plasmid systems, from premature termination of transcription (119) to inhibition of mRNA translation (4).

The other major mechanism of plasmid replication control involves binding of the replication initiation protein to the iterons within the plasmid origin (87). Saturation of iterons with bound Rep protein is an important step in plasmid replication initiation. Two models have been proposed for iteron-mediated control of plasmid replication: a titration model (161) and a handcuffing model (110). In the titration model, the Rep proteins are assumed to be rate-limiting for initiation. The iterons bind and sequester Rep, thereby

controlling the amount of initiator available. In the handcuffing model, Rep proteins complexed with iterons on one plasmid are proposed to interact with Rep-iteron complexes on another plasmid copy, resulting in steric hindrance at the origins that blocks replication initiation. This model provides a copy-sensing mechanism. A low plasmid copy number reduces the likelihood of interaction between two plasmid origins, and replication proceeds unhindered. By contrast, a high plasmid copy number increases the likelihood of steric hindrance and replication inhibition.

The titration model was disfavored when it was discovered that the Rep protein negatively autoregulates by binding to an operator at the *rep* promoter. Sequestration of Rep proteins by the iterons, as proposed in the titration model, should relieve transcriptional repression, leading to an increase in available Rep and activation of the origin. This weakness in the model was partially addressed when Rep proteins were found to occur in both monomeric and dimeric forms. Rep monomers were proposed as the rate-limiting form, which needed to saturate the iterons at the origin to initiate replication, whereas Rep dimers were thought to bind to the operator sequence and act as negative autoregulators.

However, recent studies have identified Rep dimer-iteron complexes (15, 42). One study showed that the Rep dimers bound to alternate iterons, leaving intervening iterons "empty" and never saturating the origin (15). Although the exact molecular details of plasmid copy number control are still under investigation, it is most likely that a combination of the mechanisms proposed in the titration and handcuffing models regulate initiation of plasmid replication (31).

Incompatibility

In a phenomenon known as plasmid incompatibility, two plasmids encoding similar replication or partitioning systems cannot stably coexist in the same cell in the absence of selection (118). Because plasmid origins are randomly selected during replication, two plasmids encoding similar origins and *trans*-acting copy number control elements may be replicated at a different rate. Without a means of differentiating the plasmids' replication systems, the less-replicated plasmid cannot correct the drift in copy number and is lost over the course of multiple generations. Often, a strain with two incompatible plasmids will segregate into a mixed population of single-plasmid strains.

Plasmid incompatibility was proposed as a means of classifying plasmids into families (33). Initially, the incompatibility phenotype itself served as the criterion for assigning a plasmid to a particular family. The process required extensive *in vivo* testing with mini-replicons containing characteristic incompatibility loci for each family (32). When sequencing revealed that incompatible plasmids shared similar replication genes, the classification of plasmids into families shifted to a more sequence-based approach, first with probe hybridization (29) and most recently with qualitative PCR (23, 79). Twenty-six different incompatibility families, also known as replicon families, have been identified for plasmids of Enterobacteriaceae (154). The available replicon typing methods have provided valuable means of surveying large collections of plasmids for similarities in backbone systems, but it is important to note that the currently described replicon families are not exhaustive, and plasmids are frequently isolated that do not fit into this classification hierarchy (145).

Maintenance

Plasmids use a variety of strategies to prevent segregational loss in a growing bacterial population. Some of these strategies ensure that plasmid copies are distributed evenly to each daughter cell during bacterial cell division, whereas others cause the death of daughter cells that did not inherit the plasmid.

Random diffusion

Small, high copy number plasmids rely on the passive mechanism of random diffusion to distribute plasmid copies throughout the cell and ensure inheritance by both daughter cells. If the plasmid's distribution within the cell is truly random, the probability of obtaining a plasmid-free daughter cell is $2^{(1-n)}$, where n = the number of plasmid copies (117). This formula shows that a high copy number leads to stable plasmid inheritance in a growing bacterial population. For example, for a plasmid with 30 copies per cell, only two cells in 10^9 will be plasmid-free.

Multimer resolution

Plasmid copies are not always randomly distributed as individual units throughout a bacterial cell. The processes of replication and recombination can cause the formation of plasmid multimers, which reduces the number of individual plasmids in the cell and increases the probability of obtaining a plasmid-free daughter cell (69). In addition, the random selection of plasmid origins during replication may lead to plasmid multimers increasing their numbers opposed to monomers, further increasing the rate of segregational loss (151). Therefore, many plasmids of both high and low copy number encode a multimer resolution system to ensure that plasmid copies are maintained as individual units.

Plasmid multimers are resolved by site-specific recombinases (sometimes annotated as resolvases), which catalyze DNA cleavage and rejoining reactions at a defined nucleotide sequence on the plasmid called *res* (for resolution site) (177). Multimer resolution is catalyzed by recombinases of both the serine- and tyrosine-recombinase families (69). The enzymes may be plasmid-encoded or supplied by the host. The Cre-*loxP* system of plasmid P1 (6) and the ResD/*rfs* system of the F plasmid (88) are examples of complete plasmid-encoded multimer resolution systems. In contrast, host-encoded Xer recombinases catalyze the resolution of multimers of colicin plasmid ColE1 (152).

Partitioning

Even with a multimer resolution system, large plasmids often occur at such low copy numbers that random diffusion is insufficient to ensure stable inheritance in a bacterial population. For example, for a plasmid with five copies per cell, one cell in every 16 will be plasmid-free if random diffusion and multimer resolution are the only maintenance strategies used. Stable inheritance of many large, low copy number plasmids is reinforced by the activity of plasmid-encoded partitioning loci (51, 58).

Most plasmid partitioning loci encode three elements: a *cis*-acting, centromere-like sequence characterized by multiple repeats, a DNA-binding protein that recognizes the centromere-like sequence, and an ATPase (5, 61, 123). Although many of the ATPases belong to the Walker-type ATPase superfamily, a few are part of an actin-like family. Partitioning loci are often divided into two groups based on the homology of their ATPases (62).

The general mechanism of partitioning is similar between the two groups. The partitioning complex is formed when the DNA-binding proteins bind to the centromere-like sequence. This complex can also include host proteins such as IHF (57). After the initial partitioning complex is assembled, more of the DNA-binding partition proteins are recruited to the flanking regions around the centromere-like sequence. This protein array then interacts with the partitioning ATPase. Polymerization of ATPase proteins forms a filament, which is thought to push or pull the plasmid towards the cell pole. Upon reaching either the one-quarter- or three-quarters-position in the cell, which are the mid-points of the future daughter cells, the ATPase filament dissociates.

Elements of the partitioning mechanism are still unclear. For example, an early step in many partitioning models is the pairing of two plasmid copies via interactions between the partitioning complexes at the two centromere-like sequences (52, 77). This is thought to orient the plasmid copies and ensure their localization to the opposite poles of the cell upon polymerization of the ATPase filament. This model is supported by recent evidence showing direct centromere pairing mediated by the DNA-binding partition proteins (136). By contrast, an alternative model of the partitioning mechanism of actin-like ATPases does not involve direct centromere pairing. Instead, the partitioning complex on each plasmid copy serves as an anchor for short, rapidly disassociating ATPase filaments, which appear to "search" the surrounding space (59). When two filaments contact each other, they stabilize and polymerize to form a long filament that pushes the two plasmids apart.

Localization of plasmids during partitioning orients each plasmid at the mid-cell points of the future daughter cells (54, 130). Recently, partitioning loci were shown to

restrict the mobility of the plasmid during the entire cell cycle, confining plasmids to the mid-cell point until partitioning acts to segregate them to the mid-cell points of future daughter cells (41). Although it is unclear exactly how the partitioning system "tethers" plasmids to this position, controlled plasmid localization helps to ensure effective segregation by positioning plasmids at the optimum point for equal distribution to each daughter cell.

Post-segregational killing

In addition to active partitioning loci, some plasmids also encode post-segregational killing (psk) systems as further insurance against segregational loss. Rather than actively distributing plasmid copies to each daughter cell, psk systems eliminate daughter cells that did not inherit the plasmid (70, 82, 177). All of these systems have a general "toxin-antitoxin" organization, in which the plasmid encodes a toxin with a long half-life and an anti-toxin with a short half-life. When the cell divides, both toxin and anti-toxin are distributed to each daughter cell in the cytoplasm. If one of the daughter cells did not inherit the plasmid, the cell is not capable of producing additional anti-toxin. Once the existing anti-toxin is degraded, the longer-lived toxin will kill the bacterial host cell. The specific mechanisms and bacterial host targets of psk systems vary widely.

An example of a protein-based system is the *ccd* psk system encoded by the F plasmid. In this system, the CcdB protein binds to DNA gyrase and inhibits its activity, trapping it at the DNA cleavage stage and inducing double-strand breaks in host chromosomal DNA (10, 11). Gyrase inhibition also induces oxidative damage, which contributes to cell killing (49). The CcdA anti-toxin protein inhibits CcdB by direct

binding (106). CcdA is degraded by Lon protease, which is encoded by the host cell (164).

In the *hok*/*sok* psk system of plasmid R1, the Hok protein inserts into the inner membrane and disrupts membrane integrity, causing loss of membrane potential and efflux and influx of small molecules (60). The anti-toxin Sok is an antisense RNA transcribed from the *hok* locus; it binds to the *hok* mRNA leader region and inhibits translation of *hok* mRNA. Sok-RNA is digested more rapidly by host RNAses than *hok* mRNA, therefore if the plasmid is lost, digested Sok-RNA cannot be replaced, and the *hok* mRNA will be translated and cause cell death.

Transfer

In addition to replication and maintenance functions involved in intracellular survival, many plasmids encode a suite of proteins that mediates intercellular plasmid spread by conjugative transfer (90, 173). For the plasmid, intercellular transfer can also be considered a survival function, since it allows a plasmid to increase its frequency in a population independently of the success of its original host lineage. For the recipient cell, plasmid-mediated horizontal gene transfer may expand the strain's adaptive potential, conferring a competitive advantage over other strains and/or enabling niche expansion. Conjugative transfer operons are often found on large (≥30 kb) plasmids, but not all large plasmids are capable of self-transfer. Instead, some plasmids encode mobilization loci that allow them to spread by "hitchhiking" with a coresident conjugative plasmid.

Conjugative transfer

Conjugative transfer may have arisen by adaptation of a DNA replication system and a macromolecular transport apparatus, which evolved to suit the task of intercellular

plasmid DNA transfer (102). The only novel protein needed to complete the system was a coupling protein that bridged the gap between the two components. This hypothesis is supported by the general organization of transfer functions in many conjugative plasmids. Within plasmid-encoded transfer operons, one set of genes encodes the DNA processing functions, which resemble rolling circle replication (167), and another set encodes a type IV secretion apparatus capable of transferring DNA across cell envelopes from donor to recipient (91). In addition, most transfer operons encode a coupling protein, which interacts with both the DNA processing proteins and the type IV secretion system (T4SS) proteins (22).

The mechanistic basis of conjugation is the focus of extensive ongoing research. The existing data support a general, albeit incomplete model for conjugation in Gram-negative bacteria. Our understanding of plasmid transfer in Gram-positive bacteria is not yet as detailed (67). Characterization of conjugative systems encoded by different Gram-negative plasmids has led to the classification of conjugation as F-like or P-like, named after the incompatibility family of the prototypical plasmid in each group (90). Whereas processing of plasmid DNA for transfer is similar in both groups, the structure of the T4SS apparatus and the mechanism of transfer differ in some respects. In the following section, I will briefly describe the general mechanism of conjugation for plasmids of Gram-negative bacteria, pointing out where the two kinds of conjugative systems differ.

All Gram-negative conjugative plasmids encode a T4SS apparatus, also called a mating pair formation apparatus (Figure 1.3). This apparatus is a membrane-spanning multi-protein complex that forms a channel or pore within the cell envelope of the donor cell (91). The complex also extends a pilus into the extracellular space. Pili are the

hallmarks of conjugative plasmid transfer in Gram-negative bacteria. They are assembled from pilin subunits, which are stored in the inner membrane of the donor cell (105). Processing of pilin proteins involves cleavage in both F-like and P-like conjugative systems as well as additional modifications, such as acetylation in F-like systems and head-to-tail cyclization in P-like systems (53, 113, 140). F-like conjugative systems are characterized by long, flexible pili, which are equally efficient for DNA transfer in liquid media and on solid surfaces, whereas P-like conjugative systems are characterized by short, rigid pili, which are only efficient on solid surfaces (17). These pilus morphologies may be due to differences in auxiliary proteins of the two conjugative systems rather than the pilin proteins themselves (91).

Pilus assembly begins with the formation of a tip structure, which is poorly characterized. The T4SS apparatus then mediates polymerization of pilin subunits to extend the pilus into the extracellular environment. The pilus contacts potential recipient cells, however the nature of the "mating signal" that initiates conjugative DNA transfer and the identity of pilus receptors on the recipient cell surface are unknown (for example, (129)). Once the "mating signal" is detected, the pilus is retracted in F-like conjugative systems via disassembly of pilin proteins at the base of the T4SS apparatus, which returns pilin subunits to the inner membrane (126). Retraction of the pilus pulls the donor and recipient cells into direct physical contact. Because pilus retraction has not been demonstrated for P-like conjugative plasmids, the mechanism for obtaining direct cell-to-cell contact in this system is undefined.

The exact structure of the conjugative pore or channel in the recipient cell is also not well-understood. Mutational analysis has not identified conjugative intermediates

with DNA trapped in the recipient periplasm, suggesting that the transfer apparatus spans the outer and inner membranes of the recipient cell (90). Electron microscopy has identified protein complexes at "conjugational junctions" where the donor and recipient cell envelopes are tightly appressed (48). These complexes may be assembled independently of the pilus extension machinery. Alternatively, the conjugative pilus itself may act as a needle, piercing the recipient cell envelope and crossing both membranes (102). Part of the mechanism of transversing the recipient cell envelope involves plasmid-encoded enzymes capable of digesting the peptidoglycan cell wall (43). Mating pair formation in F-like conjugative systems is stabilized by six proteins unique to these plasmids. The absence of the proteins in P-like systems may cause the reduced efficiency of these systems in liquid media, where shearing force is more likely to disrupt conjugational junctions (91).

Formation of the conjugational junction stimulates plasmid DNA processing in preparation for conjugative transfer (103). A complex of plasmid-encoded and in some cases host-encoded proteins forms a "relaxosome" which binds to a specific nucleotide sequence on the plasmid, called the origin of transfer (*oriT*) (127). Chief among the plasmid-encoded relaxosome proteins is the relaxase, which catalyzes a site- and strand-specific cleavage reaction at a *nic* site within the *oriT* (20, 128). The relaxase covalently attaches to the 5′ end of the nicked plasmid DNA strand. Displacement of the nicked strand occurs in concert with DNA synthesis, which regenerates the double-stranded, supercoiled plasmid via a rolling circle replication mechanism (89). The displaced single-stranded plasmid DNA is delivered to the transfer apparatus by interactions between the covalently attached relaxase and the coupling protein.

The nature of the conjugative transfer substrate and the mechanism of action are still under investigation. However, a generally well-accepted model proposes a two-step process that combines transport of the relaxase protein and bound plasmid DNA with active translocation of the plasmid DNA by the coupling protein (102). In this model, the relaxase protein is the actual transfer substrate of the T4SS apparatus, and the bound plasmid DNA is passively transferred due to its covalent attachment at the 5′ end. This step of the model is supported by a recent study showing that a T4SS apparatus encoded by plasmid R388 transports relaxase to the recipient cell, even in the absence of covalently bound plasmid DNA (47). Relaxase transfer has not yet been demonstrated for F-like conjugative systems, but it is considered likely that relaxase acts as a pilot protein for DNA transfer in this system as well (91).

Because of the length of the plasmid DNA, active transport of relaxase by the T4SS does not always result in complete transfer of the entire plasmid DNA strand. In the proposed model, the coupling protein confers processivity to plasmid DNA transfer at this stage. Most coupling proteins have ATPase activity (153), which suggests that they could act as pumps, actively pushing plasmid DNA into the transfer apparatus by rotational tracking along the DNA strand (21). This activity may be enhanced by a T4SS protein found in both F-like and P-like systems that also possesses an NTPase domain (139). After the plasmid DNA has been completely transferred, the translocated relaxase catalyzes the strand rejoining reaction to recircularize the transferred plasmid DNA strand in the recipient cell (47, 89).

In addition to the relaxase, P-like conjugative systems also transport plasmid-encoded DNA primase to the recipient cell, possibly in a mechanism independent of

relaxase-plasmid DNA complex transfer (169, 170). Once in the recipient cell, primase proteins are thought to generate RNA primers from the transferred plasmid DNA strand, thereby initiating the synthesis of the complementary strand. In F-like conjugative systems, proteins encoded by the recipient cell generate primers for plasmid DNA synthesis (169).

Surface and entry exclusion

Surface and entry exclusion prevent the conjugative transfer of a plasmid to a recipient cell already carrying a plasmid encoding a similar transfer operon (157). In F-like conjugative systems, this occurs via two mechanisms. In the first, a plasmid-encoded lipoprotein occurs in abundance in the outer membrane of the cell and reduces the cell's receptiveness to pili encoded by a similar conjugative system (108). In the second, a plasmid-encoded protein localizes to the inner membrane and prevents plasmid DNA entry through the conjugational junction (68).

Mobilization

Compared to conjugative plasmids, mobilizable plasmids encode limited transfer-related functions. These include an origin of transfer and the relaxase needed to form the relaxosome (173). Most mobilizable plasmids do not encode a coupling protein, but instead rely on the coupling protein of the coresident conjugative plasmid to mediate interaction of the mobilizable plasmid's relaxosome with the T4SS apparatus. This interaction affects the efficiency of plasmid mobilization by different conjugative plasmids. Often, a particular plasmid is mobilized most efficiently by a specific conjugative plasmid.

ACCESSORY GENES

The backbone and accessory genes of a plasmid genome often originate from different sources, imparting a highly mosaic structure on most plasmids, especially large ones. For example, analysis of a large plasmid from the Gram-negative proteobacterium *Rhizobium* identified some proteins with similarity to those of a plasmid from Gram-positive *Clostridium acetobutylicum*, other proteins similar to a cyanobacterial plasmid, and still others similar to a plasmid from the archaeon *Halobacterium* (95). Both backbone and accessory genes may be shuffled among plasmids, but accessory genes are often associated with transposases and integrases, which mediate intermolecular recombination and facilitate the distribution of accessory genes among many different plasmids.

Newly acquired transposons and integrons frequently insert within existing accessory elements on plasmids, resulting in complex accessory modules that evoke comparisons to nested Russian dolls (2). For example, the Tn*21* transposon of plasmid R100 consists of four separate accessory elements (99) (Figure 1.1). The main scaffold of Tn*21* is a large composite transposon encoding a mercury resistance operon. This transposon has a Class 1 integron inserted between the mercury resistance and transposition genes. The integron itself contains two insertion sequence elements, one of which is nested inside the other. The example of Tn*21* is typical for many plasmid accessory elements, which appear to provide recombinational hotspots for acquiring new gene content. The phenotypes conferred by plasmid accessory genes span an extraordinary functional diversity, including antibiotic and metal resistances, genes

involved in pathogenicity and symbiosis, and enzymes catalyzing degradation of xenobiotic compounds (158).

As mentioned in the introduction to this chapter, antibiotic resistances are possibly the most widely recognized plasmid-borne phenotype. A search of NCBI's PubMed database for the terms "plasmid antibiotic resistance" in article titles or abstracts returned 118 articles published since January 1, 2008. Such studies have catalogued the extensive range of antibiotic resistances carried by plasmids, which counteract most classes of antibiotics currently in use (8). A few recent developments are particularly noteworthy, including the emergence of plasmid-borne 16s rRNA methylases conferring aminoglycoside resistance on Gram-negative pathogens causing systemic infections (28). In addition, increasing resistance has been observed against newly developed extended spectrum cephalosporins due to acquisition and spread of resistance genes by plasmids (150). An especially alarming development saw a plasmid from *Staphylococcus aureus* acquire a transposon encoding vancomycin resistance, thereby compromising one of the last weapons available against this significant human pathogen (168).

In addition to antibiotic resistance, some plasmid accessory elements also encode resistances to other compounds, including metals and disinfectants. The Tn*21* transposon described above carries a mercury resistance operon encoding proteins that reduce toxic mercuric ions to volatile elemental mercury, which diffuses from the cell (99). Other accessory elements contain additional mercury resistance genes for detoxification of organomercurials (7). Plasmid-mediated biocide resistance is less studied than antibiotic or metal resistances, but genes encoding resistance to quaternary ammonium compounds, which are used as disinfectants, are found on a number of plasmids (for example, (13)).

Other plasmid accessory genes have a dramatic impact on the lifestyle of the host bacterium, conferring essential functions for pathogenicity or symbiosis. *Bacillus anthracis* owes its pathogenicity to two plasmids carrying virulence factors, without which the bacterium cannot cause the disease anthrax (46). A number of *Salmonella enterica* serovars harbor virulence plasmids with *spv* operons, which contribute to the bacterium's pathogenicity (for example, (98)). On the other end of the bacterial lifestyle spectrum, a group of large plasmids provide the essential functions for the establishment of a symbiotic relationship between rhizobial bacteria and leguminous plants. The plasmid-encoded proteins are important in both root nodule formation and nitrogen fixation (109).

Plasmid accessory elements can also extend the metabolic capabilities of the host bacterium and expand the range of compounds used as carbon sources. Plasmid-encoded degradative enzymes target both natural compounds, such as toluene, naphthalene and nicotine, and xenobiotic compounds, such as polychlorinated biphenyls and carbofuran (122). Rather than encoding entire metabolic pathways, plasmid-borne enzymes often catalyze peripheral steps, which degrade target compounds to intermediates that feed into existing, chromosomally encoded metabolic pathways. The degradative properties of some plasmid-borne accessory elements have shown promise in bioremediation of industrial pollutants, including especially recalcitrant compounds (159).

PLASMID ECOLOGY AND EVOLUTION

Plasmid Population Biology

The vast majority of plasmid biology literature is from a mechanistic perspective, focusing on the backbone systems and accessory phenotypes discussed above. By

comparison, the ecology and evolution of plasmids has received less attention. Forays into these areas began with efforts to investigate a major question: how are plasmids maintained in bacterial populations?

Plasmid persistence in bacterial populations is confronted with two main challenges: fitness cost and segregational loss (Figure 1.4). *In vitro* studies have shown that most plasmids impose a fitness cost on host bacteria in the absence of selection for plasmid-borne genes (101). Although the exact cause is unknown, this cost has been ascribed to the extra metabolic burdens of plasmid protein synthesis and/or DNA replication or the disruption of cellular regulatory systems. The fitness cost suffered by plasmid-bearing cells manifests as a decreased growth rate compared to plasmid-free cells, which results in the elimination of the plasmid from the population over time. In addition, even though plasmids employ multiple strategies to ensure their inheritance in a growing bacterial population, these systems are not 100% efficient, and plasmid-free segregants may still arise in the population (Figure 1.4).

Despite these challenges to plasmid persistence, surveys of natural bacterial communities frequently find plasmids in abundance. How are these plasmids maintained? There are two main hypotheses, which may not be mutually exclusive: (1) plasmids are parasitic elements that are maintained at the expense of their hosts and (2) plasmids are mutualistic elements that are maintained by carrying genes beneficial to the host (50, 141).

The argument in favor of the "plasmids as parasites" hypothesis suggests that horizontal transfer (sometimes referred to as infectious transfer) can maintain a plasmid in a population in spite of its fitness cost by converting plasmid-free cells to plasmid-

bearing cells. Clearly, this is primarily applicable for conjugative plasmids. Mobilizable plasmids, which cannot self-transfer but can "hitchhike" with coresident conjugative plasmids, are more restricted under this hypothesis since they depend on conjugative plasmids for horizontal transfer (96, 141). Small plasmids that do not encode either transfer or mobilization genes present an even greater challenge to the "plasmids as parasites" hypothesis, which may explain why these plasmids are largely ignored in theoretical considerations of plasmid persistence.

Alternatively, the "plasmids as mutualists" hypothesis suggests that under certain environmental conditions, selection for plasmid-borne genes allows the host to overcome the fitness cost of the plasmid. In these circumstances, plasmid-bearing cells experience a growth rate advantage over plasmid-free cells. The advantage may be sufficiently large that the plasmid-bearing subpopulation sweeps through and completely displaces the plasmid-free subpopulation. This hypothesis has some weaknesses; for instance, selective pressures often fluctuate temporally or spatially and therefore may not be adequate to account for plasmid persistence. In addition, some plasmids are cryptic, with no known benefit to the host. Small plasmids again present a challenge to this hypothesis, since some encode only one or two proteins with plasmid-specific functions.

The first evaluations of these hypotheses from a population biology perspective were presented in the late 1970s by Frank Stewart and Bruce Levin (149). In their initial work, they mainly considered plasmid persistence in the absence of selection, thereby testing the "plasmids as parasites" hypothesis. Their mathematical models assumed that conjugative transfer occurred as a mass action process, meaning that transfer is equal to the product of donor and recipient cell density multiplied by a rate parameter. This

parameter corresponds to the success of plasmid transfer for each donor/recipient encounter (97). From their models, they concluded that when the fitness cost of a plasmid is greater than zero (resulting in a growth disadvantage for plasmid-bearing cells), cell density and the horizontal transfer rate must exceed the rate of segregational loss and the growth rate disadvantage to maintain the plasmid in the population (149). In other words, conjugative plasmids can be maintained as parasites only when horizontal transfer is efficient enough to overcome the plasmid's fitness cost and the loss due to segregation.

This basic model was later extended to address persistence of nonconjugative but mobilizable plasmids (96, 141). Levin and Stewart concluded that mobilizable plasmids could not be maintained in a population without direct selection for a plasmid-borne gene. However, they assumed that because of surface and entry exclusion, conjugation could not occur between two cells that each harbor the conjugative plasmid; this assumption does not allow for transfer of the mobilizable plasmid to cells bearing only the conjugative plasmid. Simonsen modified Levin and Stewart's model to allow this possibility at a reduced rate. He also reduced the fitness cost of the mobilizable plasmid compared to the conjugative plasmid. These modifications revealed that under certain conditions, mobilizable plasmids can be maintained parasitically if and only if conjugative plasmids transfer at a high enough rate.

From this theoretical work, it is clear that the rate of conjugative transfer is a key factor in determining whether plasmids can be maintained as parasites in bacterial populations. Early estimates of conjugative transfer using a batch culture, end-point method showed that for the R1 plasmid in laboratory and natural strains of *E. coli*, transfer rates were too low to overcome the estimated fitness cost and segregational loss

of the plasmid over time (65, 142). These and other results (101) argued against the "plasmids as parasites" hypothesis. However, *in vitro* measurements of transfer rates fail to account for the complexities of natural bacterial communities and may therefore underestimate conjugative transfer (144, 146). For example, *in situ* methods of detecting conjugative transfer using fluorescent reporter genes have estimated higher transfer rates than culture-based methods (146).

In addition, the use of laboratory host strains neglects the range of genotypic diversity in natural strains and their indigenous plasmids, which may affect conjugative transfer, segregational loss and plasmid fitness cost (35). In one study, plasmid R1 was combined with different strains of *E. coli* as donors, including some already possessing plasmids (45). The resulting conjugative transfer rates ranged over six orders of magnitude, and some of this variation was directly related to the presence of indigenous plasmids. In another study, two plasmids were maintained in a bacterial population over time despite estimates of low conjugative transfer rates (104). The results were explained by transitory derepression of the plasmid's *tra* operon, which can occur just after plasmid transfer to a new host. These observations suggest that there are some circumstances in which conjugative transfer in natural bacterial populations may be sufficient to allow plasmids to persist as parasites.

Even low conjugative transfer rates may be sufficient for plasmid persistence when there is variability in strain genotypes, a situation that more accurately simulates natural bacterial communities. In an experimental evolution study, Turner and colleagues periodically introduced plasmid-free cells from an evolving source population into an experimental population of plasmid-bearing and plasmid-free cells (162). Surprisingly,

transconjugants of the immigrant plasmid-free cells became prevalent in the experimental population, despite the fitness cost of the plasmid. The researchers showed that the source population had evolved a fitness advantage compared to the plasmid-free cells in the experimental population. When this new, fitter genotype was introduced, it acquired the plasmid by conjugative transfer. Although the plasmid's fitness cost reduced its growth advantage slightly, the new genotype was still able to sweep through the experimental population, effectively allowing the plasmid to hitchhike to high prevalence. This observation led to the proposal that selective sweeps in a natural bacterial community may allow plasmids to persist as parasites, as long as their transfer rate is high enough to ensure transfer to the fitter genotype.

Although there are clearly some circumstances in which plasmids may be maintained by horizontal transfer despite their cost to the host, the "plasmids as parasites" hypothesis cannot exclusively explain plasmid persistence in bacterial populations. What about the alternative hypothesis, which presents plasmids as mutualists that must carry genes beneficial to the host in order to persist? In their seminal study, Stewart and Levin acknowledged that selection for plasmid-borne genes often results in a growth rate advantage for plasmid-bearing cells which is also sufficient to overcome segregational loss and ensure plasmid persistence (149). But, selective pressure for plasmid-encoded functions is temporally and spatially variable; for instance, selection for a plasmid-encoded antibiotic resistance gene in the bacterial community of the human gastrointestinal tract only occurs when the human host is taking antibiotics. How does such variation in selective pressure affect plasmid persistence?

Eberhard suggested that sporadic selection actually favors the maintenance of adaptations to local and variable conditions on plasmids (50). He argued that to consider the selective advantage of the plasmid-borne gene from the level of the bacterial population ignores selection at the level of the gene itself, a variation on Dawkins' "selfish gene" theory (34). When a bacterial population is confronted with an environmental challenge, an adaptive plasmid-borne gene can propagate itself more rapidly than the same gene carried on the chromosome due to conjugative transfer or mobilization. In addition, a plasmid-borne gene has a higher mutation rate than a chromosomal gene because of the increased gene dosage obtained with multiple plasmid copies. This increases the likelihood of rapid evolution in response to environmental challenge. Considering plasmid persistence from the standpoint of selection on individual plasmid-borne genes contributed an important perspective to the debate.

In response, Bergstrom, Lipsitch and Levin revisited the original Stewart and Levin model and argued that selection for plasmid-borne genes, even if fluctuating, will eventually give rise to a genotype in which the beneficial gene has been transferred to the chromosome, thereby eliminating the need for a costly extrachromosomal element (9). One major assumption of their model is that recombination into the chromosome at a non-deleterious location occurs at a high enough rate and is followed by sufficient expression of the beneficial gene to provide this new genotype with a growth advantage over the genotype bearing the beneficial gene on a plasmid. Although this assumption lends itself well to an experimental evolution study, such long-term experimental tests have not yet been reported, therefore it is unclear whether these authors' conclusions are valid in natural bacterial communities.

Recent experimental tests of the effects of heterogeneous selection on plasmid persistence have indicated that plasmids can be maintained in discontinuous distributions in such complex situations. Slater and colleagues constructed a heterogeneous selective environment using cellulose fibers impregnated with mercury. They sprayed these fibers onto a mixed lawn of plasmid-free bacteria and bacteria carrying a large plasmid encoding mercury resistance (144). The plasmid-free strain and the plasmid were labeled with fluorescent reporter genes. Using fluorescence microscopy, the researchers determined how the mercury gradient around each cellulose fiber affected plasmid persistence. They found that in an environment of heterogeneous selection, the plasmid conferred a sporadic benefit on host cells. When considered as a population, the plasmid was under negative frequency-dependent selection, meaning there was a greater increase in positive selection for the plasmid when plasmid-bearing cells were initially rare. Outcomes such as these suggest that, despite the objections of Bergstrom and colleagues, variable selection for plasmid-borne genes can maintain plasmids in a population under certain circumstances.

From the discussion above, we see that both the "plasmids as parasites" and "plasmids as mutualists" hypotheses may be valid depending on the situation. The theoretical and experimental explorations of these two hypotheses have revealed a spectrum of plasmid existence conditions rather than two opposing possibilities. In the last few years, a push to integrate ecological theory with studies of microbial systems (133) has increased awareness of how complexity in natural bacterial communities affects plasmid population biology (163). This awareness is reflected in recent developments of new mathematical models of plasmid persistence that incorporate

complex factors such as spatial and temporal variation (86, 100, 131). Such models will be valuable for considering plasmid maintenance in the ecologically complex conditions experienced by natural bacterial communities. This recent work also begins to account for possible variation of the plasmid itself; for example, one of the models described by Ponciano and colleagues allows the fitness cost of the plasmid to change over time. This is an important foundation for incorporating plasmid evolution into theoretical considerations of plasmid persistence.

Plasmid Evolution

We cannot really fault the existing models for seeking to simplify plasmid population biology. It is far easier to consider "single, inviolate" plasmids (149) than to incorporate the possibility that your plasmid of interest may rapidly change its accessory phenotypes or even its conjugative transfer capabilities! However, the capacity for plasmids to evolve may be one of the driving forces influencing plasmid persistence in bacterial populations (71), and it certainly impacts host adaptation and evolution. Therefore, plasmid evolution deserves a more central place in the fields of plasmid biology in particular and microbial evolution in general.

Plasmid evolution can occur in many ways. As with any DNA molecule, point mutations, insertions, deletions and rearrangements affect genotypic variation of plasmids. Homologous recombination between different plasmids or plasmids and chromosomes generates new alleles or new combinations of genes. Site-specific recombination adds and removes gene cassettes from plasmid-borne integrons. Transposition leads to the gain or loss of entire accessory elements. As mentioned

previously, these evolutionary forces result in highly mosaic plasmids, with backbone and accessory genes acquired from many different sources.

Studies of plasmid evolution have taken both prospective and retrospective approaches. Before high throughput sequencing allowed comparative analysis of plasmids to identify evolutionary events, prospective experimental evolution studies revealed that key plasmid characteristics such as fitness cost can change over time. Building on the pioneering work of Richard Lenski, various authors showed that over the course of many generations, the fitness costs of both small, nonconjugative plasmids and large, conjugative plasmids are reduced or eliminated, even in the absence of selection for plasmid-borne genes (14, 30, 36, 44, 72, 112). It is striking that this possibility was only recently incorporated into the theoretical models of plasmid persistence described above (131), even though there is now significant experimental evidence contradicting the original models' assumption that plasmids must always impose a fitness cost on their hosts.

The adaptive trajectories observed in the experimental evolution studies varied. In some cases, either the plasmid or the bacterial chromosome acquired compensatory mutations that reduced plasmid fitness cost (14, 30). In other cases, true coevolution occurred, with compensatory mutations on both the chromosome and the plasmid (112). Dionisio and colleagues observed that adaptations on the plasmid reduced its fitness cost even in naïve hosts of a different species (44), revealing that plasmid evolution in a single host may increase the chances of plasmid persistence in other members of natural bacterial communities. Top and coworkers have simulated plasmid evolution in natural bacterial communities with a "host-switching" protocol that alternates the plasmid

between two different bacterial hosts during the experimental evolution (36, 72). In two separate studies, plasmids undergoing "host-switching" evolved reduced fitness cost, and in one of the studies, frequent conjugative transfer actually accelerated plasmid adaptation. This study was the only one to identify the exact mutation responsible for plasmid fitness changes: a single point mutation in the *trbC* mating pair formation gene. Prospective plasmid evolution studies such as these continue to be valuable assets for evaluating hypotheses regarding plasmid biology and for learning more about plasmid evolution.

In addition to prospective studies, gene and whole genome sequencing has presented a wealth of data for retrospective investigations of plasmid evolution. Phylogenetic analyses of plasmid protein families have identified many instances of sequence divergence in both backbone and accessory genes (125). Although these are too numerous to detail here, it is important to note that even point mutations can have dramatic effects, for instance in the generation of new plasmid incompatibility families (155) or the reduction of plasmid fitness cost in the experimental evolution study described above (36). For the purposes of this review, I will offer examples of how recombination affects plasmid evolution.

Recombination plays a significant role in the variation of plasmid backbone genes. Nucleotide sequence analysis of IncF plasmids isolated from *E. coli* showed that the gene trees for four plasmid genes encoding replication and transfer functions were not congruent, indicating that recombination events had shuffled the backbone genes of the plasmids. In addition, the plasmid genes showed evidence of higher intragenic recombination compared to chromosomal genes (16). In a comparison of two IncQ

plasmids, a recombination event appeared to have assembled a novel suite of mobilization genes (134). As a result, although the backbones of the plasmids were generally similar, each plasmid was mobilized most efficiently by a different conjugative plasmid. Chi-mediated recombination was implicated in the formation of mosaic RepFIIA replicons, which impacted plasmid incompatibility (124). Comparative analysis of *Salmonella* virulence plasmids has indicated that four serovar-specific plasmids diverged from a common ancestor through recombination involving backbone genes (26, 75). This and other evidence suggests that the "core" components of plasmids undergo substantial evolution due to recombination events.

The evolution of plasmid backbone genes may be outpaced by evolution in plasmid accessory content. Much of a plasmid's "flexible" genome includes mobile elements such as transposons and integrons, which encode recombinases mediating their movement between plasmids and chromosomes. In a prospective experimental evolution study, Condit and Levin observed that it took only 150 generations for a population with two plasmids bearing different antibiotic resistance transposons to be replaced by a population with both transposons combined on one plasmid (27). This finding of rapid flux in accessory elements is supported by retrospective analyses of complete plasmid sequences. Revilla and colleagues reported that the ~30 kb backbone regions of five IncW plasmids were ≥97% identical at the nucleotide sequence level, but their accessory content was different (135). Three of the plasmids were practically identical with the exception of their integron, which had different gene cassettes. The other two plasmids had no integron but each possessed two transposons not found on the other plasmids. A

similar phenomenon was described for the IncHI plasmids R478, R27 and pHCM1: highly similar backbones contrasted with dramatically different accessory content (63).

Evolution of plasmid genomes alters both backbone and accessory genes. How this evolution affects the natural variation of plasmids or their distribution in bacterial communities is still unclear.

OBJECTIVES OF THE DISSERTATION

Although it is clear that the backbone functions and accessory phenotypes encoded by plasmids are quite diverse, the natural variation of plasmids in bacterial populations is not well-understood. Most of our knowledge comes from *in vitro* studies of specific plasmids isolated from enterobacteria, such as F and R1. Much less is known about plasmids from Gram-positive bacteria or even about the broader plasmid pool accessible to the well-studied enterobacteria. Without this information, our picture of the ecology and evolution of plasmids is incomplete. For my dissertation research, I investigated the natural variation of large plasmids using both a sequence-based approach and a collection-based approach. In Chapter 2, I present analyses of four novel large plasmid sequences, including three plasmids from Gram-positive bacteria. In Chapter 3, I present an assessment of natural plasmid variation in reference collections of *Escherichia coli* and *Salmonella*. The research discussed in these two chapters explores the diversity of natural plasmids and their distribution in bacterial communities. To conclude in Chapter 4, I discuss how the outcomes of the dissertation impact considerations of plasmid evolution and the role of plasmids in microbial ecology and evolution.

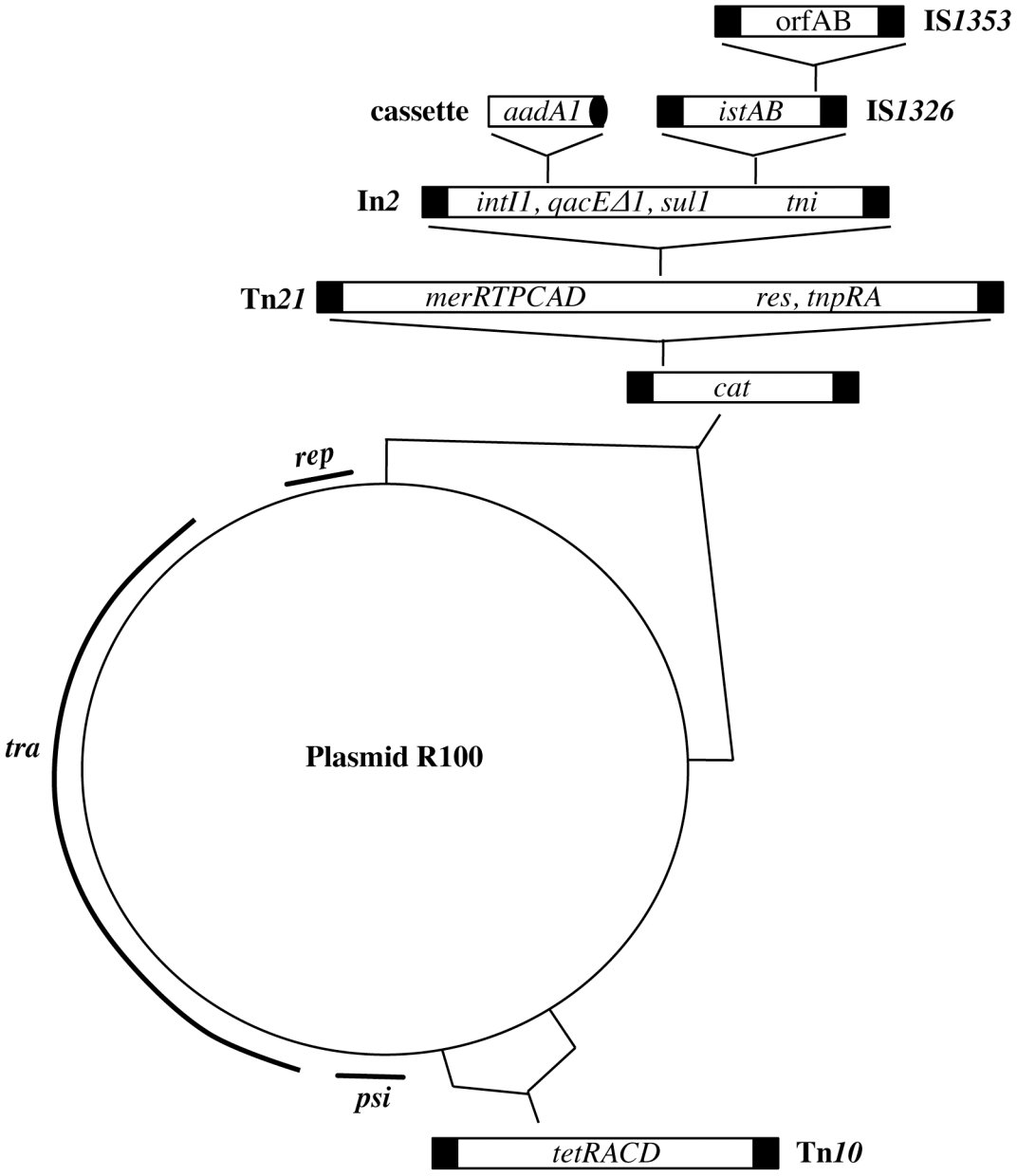


FIGURE 1.1. Modularity of backbone and accessory genes of plasmid R100. Backbone modules are denoted as bold lines, with the functions of each module in bold italics. Insertions of accessory modules are indicated by brackets. More detailed views of the accessory elements are shown to demonstrate their nested structure.

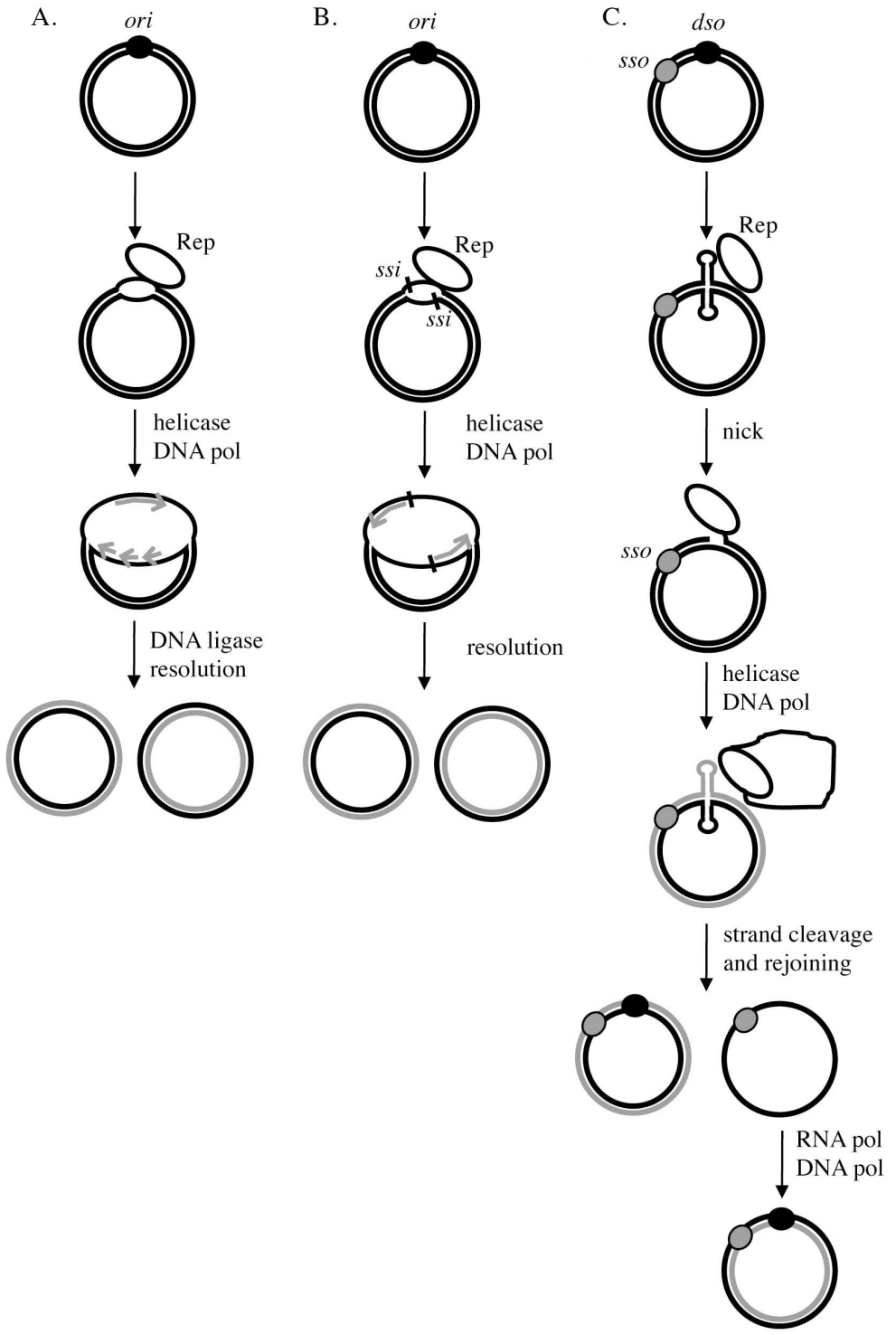


FIGURE 1.2. Modes of plasmid replication. A) Theta type replication. B) Strand displacement replication. C) Rolling circle replication.

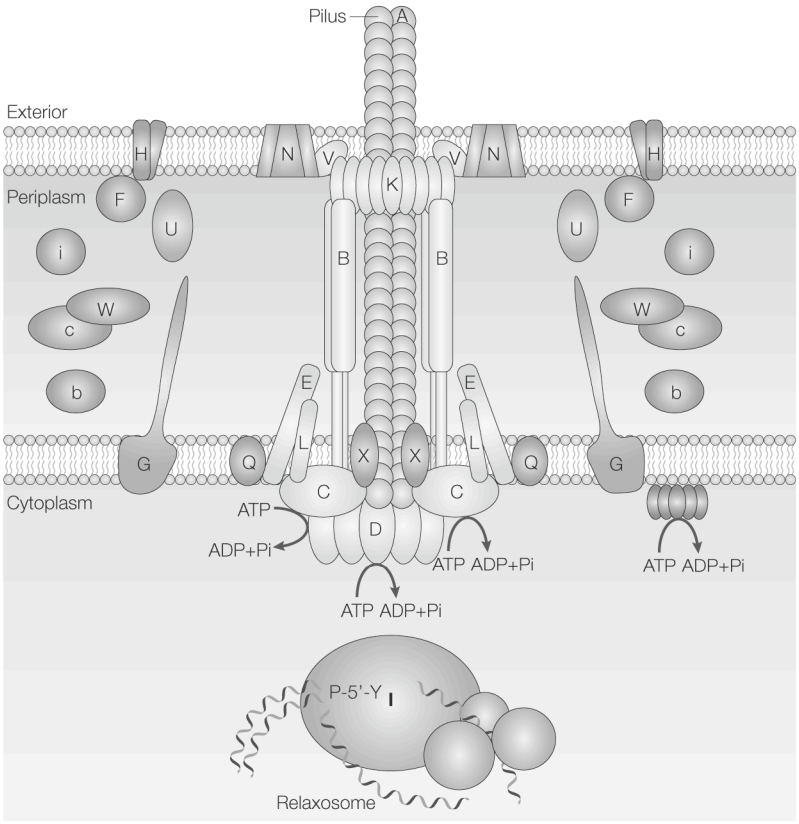


FIGURE 1.3. The F-like conjugative system. The relaxase TraI is complexed with DNA in the cytoplasm. The coupling protein TraD is attached to the T4SS apparatus on the cytoplasmic side of the inner membrane. The multiprotein complex comprising the T4SS apparatus spans the inner membrane, periplasm and outer membrance, and the pilus composed of TraA pilin subunits extends into the extracellular space. Other proteins involved in pilus assembly and mating pair stabilization are shown to the sides of the T4SS apparatus. Figure from reference 56.

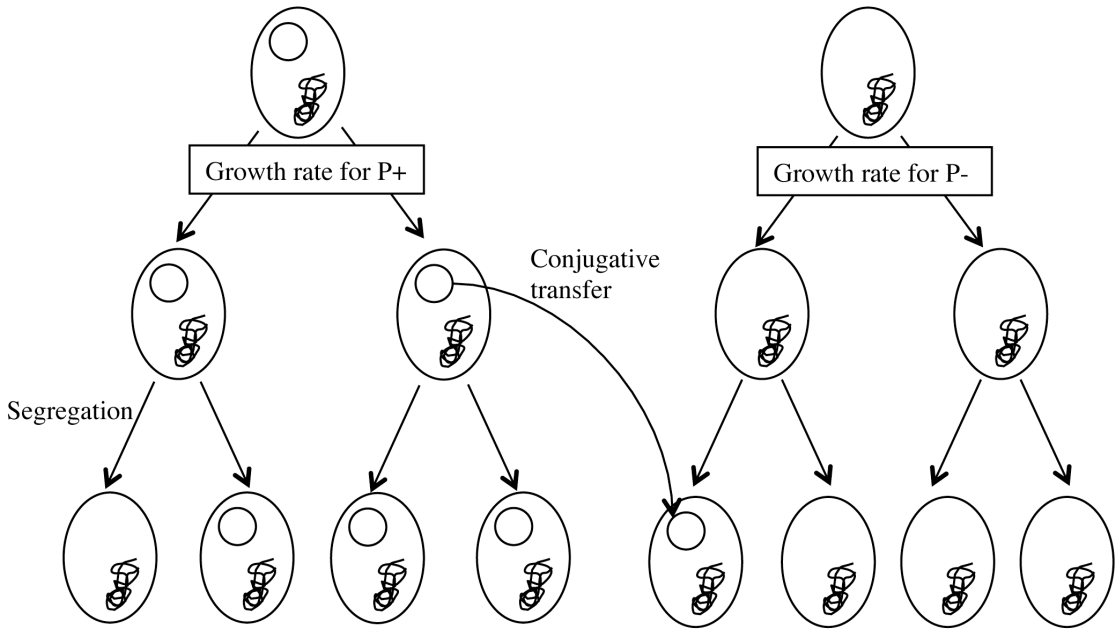


FIGURE 1.4. Factors affecting plasmid persistence in bacterial populations. Two bacterial cells are isogenic with the exception of a plasmid, shown as a circle. The three factors influencing the plasmid's maintenance in the population are emphasized: (1) the growth rate difference between the plasmid-bearing and plasmid-free cells, (2) segregrational loss of the plasmid, which converts a plasmid-bearing to a plasmid-free strain, and (3) conjugative transfer of the plasmid, which converts a plasmid-free to a plasmid-bearing strain.

REFERENCES

1. **Alonso, J. C., C. A. Stiege, R. H. Tailor, and J. F. Viret.** 1988. Functional analysis of the *dna* (Ts) mutants of *Bacillus subtilis*: plasmid pUB110 replication as a model system. Mol Gen Genet **214:**482-9.
2. **Amabile-Cuevas, C. F., and M. E. Chicurel.** 1992. Bacterial plasmids and gene flux. Cell **70:**189-99.
3. **Anand, S. P., and S. A. Khan.** 2004. Structure-specific DNA binding and bipolar helicase activities of PcrA. Nucleic Acids Res **32:**3190-7.
4. **Asano, K., and K. Mizobuchi.** 1998. Copy number control of IncIalpha plasmid ColIb-P9 by competition between pseudoknot formation and antisense RNA binding at a specific RNA site. EMBO J **17:**5201-13.
5. **Austin, S., and A. Abeles.** 1983. Partition of unit-copy miniplasmids to daughter cells. II. The partition region of miniplasmid P1 encodes an essential protein and a centromere-like site at which it acts. J Mol Biol **169:**373-87.
6. **Austin, S., M. Ziese, and N. Sternberg.** 1981. A novel role for site-specific recombination in maintenance of bacterial replicons. Cell **25:**729-36.
7. **Barkay, T., S. M. Miller, and A. O. Summers.** 2003. Bacterial mercury resistance from atoms to ecosystems. FEMS Microbiol Rev **27:**355-84.
8. **Bennett, P. M.** 2008. Plasmid encoded antibiotic resistance: acquisition and transfer of antibiotic resistance genes in bacteria. Br J Pharmacol **153 Suppl 1:**S347-57.

9. **Bergstrom, C. T., M. Lipsitch, and B. R. Levin.** 2000. Natural selection, infectious transfer and the existence conditions for bacterial plasmids. Genetics **155:**1505-19.

10. **Bernard, P., and M. Couturier.** 1992. Cell killing by the F plasmid CcdB protein involves poisoning of DNA-topoisomerase II complexes. J Mol Biol **226:**735-45.

11. **Bernard, P., K. E. Kezdy, L. Van Melderen, J. Steyaert, L. Wyns, M. L. Pato, P. N. Higgins, and M. Couturier.** 1993. The F plasmid CcdB protein induces efficient ATP-dependent DNA cleavage by gyrase. J Mol Biol **234:**534-41.

12. **Birch, P., and S. A. Khan.** 1992. Replication of single-stranded plasmid pT181 DNA *in vitro*. Proc Natl Acad Sci U S A **89:**290-4.

13. **Bjorland, J., T. Steinum, M. Sunde, S. Waage, and E. Heir.** 2003. Novel plasmid-borne gene *qacJ* mediates resistance to quaternary ammonium compounds in equine *Staphylococcus aureus*, *Staphylococcus simulans*, and *Staphylococcus intermedius*. Antimicrob Agents Chemother **47:**3046-52.

14. **Bouma, J. E., and R. E. Lenski.** 1988. Evolution of a bacteria/plasmid association. Nature **335:**351-2.

15. **Bowers, L. M., and M. Filutowicz.** 2008. Cooperative binding mode of the inhibitors of R6K replication, pi dimers. J Mol Biol **377:**609-15.

16. **Boyd, E. F., C. W. Hill, S. M. Rich, and D. L. Hartl.** 1996. Mosaic structure of plasmids from natural populations of *Escherichia coli*. Genetics **143:**1091-100.

17. **Bradley, D. E., D. E. Taylor, and D. R. Cohen.** 1980. Specification of surface mating systems among conjugative drug resistance plasmids in *Escherichia coli* K-12. J Bacteriol **143:**1466-70.

18. **Bramhill, D., and A. Kornberg.** 1988. A model for initiation at origins of DNA replication. Cell **54:**915-8.

19. **Brantl, S.** 2004. Plasmid replication control by antisense RNAs, p. 47-62. *In* B. E. Funnell and G. J. Phillips (ed.), Plasmid Biology. ASM Press, Washington, D.C.

20. **Byrd, D. R., and S. W. Matson.** 1997. Nicking by transesterification: the reaction catalysed by a relaxase. Mol Microbiol **25:**1011-22.

21. **Cabezon, E., and F. de la Cruz.** 2006. TrwB: an F(1)-ATPase-like molecular motor involved in DNA transport during bacterial conjugation. Res Microbiol **157:**299-305.

22. **Cabezon, E., J. I. Sastre, and F. de la Cruz.** 1997. Genetic evidence of a coupling role for the TraG protein family in bacterial conjugation. Mol Gen Genet **254:**400-6.

23. **Carattoli, A., A. Bertini, L. Villa, V. Falbo, K. L. Hopkins, and E. J. Threlfall.** 2005. Identification of plasmids by PCR-based replicon typing. J Microbiol Methods **63:**219-28.

24. **Chang, T. L., A. Naqvi, S. P. Anand, M. G. Kramer, R. Munshi, and S. A. Khan.** 2002. Biochemical characterization of the *Staphylococcus aureus* PcrA helicase and its role in plasmid rolling circle replication. J Biol Chem **277:**45880-6.

25. **Chattoraj, D. K.** 2000. Control of plasmid DNA replication by iterons: no longer paradoxical. Mol Microbiol **37:**467-76.

26. **Chu, C., Y. Feng, A. C. Chien, S. Hu, C. H. Chu, and C. H. Chiu.** 2008. Evolution of genes on the *Salmonella* virulence plasmid phylogeny revealed from sequencing of the virulence plasmids of *S. enterica* serotype Dublin and comparative analysis. Genomics **92:**339-43.

27. **Condit, R., and B. R. Levin.** 1990. The evolution of plasmids carrying multiple resistance genes: the role of segregation, transposition, and homologous recombination. Am Nat **135:**573-596.

28. **Courvalin, P.** 2008. New plasmid-mediated resistances to antimicrobial agents. Arch Microbiol **189:**289-91.

29. **Couturier, M., F. Bex, P. L. Bergquist, and W. K. Maas.** 1988. Identification and classification of bacterial plasmids. Microbiol Rev **52:**375-95.

30. **Dahlberg, C., and L. Chao.** 2003. Amelioration of the cost of conjugative plasmid carriage in *Eschericha coli* K12. Genetics **165:**1641-9.

31. **Das, N., M. Valjavec-Gratian, A. N. Basuray, R. A. Fekete, P. P. Papp, J. Paulsson, and D. K. Chattoraj.** 2005. Multiple homeostatic mechanisms in the control of P1 plasmid replication. Proc Natl Acad Sci U S A **102:**2856-61.

32. **Datta, N.** 1979. Plasmid classification: incompatibility grouping, p. 3-12. *In* K. N. Timmis and A. Puhlers (ed.), Plasmids of Medical, Environmental and Commercial Importance. Elseviers/North Holland Publishing Co., Amsterdam.

33. **Datta, N., and R. W. Hedges.** 1972. R factors identified in Paris, some conferring gentamicin resistance, constitute a new compatibility group. Ann Inst Pasteur (Paris) **123:**849-52.

34. **Dawkins, R.** 1976. The selfish gene. Oxford University Press, Oxford.

35. **De Gelder, L., J. M. Ponciano, P. Joyce, and E. M. Top.** 2007. Stability of a promiscuous plasmid in different hosts: no guarantee for a long-term relationship. Microbiology **153:**452-63.

36. **De Gelder, L., J. J. Williams, J. M. Ponciano, M. Sota, and E. M. Top.** 2008. Adaptive plasmid evolution results in host-range expansion of a broad-host-range plasmid. Genetics **178:**2179-90.

37. **del Solar, G., and M. Espinosa.** 2000. Plasmid copy number control: an ever-growing story. Mol Microbiol **37:**492-500.

38. **del Solar, G., R. Giraldo, M. J. Ruiz-Echevarria, M. Espinosa, and R. Diaz-Orejas.** 1998. Replication and control of circular bacterial plasmids. Microbiol Mol Biol Rev **62:**434-64.

39. **del Solar, G., M. Moscoso, and M. Espinosa.** 1993. Rolling circle-replicating plasmids from gram-positive and gram-negative bacteria: a wall falls. Mol Microbiol **8:**789-96.

40. **Dellis, S., and M. Filutowicz.** 1991. Integration host factor of *Escherichia coli* reverses the inhibition of R6K plasmid replication by pi initiator protein. J Bacteriol **173:**1279-86.

41. **Derman, A. I., G. Lim-Fong, and J. Pogliano.** 2008. Intracellular mobility of plasmid DNA is limited by the ParA family of partitioning systems. Mol Microbiol **67:**935-46.

42. **Diederix, R. E., C. Davila, R. Giraldo, and M. P. Lillo.** 2008. Fluorescence studies of the replication initiator protein RepA in complex with operator and iteron sequences and free in solution. FEBS J **275:**5393-407.

43. **Dijkstra, A. J., and W. Keck.** 1996. Peptidoglycan as a barrier to transenvelope transport. J Bacteriol **178:**5555-62.

44. **Dionisio, F., I. C. Conceicao, A. C. Marques, L. Fernandes, and I. Gordo.** 2005. The evolution of a conjugative plasmid and its ability to increase bacterial fitness. Biol Lett **1:**250-2.

45. **Dionisio, F., I. Matic, M. Radman, O. R. Rodrigues, and F. Taddei.** 2002. Plasmids spread very fast in heterogeneous bacterial communities. Genetics **162:**1525-32.

46. **Dixon, T. C., M. Meselson, J. Guillemin, and P. C. Hanna.** 1999. Anthrax. N Engl J Med **341:**815-26.

47. **Draper, O., C. E. Cesar, C. Machon, F. de la Cruz, and M. Llosa.** 2005. Site-specific recombinase and integrase activities of a conjugative relaxase in recipient cells. Proc Natl Acad Sci U S A **102:**16385-90.

48. **Durrenberger, M. B., W. Villiger, and T. Bachi.** 1991. Conjugational junctions: morphology of specific contacts in conjugating *Escherichia coli* bacteria. J Struct Biol **107:**146-56.

49. **Dwyer, D. J., M. A. Kohanski, B. Hayete, and J. J. Collins.** 2007. Gyrase inhibitors induce an oxidative damage cellular death pathway in *Escherichia coli*. Mol Syst Biol **3:**91.

50. **Eberhard, W. G.** 1990. Evolution in bacterial plasmids and levels of selection. Q Rev Biol **65:**3-22.

51. **Ebersbach, G., and K. Gerdes.** 2005. Plasmid segregation mechanisms. Annu Rev Genet **39:**453-79.

52. **Edgar, R., D. K. Chattoraj, and M. Yarmolinsky.** 2001. Pairing of P1 plasmid partition sites by ParB. Mol Microbiol **42:**1363-70.

53. **Eisenbrandt, R., M. Kalkum, E. M. Lai, R. Lurz, C. I. Kado, and E. Lanka.** 1999. Conjugative pili of IncP plasmids, and the Ti plasmid T pilus are composed of cyclic subunits. J Biol Chem **274:**22548-55.

54. **Erdmann, N., T. Petroff, and B. E. Funnell.** 1999. Intracellular localization of P1 ParB protein depends on ParA and *parS*. Proc Natl Acad Sci U S A **96:**14905-10.

55. **Fredericq, P., and E. Dehalle.** 1969. Recombination of R factors with Col factors and chromosomal genes in *Escherichia coli*. J Gen Microbiol **55:**xi-xii.

56. **Frost, L. S., R. Leplae, A. O. Summers, and A. Toussaint.** 2005. Mobile genetic elements: the agents of open source evolution. Nat Rev Microbiol **3:**722-32.

57. **Funnell, B. E.** 1991. The P1 plasmid partition complex at *parS*. The influence of *Escherichia coli* integration host factor and of substrate topology. J Biol Chem **266:**14328-37.

58. **Funnell, B. E., and R. A. Slavcev.** 2004. Partition systems of bacterial plasmids, p. 81-103. *In* B. E. Funnell and G. J. Phillips (ed.), Plasmid Biology. ASM Press, Washington, D.C.

59. **Garner, E. C., C. S. Campbell, D. B. Weibel, and R. D. Mullins.** 2007. Reconstitution of DNA segregation driven by assembly of a prokaryotic actin homolog. Science **315:**1270-4.

60. **Gerdes, K., A. P. Gultyaev, T. Franch, K. Pedersen, and N. D. Mikkelsen.** 1997. Antisense RNA-regulated programmed cell death. Annu Rev Genet **31:**1-31.

61. **Gerdes, K., and S. Molin.** 1986. Partitioning of plasmid R1. Structural and functional analysis of the *parA* locus. J Mol Biol **190:**269-79.

62. **Gerdes, K., J. Moller-Jensen, and R. Bugge Jensen.** 2000. Plasmid and chromosome partitioning: surprises from phylogeny. Mol Microbiol **37:**455-66.

63. **Gilmour, M. W., N. R. Thomson, M. Sanders, J. Parkhill, and D. E. Taylor.** 2004. The complete nucleotide sequence of the resistance plasmid R478: defining the backbone components of incompatibility group H conjugative plasmids through comparative genomics. Plasmid **52:**182-202.

64. **Giraldo, R., and R. Diaz.** 1992. Differential binding of wild-type and a mutant RepA protein to *oriR* sequence suggests a model for the initiation of plasmid R1 replication. J Mol Biol **228:**787-802.

65. **Gordon, D. M.** 1992. Rate of plasmid transfer among *Escherichia coli* strains isolated from natural populations. J Gen Microbiol **138:**17-21.

66. **Griffiths, A. J.** 1995. Natural plasmids of filamentous fungi. Microbiol Rev **59:**673-85.

67. **Grohmann, E., G. Muth, and M. Espinosa.** 2003. Conjugative plasmid transfer in gram-positive bacteria. Microbiol Mol Biol Rev **67:**277-301, table of contents.

68. **Haase, J., M. Kalkum, and E. Lanka.** 1996. TrbK, a small cytoplasmic membrane lipoprotein, functions in entry exclusion of the IncP alpha plasmid RP4. J Bacteriol **178:**6720-9.

69. **Hallet, B., V. Vanhooff, and F. Cornet.** 2004. DNA site-specific resolution systems, p. 145-180. *In* B. E. Funnell and G. J. Phillips (ed.), Plasmid Biology. ASM Press, Washington, D.C.

70. **Hayes, F.** 2003. Toxins-antitoxins: plasmid maintenance, programmed cell death, and cell cycle arrest. Science **301:**1496-9.

71. **Heuer, H., Z. Abdo, and K. Smalla.** 2008. Patchy distribution of flexible genetic elements in bacterial populations mediates robustness to environmental uncertainty. FEMS Microbiol Ecol **65:**361-371.

72. **Heuer, H., R. E. Fox, and E. M. Top.** 2007. Frequent conjugative transfer accelerates adaptation of a broad-host-range plasmid to an unfavorable *Pseudomonas putida* host. FEMS Microbiol Ecol **59:**738-48.

73. **Hill, T. M., M. L. Tecklenburg, A. J. Pelletier, and P. L. Kuempel.** 1989. *tus*, the trans-acting gene required for termination of DNA replication in *Escherichia coli*, encodes a DNA-binding protein. Proc Natl Acad Sci U S A **86:**1593-7.

74. **Honda, Y., H. Sakai, H. Hiasa, K. Tanaka, T. Komano, and M. Bagdasarian.** 1991. Functional division and reconstruction of a plasmid replication origin:

molecular dissection of the *oriV* of the broad-host-range plasmid RSF1010. Proc Natl Acad Sci U S A **88:**179-83.

75. **Hong, S. F., C. H. Chiu, C. Chu, Y. Feng, and J. T. Ou.** 2008. Complete nucleotide sequence of a virulence plasmid of *Salmonella enterica* serovar Dublin and its phylogenetic relationship to the virulence plasmids of serovars Choleraesuis, Enteritidis and Typhimurium. FEMS Microbiol Lett **282:**39-43.

76. **Iordanescu, S., and R. Basheer.** 1991. The *Staphylococcus aureus* mutation *pcrA3* leads to the accumulation of pT181 replication initiation complexes. J Mol Biol **221:**1183-9.

77. **Jensen, R. B., R. Lurz, and K. Gerdes.** 1998. Mechanism of DNA segregation in prokaryotes: replicon pairing by *parC* of plasmid R1. Proc Natl Acad Sci U S A **95:**8550-5.

78. **Jin, R., M. E. Fernandez-Beros, and R. P. Novick.** 1997. Why is the initiation nick site of an AT-rich rolling circle plasmid at the tip of a GC-rich cruciform? EMBO J **16:**4456-66.

79. **Johnson, T. J., Y. M. Wannemuehler, S. J. Johnson, C. M. Logue, D. G. White, C. Doetkott, and L. K. Nolan.** 2007. Plasmid replicon typing of commensal and pathogenic *Escherichia coli* isolates. Appl Environ Microbiol **73:**1976-83.

80. **Khan, S. A.** 2005. Plasmid rolling-circle replication: highlights of two decades of research. Plasmid **53:**126-36.

81. **Khan, S. A.** 2004. Rolling-circle replication, p. 63-78. *In* B. E. Funnell and G. J. Phillips (ed.), Plasmid Biology. ASM Press, Washington, D.C.

82. **Kobayashi, I.** 2004. Genetic addiction: a principle of gene symbiosis in a genome, p. 105-144. *In* B. E. Funnell and G. J. Phillips (ed.), Plasmid Biology. ASM Press, Washington, D.C.

83. **Kobori, H., C. W. Sullivan, and H. Shizuya.** 1984. Bacterial plasmids in Antarctic natural microbial assemblages. Appl Environ Microbiol **48:**515-518.

84. **Koepsel, R. R., R. W. Murray, W. D. Rosenblum, and S. A. Khan.** 1985. The replication initiator protein of plasmid pT181 has sequence-specific endonuclease and topoisomerase-like activities. Proc Natl Acad Sci U S A **82:**6845-9.

85. **Kramer, M. G., S. A. Khan, and M. Espinosa.** 1997. Plasmid rolling circle replication: identification of the RNA polymerase-directed primer RNA and requirement for DNA polymerase I for lagging strand synthesis. EMBO J **16:**5784-95.

86. **Krone, S. M., R. Lu, R. Fox, H. Suzuki, and E. M. Top.** 2007. Modelling the spatial dynamics of plasmid transfer and persistence. Microbiology **153:**2803-16.

87. **Kruger, R., S. A. Rakowski, and M. Filutowicz.** 2004. Participating elements in the replication of iteron-containing plasmids, p. 25-45. *In* B. E. Funnell and G. J. Phillips (ed.), Plasmid Biology. ASM Press, Washington, D.C.

88. **Lane, D., R. de Feyter, M. Kennedy, S. H. Phua, and D. Semon.** 1986. D protein of miniF plasmid acts as a repressor of transcription and as a site-specific resolvase. Nucleic Acids Res **14:**9713-28.

89. **Lanka, E., and B. M. Wilkins.** 1995. DNA processing reactions in bacterial conjugation. Annu Rev Biochem **64:**141-69.

90. **Lawley, T., B. M. Wilkins, and L. S. Frost.** 2004. Bacterial conjugation in Gram-negative bacteria, p. 203-256. *In* B. E. Funnell and G. J. Phillips (ed.), Plasmid Biology. ASM Press, Washington, D.C.

91. **Lawley, T. D., W. A. Klimke, M. J. Gubbins, and L. S. Frost.** 2003. F factor conjugation is a true type IV secretion system. FEMS Microbiol Lett **224:**1-15.

92. **Lederberg, J.** 1952. Cell genetics and hereditary symbiosis. Physiol Rev **32:**403-30.

93. **Lederberg, J.** 1998. Plasmid (1952-1997). Plasmid **39:**1-9.

94. **Lederberg, J., L. L. Cavalli, and E. M. Lederberg.** 1952. Sex compatibility in *Escherichia coli*. Genetics **37:**720-30.

95. **Leplae, R., G. Lima-Mendez, and A. Toussaint.** 2006. A first global analysis of plasmid encoded proteins in the ACLAME database. FEMS Microbiol Rev **30:**980-94.

96. **Levin, B. R., and F. M. Stewart.** 1980. The population biology of bacterial plasmids: a priori conditions for the existence of mobilizable nonconjugative factors. Genetics **94:**425-43.

97. **Levin, B. R., F. M. Stewart, and V. A. Rice.** 1979. The kinetics of conjugative plasmid transmission: fit of a simple mass action model. Plasmid **2:**247-60.

98. **Libby, S. J., M. Lesnick, P. Hasegawa, E. Weidenhammer, and D. G. Guiney.** 2000. The *Salmonella* virulence plasmid *spv* genes are required for cytopathology in human monocyte-derived macrophages. Cell Microbiol **2:**49-58.

99. **Liebert, C. A., R. M. Hall, and A. O. Summers.** 1999. Transposon Tn*21*, flagship of the floating genome. Microbiol Mol Biol Rev **63:**507-22.

100. **Lili, L. N., N. F. Britton, and E. J. Feil.** 2007. The persistence of parasitic plasmids. Genetics **177:**399-405.

101. **Lilley, A., P. Young, and M. Bailey.** 2000. Bacterial population genetics: do plasmids maintain bacterial diversity and adaptation?, p. 287-300. *In* C. M. Thomas (ed.), The Horizontal Gene Pool: Bacterial Plasmids and Gene Spread. Harwood Academic Publishers, Amsterdam.

102. **Llosa, M., F. X. Gomis-Ruth, M. Coll, and F. de la Cruz Fd.** 2002. Bacterial conjugation: a two-step mechanism for DNA transport. Mol Microbiol **45:**1-8.

103. **Lu, J., and L. S. Frost.** 2005. Mutations in the C-terminal region of TraM provide evidence for *in vivo* TraM-TraD interactions during F-plasmid conjugation. J Bacteriol **187:**4767-73.

104. **Lundquist, P. D., and B. R. Levin.** 1986. Transitory derepression and the maintenance of conjugative plasmids. Genetics **113:**483-97.

105. **Majdalani, N., and K. Ippen-Ihler.** 1996. Membrane insertion of the F-pilin subunit is Sec independent but requires leader peptidase B and the proton motive force. J Bacteriol **178:**3742-7.

106. **Maki, S., S. Takiguchi, T. Miki, and T. Horiuchi.** 1992. Modulation of DNA supercoiling activity of *Escherichia coli* DNA gyrase by F plasmid proteins. Antagonistic actions of LetA (CcdA) and LetD (CcdB) proteins. J Biol Chem **267:**12244-51.

107. **Manen, D., L. C. Upegui-Gonzalez, and L. Caro.** 1992. Monomers and dimers of the RepA protein in plasmid pSC101 replication: domains in RepA. Proc Natl Acad Sci U S A **89:**8923-7.

108. **Manning, P. A., L. Beutin, and M. Achtman.** 1980. Outer membrane of *Escherichia coli*: properties of the F sex factor *traT* protein which is involved in surface exclusion. J Bacteriol **142:**285-94.

109. **Martinez, E., D. Romero, and R. Palacios.** 1990. The *Rhizobium* genome. Crit Rev Plant Sci **9:**59-93.

110. **McEachern, M. J., M. A. Bott, P. A. Tooker, and D. R. Helinski.** 1989. Negative control of plasmid R6K replication: possible role of intermolecular coupling of replication origins. Proc Natl Acad Sci U S A **86:**7942-6.

111. **Miteva, V. I., P. P. Sheridan, and J. E. Brenchley.** 2004. Phylogenetic and physiological diversity of microorganisms isolated from a deep Greenland glacier ice core. Appl Environ Microbiol **70:**202-13.

112. **Modi, R. I., and J. Adams.** 1991. Coevolution in bacterial-plasmid populations. Evolution **45:**656-667.

113. **Moore, D., C. M. Hamilton, K. Maneewannakul, Y. Mintz, L. S. Frost, and K. Ippen-Ihler.** 1993. The *Escherichia coli* K-12 F plasmid gene *traX* is required for acetylation of F pilin. J Bacteriol **175:**1375-83.

114. **Moscoso, M., G. del Solar, and M. Espinosa.** 1995. *In vitro* recognition of the replication origin of pLS1 and of plasmids of the pLS1 family by the RepB initiator protein. J Bacteriol **177:**7041-9.

115. **Noirot, P., J. Bargonetti, and R. P. Novick.** 1990. Initiation of rolling-circle replication in pT181 plasmid: initiator protein enhances cruciform extrusion at the origin. Proc Natl Acad Sci U S A **87:**8560-4.

116. **Nordstrom, K.** 1990. Control of plasmid replication--how do DNA iterons set the replication frequency? Cell **63:**1121-4.

117. **Nordstrom, K., and S. J. Austin.** 1989. Mechanisms that contribute to the stable segregation of plasmids. Annu Rev Genet **23:**37-69.

118. **Novick, R. P.** 1987. Plasmid incompatibility. Microbiol Rev **51:**381-95.

119. **Novick, R. P., S. Iordanescu, S. J. Projan, J. Kornblum, and I. Edelman.** 1989. pT181 plasmid replication is regulated by a countertranscript-driven transcriptional attenuator. Cell **59:**395-404.

120. **Novick, R. P., and M. H. Richmond.** 1965. Nature and interactions of the genetic elements governing penicillinase synthesis in *Staphylococcus aureus*. J Bacteriol **90:**467-80.

121. **Ochman, H., J. G. Lawrence, and E. A. Groisman.** 2000. Lateral gene transfer and the nature of bacterial innovation. Nature **405:**299-304.

122. **Ogawa, N., A. M. Chakrabarty, and O. Zaborina.** 2004. Degradative plasmids, p. 341-376. *In* B. E. Funnell and G. J. Phillips (ed.), Plasmid Biology. ASM Press, Washington, D.C.

123. **Ogura, T., and S. Hiraga.** 1983. Partition mechanism of F plasmid: two plasmid gene-encoded products and a *cis*-acting region are involved in partition. Cell **32:**351-60.

124. **Osborn, A. M., F. M. da Silva Tatley, L. M. Steyn, R. W. Pickup, and J. R. Saunders.** 2000. Mosaic plasmids and mosaic replicons: evolutionary lessons from the analysis of genetic diversity in IncFII-related replicons. Microbiology **146:**2267-2275.

125. **Osborn, M., S. Bron, N. Firth, S. Holsappel, A. Huddleston, R. Kiewiet, W. Meijer, J. Seegers, R. Skurray, P. Terpstra, C. M. Thomas, P. Thorsted, E. Tietze, and S. L. Turner.** 2000. The evolution of bacterial plasmids, p. 301-361. *In* C. M. Thomas (ed.), The Horizontal Gene Pool: Bacterial Plasmids and Gene Spread. Harwood Academic Publishers, Amsterdam.

126. **Ou, J. T., and T. F. Anderson.** 1970. Role of pili in bacterial conjugation. J Bacteriol **102:**648-54.

127. **Pansegrau, W., D. Balzer, V. Kruft, R. Lurz, and E. Lanka.** 1990. *In vitro* assembly of relaxosomes at the transfer origin of plasmid RP4. Proc Natl Acad Sci U S A **87:**6555-9.

128. **Pansegrau, W., and E. Lanka.** 1996. Mechanisms of initiation and termination reactions in conjugative DNA processing. Independence of tight substrate binding and catalytic activity of relaxase (TraI) of IncPalpha plasmid RP4. J Biol Chem **271:**13068-76.

129. **Perez-Mendoza, D., and F. de la Cruz.** 2009. *Escherichia coli* genes affecting recipient ability in plasmid conjugation: are there any? BMC Genomics **10:**71.

130. **Pogliano, J., T. Q. Ho, Z. Zhong, and D. R. Helinski.** 2001. Multicopy plasmids are clustered and localized in *Escherichia coli*. Proc Natl Acad Sci U S A **98:**4486-91.

131. **Ponciano, J. M., L. De Gelder, E. M. Top, and P. Joyce.** 2007. The population biology of bacterial plasmids: a hidden Markov model approach. Genetics **176:**957-68.

132. **Pritchard, R. H., P. T. Barth, and J. Collins.** 1969. Control of DNA synthesis in bacteria, p. 263-297. *In* P. Meadow and S. J. Pirt (ed.), Microbial Growth: Nineteenth Symposium of the Society for General Microbiology. University Press, London.

133. **Prosser, J. I., B. J. Bohannan, T. P. Curtis, R. J. Ellis, M. K. Firestone, R. P. Freckleton, J. L. Green, L. E. Green, K. Killham, J. J. Lennon, A. M. Osborn, M. Solan, C. J. van der Gast, and J. P. Young.** 2007. The role of ecological theory in microbial ecology. Nat Rev Microbiol **5:**384-92.

134. **Rawlings, D. E.** 2005. The evolution of pTF-FC2 and pTC-F14, two related plasmids of the IncQ-family. Plasmid **53:**137-47.

135. **Revilla, C., M. P. Garcillan-Barcia, R. Fernandez-Lopez, N. R. Thomson, M. Sanders, M. Cheung, C. M. Thomas, and F. de la Cruz.** 2008. Different pathways to acquiring resistance genes illustrated by the recent evolution of IncW plasmids. Antimicrob Agents Chemother **52:**1472-80.

136. **Ringgaard, S., J. Lowe, and K. Gerdes.** 2007. Centromere pairing by a plasmid-encoded type I ParB protein. J Biol Chem **282:**28216-25.

137. **Scherzinger, E., M. M. Bagdasarian, P. Scholz, R. Lurz, B. Ruckert, and M. Bagdasarian.** 1984. Replication of the broad host range plasmid RSF1010: requirement for three plasmid-encoded proteins. Proc Natl Acad Sci U S A **81:**654-8.

138. **Scherzinger, E., V. Haring, R. Lurz, and S. Otto.** 1991. Plasmid RSF1010 DNA replication *in vitro* promoted by purified RSF1010 RepA, RepB and RepC proteins. Nucleic Acids Res **19:**1203-11.

139. **Shirasu, K., Z. Koukolikova-Nicola, B. Hohn, and C. I. Kado.** 1994. An inner-membrane-associated virulence protein essential for T-DNA transfer from *Agrobacterium tumefaciens* to plants exhibits ATPase activity and similarities to conjugative transfer genes. Mol Microbiol **11:**581-8.

140. **Silverman, P. M.** 1997. Towards a structural biology of bacterial conjugation. Mol Microbiol **23:**423-9.

141. **Simonsen, L.** 1991. The existence conditions for bacterial plasmids: theory and reality. Microb Ecol **22:**187-205.

142. **Simonsen, L., D. M. Gordon, F. M. Stewart, and B. R. Levin.** 1990. Estimating the rate of plasmid transfer: an end-point method. J Gen Microbiol **136:**2319-25.

143. **Sista, P. R., S. Mukherjee, P. Patel, G. S. Khatri, and D. Bastia.** 1989. A host-encoded DNA-binding protein promotes termination of plasmid replication at a sequence-specific replication terminus. Proc Natl Acad Sci U S A **86:**3026-30.

144. **Slater, F. R., K. D. Bruce, R. J. Ellis, A. K. Lilley, and S. L. Turner.** 2008. Heterogeneous selection in a spatially structured environment affects fitness tradeoffs of plasmid carriage in pseudomonads. Appl Environ Microbiol **74:**3189-97.

145. **Sobecky, P. A., T. J. Mincer, M. C. Chang, and D. R. Helinski.** 1997. Plasmids isolated from marine sediment microbial communities contain replication and incompatibility regions unrelated to those of known plasmid groups. Appl Environ Microbiol **63:**888-95.

146. **Sorensen, S. J., M. Bailey, L. H. Hansen, N. Kroer, and S. Wuertz.** 2005. Studying plasmid horizontal transfer *in situ*: a critical review. Nat Rev Microbiol **3:**700-10.

147. **Soultanas, P., M. S. Dillingham, F. Papadopoulos, S. E. Phillips, C. D. Thomas, and D. B. Wigley.** 1999. Plasmid replication initiator protein RepD increases the processivity of PcrA DNA helicase. Nucleic Acids Res **27:**1421-8.

148. **Stenzel, T. T., T. MacAllister, and D. Bastia.** 1991. Cooperativity at a distance promoted by the combined action of two replication initiator proteins and a DNA bending protein at the replication origin of pSC101. Genes Dev **5:**1453-63.

149. **Stewart, F. M., and B. R. Levin.** 1977. The population biology of bacterial plasmids: a priori conditions for the existence of conjugationally transmitted factors. Genetics **87:**209-228.

150. **Su, L. H., C. Chu, A. Cloeckaert, and C. H. Chiu.** 2008. An epidemic of plasmids? Dissemination of extended-spectrum cephalosporinases among *Salmonella* and other Enterobacteriaceae. FEMS Immunol Med Microbiol **52:**155-68.

151. **Summers, D. K., C. W. Beton, and H. L. Withers.** 1993. Multicopy plasmid instability: the dimer catastrophe hypothesis. Mol Microbiol **8:**1031-8.

152. **Summers, D. K., and D. J. Sherratt.** 1984. Multimerization of high copy number plasmids causes instability: CoIE1 encodes a determinant essential for plasmid monomerization and stability. Cell **36:**1097-103.

153. **Tato, I., S. Zunzunegui, F. de la Cruz, and E. Cabezon.** 2005. TrwB, the coupling protein involved in DNA transport during bacterial conjugation, is a DNA-dependent ATPase. Proc Natl Acad Sci U S A **102:**8156-61.

154. **Taylor, D. E., A. Gibreel, T. D. Lawley, and D. M. Tracz.** 2004. Antibiotic resistance plasmids, p. 473-491. *In* B. E. Funnell and G. J. Phillips (ed.), Plasmid Biology. ASM Press, Washington, D.C.

155. **Thomas, C. M.** 2004. Evolution and population genetics of bacterial plasmids, p. 509-528. *In* B. E. Funnell and G. J. Phillips (ed.), Plasmid Biology. ASM Press, Washington, D.C.

156. **Thomas, C. M.** 2000. Paradigms of plasmid organization. Mol Microbiol **37:**485-91.

157. **Thomas, C. M., and K. M. Nielsen.** 2005. Mechanisms of, and barriers to, horizontal gene transfer between bacteria. Nat Rev Microbiol **3:**711-21.

158. **Top, E. M., Y. Moenne-Loccoz, T. Pembroke, and C. M. Thomas.** 2000. Phenotypic traits conferred by plasmids, p. 249-285. *In* C. M. Thomas (ed.), The Horizontal Gene Pool: Bacterial Plasmids and Gene Spread. Harwood Academic Publishers, Amsterdam.

159. **Top, E. M., P. Van Daele, N. De Saeyer, and L. J. Forney.** 1998. Enhancement of 2,4-dichlorophenoxyacetic acid (2,4-D) degradation in soil by dissemination of catabolic plasmids. Antonie Van Leeuwenhoek **73:**87-94.

160. **Toussaint, A., and C. Merlin.** 2002. Mobile elements as a combination of functional modules. Plasmid **47:**26-35.

161. **Tsutsui, H., A. Fujiyama, T. Murotsu, and K. Matsubara.** 1983. Role of nine repeating sequences of the mini-F genome for expression of F-specific incompatibility phenotype and copy number control. J Bacteriol **155:**337-44.

162. **Turner, P. E., V. S. Cooper, and R. E. Lenski.** 1998. Tradeoff between horizontal and vertical modes of transmission in bacterial plasmids. Evolution **52:**315-329.

163. **Turner, S. L., M. J. Bailey, A. K. Lilley, and C. M. Thomas.** 2002. Ecological and molecular maintenance strategies of mobile genetic elements. FEMS Microbiol Ecol **42:**177-185.

164. **Van Melderen, L., P. Bernard, and M. Couturier.** 1994. Lon-dependent proteolysis of CcdA is the key control for activation of CcdB in plasmid-free segregant bacteria. Mol Microbiol **11:**1151-7.

165. **Watanabe, T.** 1963. Infective heredity of multiple drug resistance in bacteria. Bacteriol Rev **27:**87-115.

166. **Watanabe, T.** 1969. Transferable drug resistance: the nature of the problem, p. 81-101. *In* G. Wolstenholme and M. O'Connor (ed.), Bacterial Episomes and Plasmids. J & A Churchill Ltd., London.

167. **Waters, V. L., and D. G. Guiney.** 1993. Processes at the nick region link conjugation, T-DNA transfer and rolling circle replication. Mol Microbiol **9:**1123-30.

168. **Weigel, L. M., D. B. Clewell, S. R. Gill, N. C. Clark, L. K. McDougal, S. E. Flannagan, J. F. Kolonay, J. Shetty, G. E. Killgore, and F. C. Tenover.** 2003.

Genetic analysis of a high-level vancomycin-resistant isolate of *Staphylococcus aureus*. Science **302:**1569-71.

169. **Wilkins, B. M., and E. Lanka.** 1993. DNA processing and replication during plasmid transfer between Gram-negative bacteria, p. 105-136. *In* D. Clewell (ed.), Bacterial Conjugation. Plenum Press, New York.

170. **Wilkins, B. M., and A. T. Thomas.** 2000. DNA-independent transport of plasmid primase protein between bacteria by the I1 conjugation system. Mol Microbiol **38:**650-7.

171. **Zakian, V. A., B. J. Brewer, and W. L. Fangman.** 1979. Replication of each copy of the yeast 2 micron DNA plasmid occurs during the S phase. Cell **17:**923-34.

172. **Zawadzki, P., M. A. Riley, and F. M. Cohan.** 1996. Homology among nearly all plasmids infecting three *Bacillus* species. J Bacteriol **178:**191-8.

173. **Zechner, E. L., F. de la Cruz, R. Eisenbrandt, A. M. Grahn, G. Koraimann, E. Lanka, G. Muth, W. Pansegrau, C. M. Thomas, B. M. Wilkins, and M. Zatyka.** 2000. Conjugative DNA transfer processes, p. 87-174. *In* C. M. Thomas (ed.), The Horizontal Gene Pool: Bacterial Plasmids and Gene Spread. Harwood Academic Publishers, Amsterdam.

174. **Zhao, H. X., A. G. Wood, F. Widdel, and M. P. Bryant.** 1988. An extremely thermophilic *Methanococcus* from a deep-sea hydrothermal vent and its plasmid. Arch Microbiol **150:**178-183.

175. **Zhou, L., M. Zhou, C. Sun, J. Han, Q. Lu, J. Zhou, and H. Xiang.** 2008. Precise determination, cross-recognition, and functional analysis of the double-

strand origins of the rolling-circle replication plasmids in haloarchaea. J Bacteriol **190:**5710-9.

176. **Ziegelin, G., T. Niedenzu, R. Lurz, W. Saenger, and E. Lanka.** 2003. Hexameric RSF1010 helicase RepA: the structural and functional importance of single amino acid residues. Nucleic Acids Res **31:**5917-29.

177. **Zielenkiewicz, U., and P. Ceglowski.** 2001. Mechanisms of plasmid stable maintenance with special focus on plasmid addiction systems. Acta Biochim Pol **48:**1003-23.

CHAPTER 2

FACILE RECOVERY OF INDIVIDUAL HIGH-MOLECULAR-WEIGHT, LOW-COPY-NUMBER NATURAL PLASMIDS FOR GENOMIC SEQUENCING[1]

---

[1] Williams, L. E., C. Detter, K. Barry, A. Lapidus, and A. O. Summers. 2006. *Applied and Environmental Microbiology*. 72(7): 4899-4906.

ABSTRACT

Sequencing of the large (>50 kb), low-copy-number (<5 per cell) plasmids that mediate horizontal gene transfer has been hindered by the difficulty and expense of isolating DNA from individual plasmids of this class. We report here that a kit method previously devised for purification of bacterial artificial chromosomes (BACs) can be adapted for effective preparation of individual plasmids up to 220 kb from wild gram-negative and gram-positive bacteria. Individual plasmid DNA recovered from less than 10 ml of *Escherichia coli*, *Staphylococcus*, and *Corynebacterium* cultures was of sufficient quantity and quality for construction of high-coverage libraries, as shown by sequencing five native plasmids ranging in size from 30 kb to 94 kb. We also report recommendations for vector screening to optimize plasmid sequence assembly, preliminary annotation of novel plasmid genomes, and insights on mobile genetic element biology derived from these sequences. Adaptation of this BAC method for large plasmid isolation removes one major technical hurdle to expanding our knowledge of the natural plasmid gene pool.

INTRODUCTION

Recent genome sequencing and analysis has revealed extensive horizontal gene transfer among bacterial genomes (4, 14). Remnants of mobile genetic elements (MGEs), such as plasmids and bacteriophages (23), are often found adjacent to horizontally transferred chromosomal regions, indicating that these elements are important mediators of gene transfer between bacterial chromosomes. The MGEs themselves also typically carry genes for virulence factors, antibiotic resistances, and novel metabolic processes that enable bacterial hosts to adapt to new environmental conditions (11, 12).

Despite the recognized importance of these elements, genomic analysis of MGEs has been limited. Whereas the total size of sequenced bacterial genomes is 1.3 Gb, only 61 Mb of plasmid genomes and 30 Mb of phage genomes have been sequenced previously (11). Most MGE sequences have been obtained fortuitously during sequencing of their hosts' genomes, resulting in a bias towards MGEs associated with a limited selection of organisms. Large (>50 kb), conjugative plasmids are especially poorly represented in current sequence databases, constituting only 20% of all plasmid sequences in GenBank at present. Most commercial kits for plasmid DNA preparation are designed for small, high-copy-number plasmids. The traditional methods of high-molecular-weight plasmid isolation, such as cesium chloride density gradient centrifugation (26) and pulsed-field gel electrophoresis, require equipment and expertise that are not widely available. Moreover, these and other techniques, such as Eckhardt in-well lysis (6), are time and labor intensive and thus unsuitable for a high-throughput approach.

In the course of large-scale sequencing projects such as the Human Genome Project, magnetic beads were modified for purification of nucleic acids, including bacterial artificial chromosome (BAC) clones (7, 15, 28). The method used here, termed Solid-Phase Reversible Immobilization (SPRI), employs magnetic beads with carboxylated surfaces to bind plasmid DNA under proprietary buffer conditions. Magnetic immobilization of the beads and bound DNA allows removal of cellular debris and chromosomal DNA. We inferred that SPRI should enable isolation of large, natural plasmids similar in size and copy number to BAC clones. The rapidity, ease and low cost of SPRI BAC purification suggested that it might provide an advantage over traditional costly and laborious methods of high-molecular-weight plasmid isolation. However, there are some important differences between BACs and natural plasmids.

Whereas BACs are maintained individually in laboratory strains of *Escherichia coli*, wild bacteria typically have several plasmids in a wide range of sizes and copy numbers. Ideally, these should each be recovered separately because the abundance of repetitive elements in plasmids can make computer assembly of libraries constructed from pooled supercoiled DNA, such as obtained from CsCl gradients, difficult or impossible. In addition, it is preferable to recover plasmids from their native hosts (when culturable) rather than having to transfer them to a laboratory strain, which might result in changes (see "Plasmid pLEW517" below). Thus, the ideal plasmid isolation method should be applicable to many types of culturable bacteria and not just *E. coli*.

We describe here a protocol for the use of SPRI for rapid, efficient and inexpensive isolation of sequencing-quality DNA of individual large, low-copy-number plasmids from gram-negative and gram-positive bacterial strains. As proof of the efficacy

of this method, we report on the completion and closure of full-length sequences of five natural bacterial plasmids isolated using this protocol: the previously sequenced 94-kb *Shigella flexneri* plasmid NR1; a novel 65-kb *E. coli* plasmid, pLEW517; a novel 52-kb *Staphylococcus* plasmid, pLEW6932; and two novel *Corynebacterium* plasmids, the 35-kb pLEW279a and the 30-kb pLEW279b. We chose these strains and plasmids to answer the following salient questions about plasmid DNA isolation using SPRI technology: (i) does the SPRI isolation method work on gram-negative and gram-positive bacteria? (ii) What is the yield of DNA from a single plasmid band? (iii) What is the best way to remove the plasmid band from the gel? (iv) How much do chromosomal DNA and DNA from other plasmids contaminate the library? We also addressed some of the bioinformatics problems unique to sequencing plasmid DNA: (i) what are the best strategies for vector screening during assembly? (ii) How effective are default BLAST algorithms in identifying DNA and/or protein sequences in plasmids?

MATERIALS AND METHODS

Plasmid DNA isolation.

Bacterial strains and plasmids are listed in Table 2.1. Overnight cultures were inoculated from -70°C stocks and grown without antibiotics. Gram-negative strains were grown overnight in Luria-Bertani (LB) broth at 37°C with shaking at 225 rpm and then diluted 1:50 in LB broth and incubated at 37°C with shaking at 225 rpm for 2.5 to 3 h to obtain cells in logarithmic phase for more effective lysis. Cells were harvested from 1.5 ml culture by centrifugation for 8 min at 11,750 × *g*. Gram-positive strains were grown overnight in brain heart infusion broth at 30°C with shaking at 225 rpm and then diluted in brain heart infusion broth to an optical density at 600 nm of 0.25 to 0.65 to prevent

oversaturation of the beads. Cells were harvested from 1.5 ml culture as described above. Gram-positive strains were not subcultured, because the stationary-phase cells lysed as efficiently as cells in logarithmic phase.

Supercoiled DNA was prepared using a CosMCPrep kit for high- and low-copy-number plasmid purification (Agencourt Biosciences Corp., Beverly, MA) and a microcentrifuge tube protocol supplied by the manufacturer. Pelleted cells were suspended in 100 μl CosMCPrep resuspension buffer. Modifications were made to the cell resuspension step for gram-positive organisms to improve lysis. For *Staphylococcus* preparations, resuspension buffer was supplemented with 200 μg/ml lysostaphin (Sigma, Inc., St. Louis, MO) and 6% polyethylene glycol (Sigma, Inc.), and cell suspensions were incubated at room temperature for 5 min. For *Corynebacterium* preparations, resuspension buffer was supplemented with 5 mg/ml lysozyme (Sigma, Inc.) and 6% polyethylene glycol, and cell suspensions were incubated at 37°C for 30 min.

Following cell resuspension, 100 μl CosMCPrep lysis buffer was added. Preparations were mixed gently by inverting five times and held at room temperature for 5 min. All preparations were handled gently during and after lysis to prevent shearing of the supercoiled DNA. CosMCPrep neutralization buffer (100 μl) was added, and preparations were rotated on an orbital shaker for 10 min at 170 rpm to flocculate cellular debris.

Preparations were then centrifuged for 12 min at 16,000 × *g* at room temperature. The cleared cell lysates (200 μl) were transferred to microcentrifuge tubes and mixed with 139 μl isopropanol and 10 μl CosMCPrep magnetic bead suspension by mixing gently four to six times with a pipette tip. Tubes were placed in a magnetic stand at room

temperature for 10 min to trap the beads with their bound DNA against the sides of the tubes. Unless otherwise noted, all subsequent steps were conducted at room temperature with the tubes in the magnetic stand. The lysate was aspirated without disturbing the beads, and the beads were washed three times with 200 µl 70% ethanol-30% autoclaved MilliQ water. Beads were dried at 37°C in a forced air incubator for 2 min. DNA was then eluted by pipetting 40 µl CosMCPrep resuspension buffer over the beads. Preparations were returned to 37°C in closed tubes for 5 min to separate any beads from the eluate and allow complete elution of bound supercoiled DNA. The bead-free supercoiled DNA eluates (40 µl) were transferred to microcentrifuge tubes and stored at -20°C. Electrophoresis and gel extraction of plasmids were usually done within 24 h, but eluates from gram-negative strains were stored for up to 20 days without appreciable loss of supercoiled DNA.

Electrophoresis and gel extraction of plasmids.

All 40 µl of eluate was loaded onto a 0.3% or 0.5% SeaKem Gold agarose (Cambrex BioScience, Walkersville, MD) gel, with 10 µl of High Range MassRuler DNA ladder (Fermentas, Inc., Hanover, MD) as a size and mass standard. The gel was electrophoresed in 1x TAE (40 mM Tris-acetate, 2 mM $Na_2EDTA\text{-}2H_2O$) at 80 V for 3 to 4 h. Gels were stained with SYBR green I nucleic acid gel stain (Molecular Probes, Inc., Eugene, OR) to increase sensitivity and reduce the DNA damage typically observed with ethidium bromide. Gels were visualized on a Molecular Dynamics FluorImager, and the mass of DNA in each plasmid band was determined by densitometry. To reduce exposure to damaging UV light, gels were placed on a DarkReader Transilluminator (Clare Chemical Research, Denver, CO), and gel slices containing individual plasmid bands

were excised with a razor blade. Supercoiled plasmid DNA was extracted from gel slices using either a GeneClean Turbo glass milk spin kit (Qbiogene, Inc., Carlsbad, CA) or dialysis tubing electroelution.

A GeneClean Turbo kit was used following the manufacturer's instructions, except that 30 μl TE buffer (10 mM Tris, 1 mM EDTA, pH 8) was used for elution. Dialysis tubing electroelution was done as previously described (27, 29), using 1-inch-diameter SpectraPor dialysis tubing that had been boiled in 25 mM EDTA for 10 min, rinsed once with water, and stored at 4°C in 30% ethanol. Gel slices with plasmid DNA were placed in dialysis tubing bags with 250 to 500 μl of 0.1x TAE and electrophoresed at 100 V for 2 h in 0.1x TAE. Polarity was reversed for 2 min, and then the electroeluted DNA was transferred by pipette to a clean 1.5 ml microcentrifuge tube.

DNA yield was quantified by $A_{260/280}$ of a 1:10 dilution in either TE (for GeneClean eluates) or 0.1x TAE (for dialysis tubing electroeluates).

Whole-genome shotgun sequencing and assembly.

For each plasmid, four preparations of DNA, each isolated from 1.5 ml culture using the appropriate CosMCPrep kit method, excised from a gel as a single plasmid band, and eluted using either the GeneClean kit or dialysis tubing electroelution, were pooled and then sheared on a HydroShear (Genomic Solutions, Ann Arbor, MI) at setting 9 into 2- to 3-kb fragments, which were blunt end ligated into pMCL200, a pUC18-based cloning vector (21). Library construction and template preparation were conducted following standard Joint Genome Institute protocols (http://www.jgi.doe.gov/sequencing/protocols/prots_production.html). End sequencing reactions were carried out using a 1/16 dilution of BigDye Terminator v3.1 (Applied

Biosystems, Foster City, CA) and resolved on ABI PRISM 3730 sequencers. Electropherograms were analyzed with PHRED basecalling software (8). The average sequencing read length was 689 ± 30 bp. Sequencing reads were screened using Cross-Match SPS-3.57 (Southwest Parallel Software) to identify and remove vector sequence. For the *Corynebacterium* plasmids pLEW279a and pLEW279b, the entire vector sequence was used for screening. For the *E. coli* plasmids NR1 and pLEW517 and the *Staphylococcus* plasmid pLEW6932, screening with the complete cloning vector sequence introduced artificial gaps into the assemblies, possibly due to similarity between the origins of replication and antibiotic resistance genes on the natural plasmids and those on the cloning vector. Consequently, only sequences identical to the cloning vector spanning the insert site to the sequencing primer annealing site were removed from sequencing reads for these plasmids. After the vector sequences were removed, reads were assembled by PHRAP (www.phrap.org), and gap closure was accomplished by directed PCR of library clones or purified plasmid DNA, resulting in a single, circularized contiguous sequence (contig) for each plasmid.

Analysis of plasmid sequences.

BLAST (1) analysis used the default parameters for the NCBI BLASTN program (+1/-3 match/mismatch penalty; word size 11) and the BLASTX program (word size 3), with the exception that the bacterial translation table was used for BLASTX (www.ncbi.nlm.nih.gov/BLAST/). Pairwise, whole genome alignments of plasmid sequences were generated using the MUMmer software package (19). The "nucmer" script was used to obtain percent identity between the two sequences and to identify the exact locations of disagreements.

GenBank accession numbers.

The plasmid sequences determined in this study have been deposited in the GenBank database under the following accession numbers: for NR1, DQ364638; for wild pLEW517, DQ390454; for transconjugant pLEW517, DQ390455; for pLEW6932, DQ390456; for pLEW279a, DQ390458; and for pLEW279b, DQ390457.

RESULTS

Recovery of supercoiled plasmid DNA.

The standard CosMCPrep SPRI protocol was effective for plasmid isolation from gram-negative proteobacteria *E. coli*, *Rhizobium* (Figure 2.1), *Salmonella* and *Pseudomonas* (data not shown), and modifications (see Materials and Methods) enabled plasmid isolation from low-G+C (*Staphylococcus*) and high-G+C (*Corynebacterium*) gram-positive bacteria (Figure 2.1). Plasmids from 4 to 220 kb in size were readily visible on agarose gels, and supercoiled DNA was recovered equally well from small, high-copy-number plasmids such as pBR322 (3) and from large, low-copy-number plasmids such as NR1 (33) (Figure 2.1). For the well-characterized, laboratory standard plasmids pBR322, R388, RP4 and NR1, the average yield of DNA of each plasmid recovered from 1.5 ml *E. coli* cultures and detected as a single band on a gel was 0.865 μg (range, 0.5 to 1.3 μg). Host-plasmid combinations were chosen to address specific critical questions concerning the ability of this method to optimize the sequencing of large, low-copy-number natural plasmids directly from their original host bacteria.

Plasmid NR1.

To determine whether the recovered plasmid DNA was suitable in amount and quality for library construction, we sequenced plasmid NR1, a 94-kb IncFII plasmid

presumably identical to the previously sequenced plasmid R100 (NC_002134). We found that using a GeneClean kit for gel extraction of NR1 DNA resulted in extensive random shearing (Figure 2.2). However, pooling four preparations yielded 1.1 μg DNA, which was sufficient for construction of a library of 768 clones without amplification. Sequencing and assembly of NR1 yielded one contig of 94,289 bp and another of 2,198 bp (Table 2.2). MUMmer (19) analysis showed that the larger contig had only 15 single-nucleotide disagreements with the 94,281-bp R100 reference sequence. The sequence quality for all reads was sufficiently high to confirm that these were real polymorphisms in the plasmid sequences.

The smaller contig from the NR1 project (Table 2.2) was 100% identical to the pMCL200 cloning vector and was likely a result of the modified vector screening used during assembly. The narrowed vector screening was required to prevent introduction of artificial gaps in the NR1 assembly in regions of identity between the natural plasmid and the cloning vector (e.g., *cat*, the chloramphenicol acetyltransferase gene). However, allowing some vector sequences to remain resulted in the assembly of the small but readily identifiable vector contig.

Plasmid pLEW517.

To determine whether a single plasmid in a multiplasmid host could be reliably sequenced from supercoiled DNA recovered from a gel, we sequenced plasmid pLEW517, previously shown to confer resistance to ampicillin, streptomycin, sulfonamides, and mercury (32), as derived both from its native host, the multiplasmid primate intestinal *E. coli* strain 517-2H1 (Figure 2.1) (30), and from an otherwise plasmid-free laboratory *E. coli* strain 690FNR into which it had been transferred by

conjugation. Dialysis tubing electroelution, which preserved the supercoiled conformation of extracted DNA better than glass milk extraction (Figure 2.2), was used to recover pLEW517 DNA and in all subsequent work. Pooling four preparations yielded 3.5 μg for wild pLEW517 and 4.0 μg for transconjugant pLEW517, amounts which were sufficient for library construction.

Sequencing and assembly of wild pLEW517 yielded a single major contig of 63,946 bp (Table 2.2), and transconjugant pLEW517 yielded a single major contig of 65,288 bp. These contigs were 100% identical except for a 1,342-bp segment found on transconjugant pLEW517 but not wild pLEW517 (discussed below). Overall, 98% of the sequence of pLEW517 returned significant BLASTN hits (Table 2.3). BLASTN analysis indicated that pLEW517 is a variant of plasmid R46 (NC_003292), a 50-kb IncN plasmid previously observed in *Salmonella enterica* serovar Typhimurium. Significant similarity (e-value, 0.0) was detected to regions encoding replication, maintenance and transfer functions on plasmid R46. However, an 18-kb segment of R46 that included the In*1* integron was absent from pLEW517. pLEW517 also had sequences with very significant similarity to those of transposon Tn*21*, which contains a class 1 integron (In*2*) and a mercury resistance (*mer*) operon (20), and of transposon Tn*3*, which encodes a ß-lactamase (18).

Alignments of the wild and transconjugant pLEW517 sequences to the R46 sequence indicated that the 1,342-bp difference between the two sequences lies in a repeat region of conserved upstream (CUP) elements, which may regulate expression of adjacent genes during conjugative transfer (5). Repeat regions frequently present difficulties during assembly, and this repeat region may have been overcollapsed during

assembly of the wild pLEW517 sequence, thus causing the observed difference in length. However, the high sequencing quality and coverage suggests that this may be a real difference between the wild and transconjugant plasmids arising during conjugation to the new host. In either case, neither chromosomal DNA nor the two other plasmids of very different molecular weights and copy numbers in the same strain (Figure 2.1) interfered with the assembly of the target plasmid extracted from the gel.

Further examination of the pLEW517 region showing similarity to Tn*21* revealed that although the hallmarks of Tn*21*, which include the transposition genes (*tnpAR*) and the *mer* operon, were present on the pLEW517 transposon, there were dramatic differences in the content of its version of the integron (Figure 2.3). Whereas the class 1 integron In*2* of Tn*21* has a single cassette encoding an aminoglycoside adenyltransferase (*aadA1*), the class 1 integron of pLEW517 has three cassettes carrying *dfrA12* (previously called *dhfrXII*), a dihydrofolate reductase; an open reading frame (ORF) of unknown function; and *aadA2*, an aminoglycoside adenyltransferase. These cassettes were identified by BLASTN analysis based on similarity to an integron found on the 89.5-kb *Citrobacter freundii* plasmid pCTX-M3 (NC_004464), which contains the three cassettes in the same order. This cassette arrangement was previously reported on a Tn*21*-like element carried on a 70-kb plasmid from a pathogenic *E. coli* strain (16). Note that the insertion of the pLEW517 integron into the ancestral *mer* transposon "backbone" is in exactly the same position as in the prototypical Tn*21* of NR1 and R100.

In addition to these differences at the *attI* insertion point, there was an insertion into the pLEW517 integron that truncated both orf5 in the 3′ conserved segment and the transposition (*tni*) module compared to the corresponding sequences in the prototypical

Tn*21*. BLAST identified similarity to a region on the 48-kb plasmid pRSB101 (NC_006385) from an uncultured host, including a homolog of *chrA* resembling a chromate ion transporter, an ORF of unknown function, an operon carrying a macrolide phosphotransferase (*mphA*) and its repressor (*mphR*), and two insertion sequence (IS) elements. The macrolide resistance operon was previously observed inserted into the integron of a Tn*21*-like transposon in a strain of *Aeromonas hydrophila* isolated from swine (25). This integron also contained the *dfrA12*, orfF, and *aadA2* cassettes identified on the pLEW517 transposon; however, it lacked the *chrA* homolog, the IS elements, and the ORF of unknown function. To our knowledge, this is the first observation of a Tn*21*-like transposon in an IncN plasmid, demonstrating that Tn*21* is not limited to the R100-like IncFII plasmids with which it is commonly associated. The differences in accessory element content between the prototypical Tn*21*, the other Tn*21*-like elements as described above, and the pLEW517 transposon support the idea that Tn*21*-like elements and their host plasmids serve as hotspots for recombination involving integron cassettes, insertion sequences, and other transposons.

Plasmid pLEW6932.

We then assessed the effectiveness of the SPRI method on low-G+C gram-positive bacteria using the multiplasmid *Staphylococcus* strain 693-2 obtained from poultry litter (Figure 2.1) (22). This strain had nine visible plasmid bands, the largest of which was chosen for sequencing. Four pooled preparations yielded 2.9 μg DNA for library construction. Sequencing and assembly produced a major contig of 51,514 bp, similar to the 51 kb estimated for pLEW6932 from agarose gels. BLASTN identified two small regions of significant similarity (e-value, 0.0): one similar to the arsenic resistance

operon of *Staphylococcus saprophyticus* plasmid pSSP1 (NC_007351) and another to ß-lactamase genes from other *Staphylococcus* chromosomes and plasmids (e.g., BX571857) (Table 2.3). The arsenic resistance (*ars*) operon on pLEW6932 may be selected due to the common use of organic arsenic coccidiostats such as roxarsone for growth promotion; roxarsone naturally degrades to inorganic arsenate and arsenite (13), to which this locus confers resistance.

To look for genes conserved only at the amino acid sequence level, we used BLASTX analysis, in which the nucleotide sequence is translated into all six possible reading frames and compared to the protein sequence database. BLASTX identified a region of significant similarity (e-value, $\leq e^{-90}$) to replication initiation proteins of various *Staphylococcus* plasmids (e.g., NP_932180, CAA63141). Other regions of similarity included hits to glycine betaine transporters (e.g., ZP_00233406) and cation transport ATPases (e.g., ZP_00063375) found on the chromosomes of a variety of bacterial species. Of the 125 hits identified using default parameters of BLASTX, most were repetitive hits on these and a few other loci, leaving approximately 60% of the pLEW6932 sequence with no known protein or nucleic acid similarities as reported with the default parameters of these BLAST programs.

Plasmids pLEW279a and pLEW279b.

Last, *Corynebacterium* strain L2-79-05 provided the opportunity to assess the effectiveness of SPRI on high-G+C gram-positive bacteria and also to investigate the occurrence of plasmid band cross-contamination in multiplasmid strains. In contrast to *E. coli* 517-2H1, the plasmid profile of this poultry litter (22) strain shows two plasmids of very similar sizes and copy numbers (Figure 2.1). Isolation of the larger plasmid

pLEW279a by electroelution from the gel yielded 3.1 μg DNA for library construction. Sequencing and assembly produced two major closed circular contigs (Table 2.2). The 34,606-bp contig corresponded in size to the 35-kb band extracted from agarose gels, and the second contig of 29,854 bp was similar in size to the 30-kb plasmid (named pLEW279b) also observed in the L2-79-05 plasmid profile. This suggests that for plasmids of very similar sizes and copy numbers, gel slices of apparently single-plasmid bands may contain some of the other plasmid's DNA; however, current methods of sequence assembly are capable of resolving the two plasmids into separate contigs. Although the fortuitously sequenced pLEW279b has fewer sequencing reads than pLEW279a (Table 2.2), it still has very good depth.

For pLEW279a, BLASTN identified two regions with significant similarity (e-value, 0.0) to other plasmids. One large region resembles part of the 28-kb *Corynebacterium glutamicum* plasmid pTET3 (NC_003227) (31), including a tetracycline resistance determinant and a class 1 integron with a truncated integrase and an aminoglycoside adenyltransferase cassette, *aadA9* (Table 2.3). An adjacent, smaller region resembles the 9-kb *Arcanobacterium pyogenes* plasmid pAP2 (NC_005206) (17), including a macrolide resistance determinant. In addition, BLASTX identified a region of pLEW279a with approximately 60% amino acid identity to RepA replication initiation proteins of *C. glutamicum* plasmids pTET3 and pCG4 (29 kb; NC_004945) and a region with approximately 50% amino acid identity to TraA transfer proteins of *Corynebacterium* plasmids pGA2 (19 kb; NC_004535) and pNG2 (15 kb; NC_005001). Of the 148 BLASTX hits, all are repeated hits on these and a few other loci, leaving 40%

of the plasmid sequence with no known protein or nucleic acid similarities as reported by the default parameters of the BLAST programs.

For pLEW279b, BLASTN identified a large region of significant similarity (e-value, 0.0) to *C. glutamicum* ATCC 13032 chromosomal DNA (BA000036) containing genes involved in copper metabolism and genes encoding a two-component-type response regulator and kinase. BLASTN also identified two small regions of similarity to *Corynebacterium* plasmids: one region of approximately 2 kb similar to a putative ABC transporter found on the 12-kb *C. jeikeium* plasmid pA505 (NC_004773) and another region of approximately 0.5 kb similar to part of a transposase found on *C. glutamicum* plasmids pTET3, pAG1 (20 kb; NC_001415), and pGA2. Last, BLASTX identified a region with approximately 69% amino acid identity to the RepA replication protein of *C. striatum* plasmid pTP10 (51.5 kb; NC_004939) and a region with approximately 31% amino acid identity to the TraA-like transfer protein of *Rhodococcus equi* virulence plasmid p103 (80.5 kb; NC_002576). The 103 BLASTX hits were repeated instances of these and a few other sequences. Approximately 40% of this plasmid sequence had no known protein or nucleic acid similarities detectable by the default parameters of these two BLAST programs.

DISCUSSION

The SPRI method described here can be combined with lysis methods for other bacteria that do not perturb the attachment of plasmid DNA to the carboxylated microspheres. This approach, in addition to being facile and inexpensive (ca. $0.50 per preparation), enables direct recovery of plasmid DNA from wild bacterial strains without the need to transfer them to a laboratory strain of *E. coli*. The yield of plasmid DNA after

isolation, electrophoresis, and elution is sufficient in that pooling a small number of preparations provides enough sequencing-quality DNA to construct a library and obtain sequence at excellent coverage (> 9x). We did not detect any chromosomal contamination of the plasmid sequences, and contamination of plasmid DNA preparations with other plasmid DNA was only detected when the two plasmids had similar sizes and copy numbers. Even in this instance, current methods of sequence assembly were able to differentiate the two sequences.

We found that vector screening during assembly of plasmid sequences is complicated by the presence of similar genes on both the natural plasmid and the cloning vector, resulting in the introduction of artificial gaps in the assembly. Rather than using the entire cloning vector sequence for screening, we screened using only the cloning vector sequence from the insert site to the primer annealing site. This resolved the gaps in the assemblies; however, in one case it resulted in the generation of a small contig identical to the cloning vector's origin of replication and antibiotic resistance marker. Since this contig was easily identified as cloning vector sequence, the strategy of narrowed vector screening was considered suitable for plasmid sequence assembly. As a general rule in plasmid sequencing work, any antibiotic resistance phenotype information available for the plasmid of interest could guide selection of a cloning vector that encodes a different antibiotic resistance than those found on the natural plasmid.

Our preliminary analysis of the four novel plasmid sequences illustrates the variety of genes that can be carried by plasmids as well as the limitations in plasmid genome information currently available. BLAST analysis successfully identified plasmid replication, transfer and maintenance genes as well as mobile and chromosomal genes

with diverse functions in all novel plasmid sequences. The gram-negative plasmid pLEW517, which had high-scoring BLAST hits for all but a small fraction of the genome, was characterized as a variant of plasmid R46. pLEW517 encoded the same replication and transfer genes as R46 but lacked a characteristic R46 integron and possessed two transposons not found on R46. This illustrates how the insertion and removal of accessory elements such as transposons, integrons, and insertion sequences on a plasmid backbone leads to extensive variation in plasmid genomes, similar to the chromosomal variation observed among strains of a bacterial species. Further sequencing of natural plasmids will contribute to our understanding of the degree to which plasmid backbones vary in accessory elements. Sequencing more plasmids will also uncover variation in the accessory elements themselves, such as the novel Tn*21*-like transposon found on pLEW517.

In contrast to pLEW517, approximately 60% of the pLEW6932 sequence and 40% of the pLEW279a and pLEW279b sequences had no apparent similarity to known sequences in the nucleotide and protein sequence databases. BLAST hits obtained for these genomes were from a variety of gram-positive bacterial chromosomes and plasmids, suggesting that these plasmid genomes are a mosaic of genes and accessory elements acquired from related sources. The underrepresentation of environmental gram-positive plasmids in current sequence databases may account for the observed lack of similarity. Thorough manual annotation of these plasmids is beyond the expertise of a single laboratory; however, the initial annotation accompanying the sequences as deposited in GenBank can provide a starting point for deeper analysis by those with relevant expertise. Much more extensive sequencing of plasmid genomes as well as

inclusion of plasmid and other mobile element terms in the genome and sequence ontologies is essential to addressing fundamental questions of prokaryotic evolution and to dealing with critical problems such as the spread of antibiotic resistance. The SPRI plasmid isolation method coupled with simple electrophoretic elution of individual supercoiled plasmids provides a facile and inexpensive approach to obtain DNA of sufficient amount and quality to generate libraries with excellent coverage that result in readily finishable genomic sequences.

ACKNOWLEDGMENTS

We thank NSF-REU (grant DBI0453353) summer student David Miller for assistance in applying the SPRI method to gram-positive bacteria, colleagues Tamar Barkay, Margie Lee, John Maurer, Mark Schell and Patricia Sobecky for advice on techniques and critical reading of the manuscript, and three anonymous reviewers for their help in improving the manuscript. We also thank Agencourt, Inc. for providing us with a small CosMCPrep starter kit.

This work was supported by the Office of Science (BER), U.S. Department of Energy (grant DE-FG02-04ER63770 to A.O.S.).

TABLE 2.1. Bacterial strains and plasmids.

Species and host strain (reference)[a]	Plasmid (size in kb)	Plasmid source or reference
Escherichia coli		
DU1040 (10)	NR1[c] (94)	33
J53[b]	RP4 (60)	24
CB454	R388 (33)	2
W1485[b]	F (100)	9
	pBR322 (4)	3
517-2H1 (30)	pLEW517 (65)	32
690FNR (32)	pLEW517 (65)	32
Corynebacterium sp.		
L2-79-05 (22)	pLEW279a (35)	This study
	pLEW279b (30)	This study
Staphylococcus sp.		
693-2 (22)	pLEW6932 (52)	This study

[a] The *E. coli* strains are gram negative, the *Corynebacterium* sp. strain is high-G+C gram positive, and the *Staphylococcus* sp. strain is low-G+C gram positive.

[b] Strains are deposited in the *E. coli* Genetic Stock Center (http://cgsc.biology.yale.edu).

[c] Plasmid NR1 is also referred to as R100.

TABLE 2.2. Sequencing coverage resulting in single, circularized contigs.

Plasmid and type	Contig	Length (bp)	No. of sequencing reads	Depth (×)
De novo NR1	1	94,289	1,141	11
	2	2,198	306	108
pLEW517				
Wild	1	63,946	1,243	13
Transconjugant	1	65,288	1,271	15
pLEW6932	1	51,514	2,648	44
pLEW279a	1	34,606	3,759	90
pLEW279b	1	29,854	303	9

TABLE 2.3. Highlights of the chimeric character of novel plasmid genomes[a]

Plasmid and BLAST program	Selected BLAST hit[b] (% identity)	% of plasmid genome similar to hit	Gene(s) and/or region(s) of interest within hit
pLEW517			
BLASTN	*S. enterica* serovar Typhimurium plasmid R46 (99% nt)	49.1	*oriT*, Tra proteins, *repA*, iterons, *korAB*
	Tn*21* (99% nt)	18.3	*mer* operon, class 1 integron, *tnpA*
	Tn*3* (97% nt)	7.7	β-lactamase
pLEW6932			
BLASTN	*Staphylococcus saprophyticus* pSSP1 (97% nt)	19.6	Arsenic resistance operon
	S. aureus MSSA476 (84% nt)	6.7	Metallo-β-lactamase superfamily protein
BLASTX	*Staphylococcus* plasmids (50-65% aa)	1.8	RepA
	Multiple species (30-40% aa)	1.2	Cation transport ATPase
	Multiple species (50-70% aa)	1.0	Glycine betaine transporter
pLEW279a			
BLASTN	*C. glutamicum* pTET3 (99% nt)	30.6	IS*6100*, *tetA*, *tetR*, class 1 integron w/ truncated *intI1*
	Arcanobacterium pyogenes pAP2 (99% nt)	8.1	*ermX*
BLASTX	*C. glutamicum* pTET3 and pCG4 (60% aa)	4.1	RepA
	C. diphtheriae pNG2, *C. glutamicum* pGA2 (50% aa)	9.6	TraA
pLEW279b			
BLASTN	*C. glutamicum* ATCC 13032 chromosomal DNA (97% nt)	22	Two-component system response regulator and kinase, Cu^{2+} transport ATPase, Cu^{2+} oxidase
	C. jeikeium pA505 (99% nt)	6.6	Putative ABC transporter

BLASTX	*C. striatum* pTP10 (69% aa)	3.1	RepA
	Rhodococcus equi p103 (31% aa)	9.9	TraA-like transfer protein

[a] Data do not represent an exhaustive list of BLAST hits for these plasmids.

[b] Where appropriate, multiple hits with similar BLAST scores are listed.

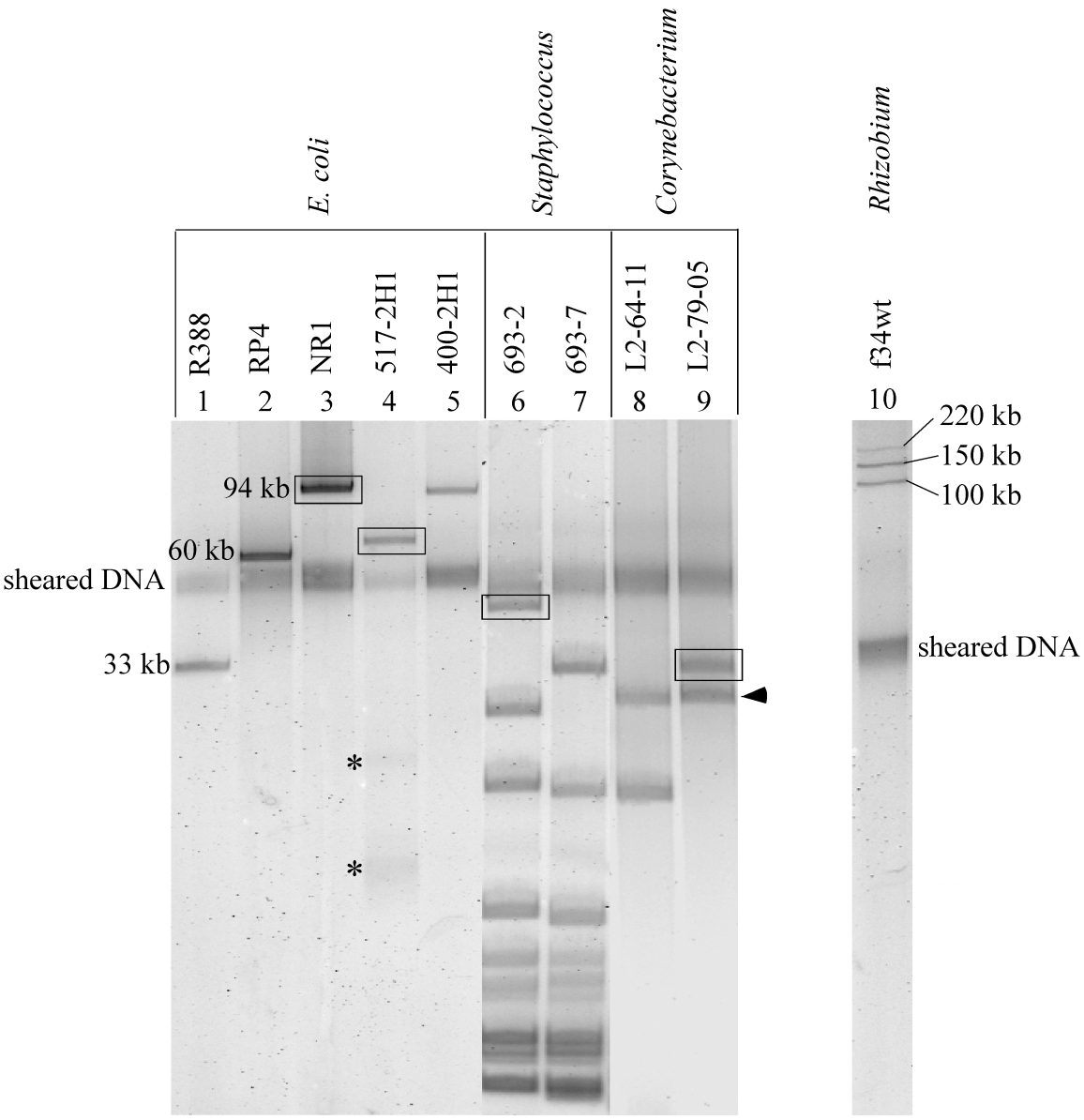


FIGURE 2.1. Plasmid DNA prepared using magnetic bead-based SPRI. Lanes 1 to 9, 0.5% SeaKem Gold agarose-1x TAE. Lane 1 = CB454(R388). Lane 2 = J53(RP4). Lane 3 = DU1040(NR1). Lanes 4 and 5 = wild *E. coli* strains (30). Lanes 6 and 7 = wild *Staphylococcus* strains (22). Lanes 8 and 9 = wild *Corynebacterium* strains (22). Lane 10 = *Rhizobium meliloti*, 0.75% Sigma agarose-1x TBE, run at 80 V for 8 h. The sheared DNA band contains any linearized DNA larger than 20 kb. Boxes indicate plasmids whose sequencing is reported here. Asterisks indicate additional faint plasmid bands in *E.*

coli strain 517-2H1, and an arrowhead indicates the plasmid whose sequence was obtained adventitiously along with that of the boxed plasmid of *Corynebacterium* strain L2-79-05.

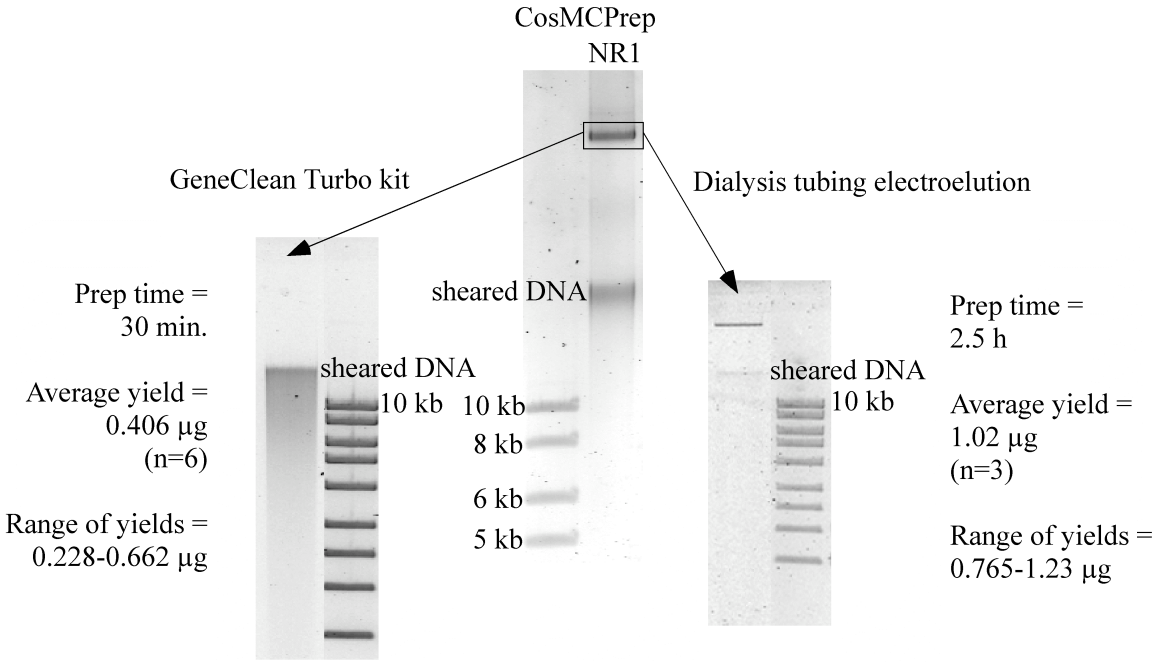


FIGURE 2.2. Extraction of NR1 plasmid DNA from agarose gels. A 10 μl aliquot of the extracted DNA was run on a 0.75% Sigma agarose-1x TBE gel.

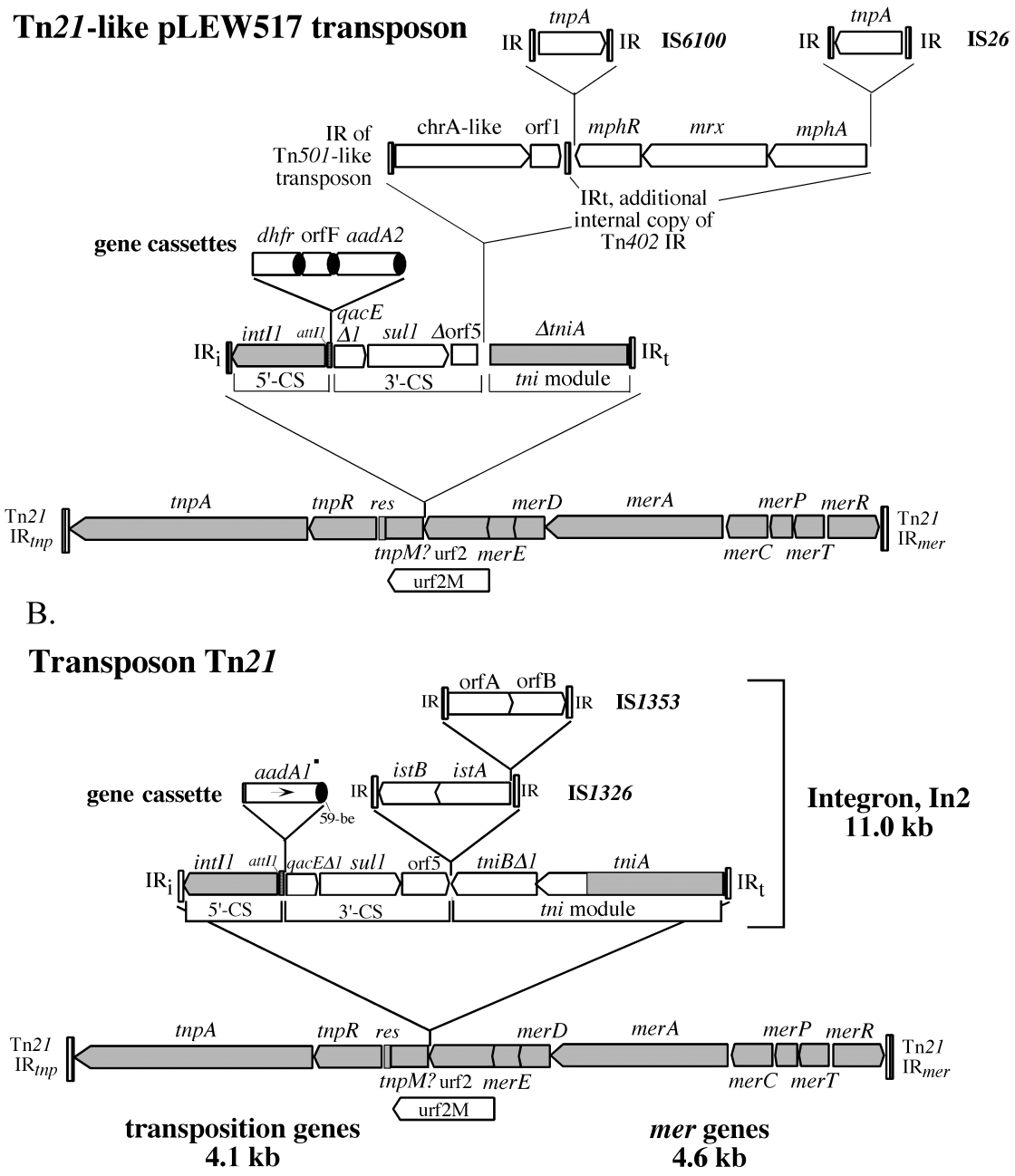


FIGURE 2.3. (A) The Tn*21*-like transposon found on *E. coli* plasmid pLEW517. Shading indicates regions where the pLEW517 transposon shares highest similarity with Tn*21*. Inverted repeats (IR) of transposons and insertion sequences are indicated by vertical bars. The transposition genes include the transposase (*tnpA*), the resolvase (*tnpR*),

and the putative transposition regulator (*tnpM*). The transposition resolution site is encoded by *res*. The 5' conserved segment (CS) of the integron consists of the integrase (*intI1*) and the *aatI1* recombination site. Three gene cassettes carry dihydrofolate reductase (*dhfr*), an ORF of unknown function (orfF), and aminoglycoside adenyltransferase (*aadA2*). The 3' conserved segment (CS) consists of genes encoding resistance to quaternary ammonium compounds (*qacE*Δ1), sulfonamide resistance (*sul1*), and an ORF of unknown function (orf5), which is truncated compared to orf5 in Tn*21*. An additional element has inserted into the integron. This element contains a *chrA*-like gene resembling a chromate ion transporter, an ORF of unknown function (orf1), genes encoding a macrolide phosphotransferase (*mphA*) and its repressor (*mphR*), a gene encoding a hydrophobic protein (*mrx*), and two insertion sequence elements (IS*6100* and IS*26*). The integron transposition (*tni*) module of this transposon has also been truncated compared to Tn*21*, with only a truncated *tniA* remaining. The mercury resistance operon includes regulatory genes (*merR* and *merD*) and structural genes (*merT*, *merP*, *merC* and *merA*). (B) Tn*21* as depicted in reference 20. Figure 2.3 was adapted from reference 20.

REFERENCES

1. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. **215**:403–410.

2. **Avila, P., and F. de la Cruz.** 1988. Physical and genetic map of the IncW plasmid R388. Plasmid **20**:155–157.

3. **Bolivar, F., R. L. Rodriguez, P. J. Greene, M. C. Betlach, H. L. Heyneker, and H. W. Boyer.** 1977. Construction and characterization of new cloning vehicles. II. A multipurpose cloning system. Gene **2**:95–113.

4. **Boucher, Y., C. J. Douady, R. T. Papke, D. A. Walsh, M. E. Boudreau, C. L. Nesbo, R. J. Case, and W. F. Doolittle.** 2003. Lateral gene transfer and the origins of prokaryotic groups. Annu. Rev. Genet. **37**:283–328.

5. **Delver, E. P., and A. A. Belogurov.** 1997. Organization of the leading region of IncN plasmid pKM101 (R46): a regulation controlled by CUP sequence elements. J. Mol. Biol. **271**:13–30.

6. **Eckhardt, T.** 1978. A rapid method for the identification of plasmid desoxyribonucleic acid in bacteria. Plasmid **1**:584–588.

7. **Elkin, C. J., P. M. Richardson, H. M. Fourcade, N. M. Hammon, M. J. Pollard, P. F. Predki, T. Glavina, and T. L. Hawkins.** 2001. High-throughput plasmid purification for capillary sequencing. Genome Res. **11**:1269–1274.

8. **Ewing, B., L. Hillier, M. C. Wendl, and P. Green.** 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res. **8**:175–185.

9. **Firth, N., K. Ippen-Ihler, and R. A. Skurray.** 1996. Structure and function of

the F factor and mechanism of conjugation, p. 2377–2401. In F. C. Neidhardt, R. Curtiss III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella*: cellular and molecular biology, 2nd ed., vol. 2. ASM Press, Washington, D.C.

10. **Foster, T. J., H. Nakahara, A. A. Weiss, and S. Silver.** 1979. Transposon A-generated mutations in the mercuric resistance genes of plasmid R100-1. J. Bacteriol. **140**:167–181.

11. **Frost, L. S., R. Leplae, A. O. Summers, and A. Toussaint.** 2005. Mobile genetic elements: the agents of open source evolution. Nat. Rev. Microbiol. **3**:722–732.

12. **Funnell, B. E., and G. J. Phillips** (ed.). 2004. Plasmid biology. ASM Press, Washington, D.C.

13. **Garbarino, J. R., A. J. Bednar, D. W. Rutherford, R. S. Beyer, and R. L. Wershaw.** 2003. Environmental fate of roxarsone in poultry litter. I. Degradation of roxarsone during composting. Environ. Sci. Technol. **37**:1509–1514.

14. **Gogarten, J. P., and J. P. Townsend.** 2005. Horizontal gene transfer, genome innovation and evolution. Nat. Rev. Microbiol. **3**:679–687.

15. **Hawkins, T. L., T. O'Connor-Morin, A. Roy, and C. Santillan.** 1994. DNA purification and isolation using a solid-phase. Nucleic Acids Res. **22**:4543–4544.

16. **Heikkila, E., M. Skurnik, L. Sundstrom, and P. Huovinen.** 1993. A novel dihydrofolate reductase cassette inserted in an integron borne on a Tn*21*-like

element. Antimicrob. Agents Chemother**. 37**:1297–1304.

17. **Jost, B. H., A. C. Field, H. T. Trinh, J. G. Songer, and S. J. Billington.** 2003. Tylosin resistance in *Arcanobacterium pyogenes* is encoded by an ermX determinant. Antimicrob. Agents Chemother. **47**:3519–3524.

18. **Kostriken, R., C. Morita, and F. Heffron.** 1981. Transposon Tn*3* encodes a site-specific recombination system: identification of essential sequences, genes, and actual site of recombination. Proc. Natl. Acad. Sci. USA **78**:4041–4045.

19. **Kurtz, S., A. Phillippy, A. L. Delcher, M. Smoot, M. Shumway, C. Antonescu, and S. L. Salzberg.** 2004. Versatile and open software for comparing large genomes. Genome Biol. **5**:R12.

20. **Liebert, C. A., R. M. Hall, and A. O. Summers.** 1999. Transposon Tn*21*, flagship of the floating genome. Microbiol. Mol. Biol. Rev. **63**:507–522.

21. **Nakano, Y., Y. Yoshida, Y. Yamashita, and T. Koga.** 1995. Construction of a series of pACYC-derived plasmid vectors. Gene **162**:157–158.

22. **Nandi, S., J. J. Maurer, C. Hofacre, and A. O. Summers.** 2004. Gram-positive bacteria are a major reservoir of class 1 antibiotic resistance integrons in poultry litter. Proc. Natl. Acad. Sci. USA **101**:7118–7122.

23. **Ochman, H., J. G. Lawrence, and E. A. Groisman.** 2000. Lateral gene transfer and the nature of bacterial innovation. Nature **405**:299–304.

24. **Pansegrau, W., E. Lanka, P. T. Barth, D. H. Figurski, D. G. Guiney, D. Haas, D. R. Helinski, H. Schwab, V. A. Stanisich, and C. M. Thomas.** 1994. Complete nucleotide sequence of Birmingham IncP alpha plasmids. Compilation

and comparative analysis. J. Mol. Biol. **239**:623–663.

25. **Poole, T. L., T. R. Callaway, K. M. Bischoff, C. E. Warnes, and D. J. Nisbet.** 2006. Macrolide inactivation gene cluster *mphA-mrx-mphR* adjacent to a class 1 integron in *Aeromonas hydrophila* isolated from a diarrhoeic pig in Oklahoma. J. Antimicrob. Chemother. **57**:31–38.

26. **Radloff, R., W. Bauer, and J. Vinograd.** 1967. A dye-buoyant-density method for the detection and isolation of closed circular duplex DNA: the closed circular DNA in HeLa cells. Proc. Natl. Acad. Sci. USA **57**:1514–1521.

27. **Sambrook, J., and D. W. Russell**. 2001. Molecular cloning: a laboratory manual, 3rd ed., vol. 1. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

28. **Skowronski, E. W., N. Armstrong, G. Andersen, M. Macht, and P. M. McCready.** 2000. Magnetic, microplate-format plasmid isolation protocol for high-yield, sequencing-grade DNA. BioTechniques **29**:786–792.

29. **Strong, S. J., Y. Ohta, G. W. Litman, and C. T. Amemiya.** 1997. Marked improvement of PAC and BAC cloning is achieved using electroelution of pulsed-field gel-separated partial digests of genomic DNA. Nucleic Acids Res. **25**:3959–3961.

30. **Summers, A. O., J. Wireman, M. J. Vimy, F. L. Lorscheider, B. Marshall, S. B. Levy, S. Bennett, and L. Billard.** 1993. Mercury released from dental "silver" fillings provokes an increase in mercury- and antibiotic-resistant bacteria in oral and intestinal floras of primates. Antimicrob. Agents Chemother. **37**:825–834.

31. **Tauch, A., S. Gotker, A. Puhler, J. Kalinowski, and G. Thierbach.** 2002. The 27.8-kb R-plasmid pTET3 from *Corynebacterium glutamicum* encodes the aminoglycoside adenyltransferase gene cassette *aadA9* and the regulated tetracycline efflux system Tet 33 flanked by active copies of the widespread insertion sequence IS6100. Plasmid **48**:117–129.

32. **Wireman, J., C. A. Liebert, T. Smith, and A. O. Summers.** 1997. Association of mercury resistance with antibiotic resistance in the gram-negative fecal bacteria of primates. Appl. Environ. Microbiol. **63**:4494–4503.

33. **Womble, D. D., and R. H. Rownd.** 1988. Genetic and physical map of plasmid NR1: comparison with other IncFII antibiotic resistance plasmids. Microbiol. Rev. **52**:433–451.

CHAPTER 3

LARGE PLASMIDS IN *ESCHERICHIA COLI* AND *SALMONELLA* REFERENCE COLLECTIONS ARE HIGHLY VARIABLE DESPITE COMMON REPLICON FAMILIES[1]

[1] Williams, L. E., V. Hilliard, and A. O. Summers. To be submitted.

ABSTRACT

Plasmids play key roles in the evolution and adaptation of their host bacteria, yet we lack a comprehensive picture of their own natural variation. We assessed the diversity and distribution of plasmids in the ECOR, SARA, SARB and SARC reference collections of *Escherichia coli* and *Salmonella*. Plasmids, especially large (≥30 kb) plasmids, are abundant in these collections. PCR-based replicon typing showed that the IncF and IncI1 replicon families are prevalent in both enterobacteria, but the presence of many untypable plasmids indicated that plasmid backbone diversity may be extensive, even in the well-characterized Enterobacteriaceae. The large plasmids of ECOR and SAR (n=164) were highly variable by RFLP analysis. Only one RFLP pattern, corresponding to the pSLT virulence plasmid of serovar Typhimurium, was found in more than four strains. The conservation of the pSLT RFLP pattern contrasted with the variability of the other ECOR and SAR plasmids, suggesting that the virulence plasmid is experiencing different evolutionary constraints than most enterobacterial plasmids. These results provide important information on natural plasmid variation and make a valuable contribution to population-based studies using these reference collections.

INTRODUCTION

Plasmids are an important pool of potentially adaptive and transferable genetic information in bacterial populations (25). Plasmid-borne genes can have dramatic effects on host bacteria, sometimes significantly altering bacterial physiology or lifestyle (55). Transfer operons encoding conjugative machinery on some plasmids facilitate their spread among bacterial strains, species and even genera (51). Although the important roles of plasmids in bacterial adaptation and horizontal gene transfer are recognized, questions remain about the ecology and evolution of these mobile genetic elements. How diverse are the plasmids of different bacterial genera? How broadly are particular plasmids distributed in nature? Previous studies of plasmid diversity and distribution have been limited to plasmids encoding traits of wide interest, such as antibiotic resistance (21, 49). A comprehensive view of naturally occurring plasmid variation is required to understand the roles of plasmids in bacterial evolution and adaptation. To this end, we isolated and characterized plasmids from the ECOR and SAR reference collections of *Escherichia coli* and *Salmonella* strains.

Plasmids are extrachromosomal, autonomously replicating DNA elements that range in size from 1 kb to over 400 kb. Plasmid genomes are composed of backbone and accessory genes. Backbone genes are considered the "core" components of the plasmid genome and encode plasmid-specific functions such as replication initiation and copy number control (53). Accessory genes include antibiotic and heavy metal resistances, virulence factors and catabolic functions, which can confer an adaptive advantage to the host bacterium in certain environmental conditions. Accessory genes are often located on transposons and integrons, which enable them to move among plasmids and host

chromosomes. Large plasmids (≥30 kb) frequently carry multiple accessory elements and encode transfer operons, making them important mediators of the spread of antibiotic resistances and other adaptive functions. For this reason, we focused our analysis on the diversity and distribution of large plasmids within the ECOR and SAR collections.

E. coli and *Salmonella* are gamma-proteobacteria of the Family Enterobacteriaceae. Whereas *E. coli* is primarily a commensal of the mammalian gastrointestinal tract, *Salmonella* is a facultative intracellular pathogen, causing diseases ranging from gastroenteritis to typhoid fever (27). The ECOR and SAR reference collections were assembled for population studies of *E. coli* and *Salmonella*, therefore the strains are genotypically diverse. Because strain selection was not biased for plasmid-related traits such as antibiotic resistance, these collections are an ideal resource for studying natural plasmid variation. The chromosomal phylogeny and variation of the ECOR and SAR strains have been extensively studied, and the existing genotypic data allow us to evaluate plasmid diversity and distribution in the context of plasmid-host relationships. With this information, we can assess how the plasmids of these collections fit into previously proposed evolutionary and ecological models of plasmid dynamics. Our analysis can also inform future population-based studies of these enterobacteria by providing important missing information about these collections' plasmids.

The ECOR collection includes 72 *E. coli* strains isolated from a variety of mammalian hosts and geographic locations (43). The strains were selected on the basis of multilocus enzyme electrophoresis (MLEE) analysis, with the intention of representing the range of genotypic diversity in natural *E. coli*. Although recent analyses have indicated that ECOR encompasses only a subset of the true diversity in natural *E. coli*

(57), the collection can still be considered a reasonable approximation of *E. coli* strain diversity for the purposes of investigating natural plasmid variation. Phylogenetic analyses of the ECOR collection using different methods, such as MLEE and multi-locus sequence typing, have consistently identified four ECOR subgroups: A, B1, B2 and D (28, 35, 57). Strains assigned to subgroups A and B1 are associated with intestinal infections (i.e., EHEC and EPEC), whereas strains assigned to subgroups B2 and D are strongly associated with extraintestinal infections (ExPEC) (6, 23).

The three *Salmonella* reference collections are SARA, SARB, and SARC. They range from a narrow to a broad representation of the salmonellae, which are classified by species, subspecies and serovar. There are two species of *Salmonella*: *bongori* and *enterica*. The species *S. enterica* is further subdivided into six subspecies: I, II, IIIa, IIIb, IV, and VI (36, 54). A seventh subspecies (VII) was identified in MLEE studies of the SAR collections (11). Strains of this subspecies consistently form a separate clade in nucleotide sequence analyses, although they appear closely related to subspecies IV (40). For consistency with the SAR collection literature, we will use the subspecies VII classification. Within each of the subspecies of *S. enterica*, strains are classified into serovars, such as the subspecies I serovars Typhimurium and Enteritidis. Among all the salmonellae, strains of subspecies I cause the vast majority of infections in warm-blooded animals (17). The other six subspecies, as well as the species *S. bongori*, are often found in cold-blooded hosts and rarely cause human disease (2, 17).

Of the three SAR reference collections, SARA includes the narrowest subset of the salmonellae (3). It is composed of 72 strains from serovars Typhimurium, Saintpaul, Heidelberg, Paratyphi B and Muenchen, all of which belong to subspecies I. The SARB

collection also includes 72 strains of subspecies I, but it has representatives of 37 serovars, providing a broader view of subspecies I strain diversity (10). Finally, the SARC collection of 16 strains has the broadest representation of the salmonellae, including two strains from each of the seven *S. enterica* subspecies and two from *S. bongori* (11).

To investigate natural plasmid variation in the ECOR and SAR collections, we obtained plasmid profiles of strains using a magnetic bead-based BAC preparation method adapted for isolation of large and small plasmids from natural bacterial strains (56). We typed plasmid replicon families using a PCR-based method that targets replication and partitioning genes on the plasmid backbone. We also assessed variation across the plasmid genome, encompassing both backbone and accessory genes, by comparing restriction fragment length polymorphism (RFLP) patterns of large plasmids.

MATERIALS AND METHODS

Bacterial strains

The ECOR collection (n=72) was obtained from Dan Dykhuizen in 1986. The SARA (n=72), SARB (n=72), and SARC (n=16) collections were obtained between 2004 and 2007 from the *Salmonella* Genetic Stock Center (SGSC) at the University of Calgary, along with 71 additional non-subspecies I *Salmonella* strains. One SARB strain, two SARC strains and one ECOR strain were omitted from this analysis for the following reasons. SARB66 was omitted because it had a different plasmid profile from SARA9, even though they are the same strain (originally National Veterinary Services Laboratories number 2816) included in both collections. Based on discussions with Ken Sanderson at the SGSC, we chose to omit SARB66 and retain SARA9. We also omitted

the two subspecies I strains in the SARC collection (SARC1 and SARC2), since SARC is included here to represent non-subspecies I *Salmonella*. Finally, ECOR30 was omitted due to changes in the plasmid profile over the course of data collection. Excluding the omitted strains, we report data for 71 SARA strains, 72 SARB strains, 14 SARC strains and 71 ECOR strains (total = 228).

Plasmid profiling

For all strains, cultures were inoculated from -70°C stocks into 2 ml lysogeny broth (LB) and grown overnight at 37°C with shaking and without antibiotics. Cultures were then diluted 1:50 in LB and incubated at 37°C with shaking for 2.5-3 hours. Total plasmid DNA was isolated following the procedure for Gram-negative organisms described previously (56) using the CosMCPrep kit for high and low copy plasmid purification (Agencourt Biosciences Corp., Beverly, MA). To obtain plasmid profiles, plasmid DNA and appropriate standards (see below) were loaded onto a 0.5% SeaKem Gold agarose (Cambrex BioScience, Walkersville, MD)/1x TAE gel and electrophoresed in 1x TAE (40 mM Tris-acetate, 2 mM $Na_2EDTA\text{-}2H_2O$) at 80 V for 3 hours. Gels were stained with SYBR Green I nucleic acid gel stain (Molecular Probes, Inc., Eugene, OR) and visualized on a Molecular Dynamics FluorImager. For each strain, total plasmid DNA isolation with the CosMCPrep kit was repeated at least once, and plasmid DNA was electrophoresed on a 0.8% SeaKem Gold/1x TAE gel to reveal any plasmid bands obscured by the sheared region in the 0.5% gel.

Differences in the rate of migration of very small plasmids and very large plasmids resulted in inaccurate size estimates using a single standard curve, therefore we estimated plasmid sizes by comparison with standard curves tailored to three different

size ranges. For plasmids <20 kb, we used pTYB1 (7,477 bp), pACYC184 (4,245 bp) or pBR322 (4,361 bp), and pUC18 (2,686 bp) as standards. For plasmids between 20 and 100 kb, we used NR1 (94 kb), RP4 (60 kb), and R388 (33 kb). For plasmids ≥100 kb, we used the 165 kb, 120 kb and 95 kb bands of the BACTracker Supercoiled DNA ladder (Epicentre, Madison, WI).

Restriction fragment length polymorphism (RFLP) analysis of plasmids

RFLP analysis was used for two purposes: 1. To identify the supercoiled forms of small plasmids within each strain's plasmid profile, and 2. To compare the large plasmids among the strain collections.

Identification of supercoiled small plasmids within strains' plasmid profiles

The electrophoresis conditions used for plasmid profiling allowed different forms of small (<20 kb) plasmids, such as multimers or nicked forms, to electrophorese easily through the gel, resulting in multiple bands of the same plasmid. To identify the supercoiled forms of small plasmids, we isolated plasmid DNA from 5 ml of overnight cultures using the Qiagen Plasmid Miniprep kit following the manufacturer's instructions. Plasmid DNA was electrophoresed on 0.8% SeaKemGold/1x TAE gels at 80 V for 3 hours. Gels were stained with SYBR Green I, and all small plasmid bands <20 kb were excised. We then digested the agarose-embedded plasmid DNA. Gel slices were placed into microfuge tubes, and 1 ml of TE (10 mM Tris, 1 mM EDTA, pH 8) was added to equilibrate the slices. Tubes were incubated at 4°C on a gently rotating platform for at least 3 hours. TE was removed by pipette, and 100 μl of restriction buffer (1x restriction buffer, 1x BSA in autoclaved MilliQ water) was added. Microfuge tubes were returned to the rotating platform at 4°C and incubated for at least 30 minutes. Restriction

buffer was removed by pipette, and fresh restriction buffer (100 μl) was added, along with 15 U of restriction enzyme. For small plasmid analysis, we did a simultaneous digestion with 15 U each of *AccI* and *NcoI*. After gently flicking to mix solution, microfuge tubes were incubated at 37°C with 50 rpm rotation for at least 16 hours.

Following digestion, tubes were cooled on ice for at least 15 minutes. The restriction buffer was removed by pipette and replaced with 1 ml of 1x TBE. The tubes were then held on ice for at least 30 minutes. Slices were loaded into the wells of a 0.75% Sigma/1x TBE gel, along with the GeneRuler 1 kb Plus DNA ladder (Fermentas, Glen Burnie, MD), and electrophoresed at 35 V for 15 h. Gels were stained with SYBR Green I and imaged on either a FluorImager or a Typhoon (GE Healthcare, Piscataway, NJ). The RFLP patterns of small plasmid bands within a strain's plasmid profile were visually compared. When the patterns of different bands matched, the smallest band was identified as the supercoiled form of the small plasmid and included in the final data for that strain's plasmid profile. The other bands were not reported in the plasmid profile.

Analysis of large plasmids

RFLP patterns of large (≥30 kb) plasmids from the ECOR and SAR collections were obtained by in-slice digestion. Total plasmid DNA was isolated from strains with the CosMCPrep kit as described above for plasmid profiling, with some modifications. The CosMCPrep procedure was scaled up by two, so that we used 3 ml of subculture and twice the volume of reagents for all steps except the final elution step. Rather than double the volume of elution buffer, we did two sequential elutions with the standard volume and pooled the eluates. Also, incubation time for the lysis step was increased to 10 minutes for *Salmonella* strains to improve plasmid DNA yield.

Total plasmid DNA was electrophoresed on 0.5% SeaKemGold/1x TAE gels at 35 V for 15 hours, stained with SYBR Green I, and imaged on the FluorImager. Large plasmid bands were excised from the gel, and agarose-embedded plasmid DNA was digested with a single enzyme following the in-slice digestion protocol described above for small plasmid analysis. Digested slices were loaded into the wells of a 0.75% Sigma/1x TBE gel, along with GeneRuler 1 kb Plus DNA ladder, and electrophoresed at 35 V for 15 h. Gels were stained with SYBR Green I and imaged on either a FluorImager or a Typhoon (GE Healthcare, Piscataway, NJ).

Cluster analysis of plasmid RFLP patterns

To compare RFLP patterns of large plasmids, we used the GeneTools/GeneDirectory software suite (SynGene, Frederick, MD). Images of the RFLP gels were imported to GeneTools, and restriction fragments were marked by automatic detection using a rolling disk baseline correction with a 30 pixel radius and peak detection minimums of 7 pixels for peak width, 3 pixels for peak height, and 2% for peak volume. Marked fragments larger than 20 kb were manually deleted because of the difficulty in accurately sizing such large linear fragments. Marked fragments smaller than 2.5 kb were also manually deleted because these small fragments were not reproducible. Any fragments between 2.5-20 kb that were missed by automatic detection were added manually. The sizes of the marked fragments were then calculated by comparison to the bands of the GeneRuler ladder using a logarithmic piecewise linear calibration. The information from GeneTools was imported into GeneDirectory, and cluster analyses of RFLP patterns were done using the unweighted pair group method with arithmetic mean (UPGMA) with the Jaccard coefficient and a tolerance of 4.5%.

Plasmid replicon typing

For all ECOR and SAR strains with large plasmids (n=129), total plasmid DNA was isolated with the CosMCPrep kit as described above for plasmid profiling. An aliquot of the plasmid prep was electrophoresed on a 0.5% SeaKemGold/1x TAE gel at 80 V for 3 hours, stained with SYBR Green I and imaged on a FluorImager to confirm that no changes had occurred in the plasmid profiles. Plasmid preps were then diluted to a concentration of <50 ng/μl with the CosMCPrep elution buffer. This dilution optimizes the replicon typing results by limiting the total amount of plasmid DNA in the PCR reactions to <100 ng (A. Carattoli, personal communication).

The replicon typing protocol consists of 18 primer pairs divided into 5 multiplex and 3 simplex reactions (16). Because of difficulties with false negatives and/or nonspecific amplification, we omitted the IncP, IncK, and IncB/O primer pairs. PCR reactions included 2 μl of the diluted plasmid prep as template, 1.25 U of Taq polymerase (Promega, Madison, WI), 1x magnesium-free reaction buffer, 4 mM magnesium, 0.25 mM each dNTP, 0.3 μM each primer and nuclease-free water to a final volume of 25 μl. PCR cycle conditions were 94°C for 5 minutes; 30 cycles of 94°C for 1 minute, 60°C (or 52°C for F simplex reaction) for 30 seconds, 72°C for 1 minute; and 72°C for 5 minutes.

A 15 μl aliquot of the reaction was loaded on a 1.5% Metaphor agarose (Cambrex BioScience, Walkersville, MD)/1x TBE gel, along with a 100 bp ladder (New England Biolabs, Ipswich, MA), and electrophoresed at 100 V for 1.5 hours. Gels were then stained with ethidium bromide and visualized on the FluorImager. Strains were scored as positive for a plasmid replicon family when an amplicon of the appropriate size was detected.

In silico *analysis of* E. coli *and* Salmonella *plasmid sequences from GenBank*

In silico plasmid replicon typing

We used the Pattern Locator program (41) to search plasmid sequences for the 18 primer pairs described in the PCR-based replicon typing protocol (16), including those that we omitted from the *in vitro* analysis. The search command was formatted as follows:

<>{forward primer sequence}[2](N)[0:850]{reverse complement of reverse primer sequence}[2]

This command allows for 2 mismatches in each primer sequence and requires that forward and reverse primers occur within 850 bp of each other. The gap length includes all predicted amplicon sizes. The search command was tested with the target sequences used by Carattoli and colleagues to design the replicon typing primers (16). For each target sequence, the search command identified the corresponding primer pair at a distance equal to the predicted amplicon size. Primer pairs were specific for their target sequences.

In silico RFLP analysis

For each plasmid sequence, *AccI* fragments were predicted using the Gene Construction Kit (Textco Biosoftware, Inc., West Lebanon, NH). Predicted fragment patterns were displayed in the Gel format, along with a lane of automatically generated size markers. These Gel files were imported into GeneTools, and fragments were automatically marked using the "track borders" baseline correction method and peak detection minimums of 2 pixels for peak height, 1 pixel for peak width and 1% for peak volume. To be consistent with the *in vitro* RFLP analysis, marked fragments larger than

20 kb and smaller than 2.5 kb were deleted. The sizes of the marked fragments were then calculated by comparison to the size markers using a logarithmic piecewise linear calibration. The information from GeneTools was imported into GeneDirectory, and cluster analyses of the predicted RFLP patterns were done using the unweighted pair group method with arithmetic mean (UPGMA) with the Jaccard coefficient and a tolerance of 1%. Tolerance was lower than that of the *in vitro* analysis to account for the sharpness of the computer-generated fragment bands and the lack of gel-to-gel variation.

RESULTS

Plasmid profiles

Plasmid prevalence in ECOR and SAR collections

Plasmids in general and large (≥30 kb) plasmids in particular are abundant in the natural *E. coli* and *Salmonella* strains of these collections (Table 3.1). The overall ECOR plasmid prevalence observed here agrees with previous work showing plasmids in 86% of ECOR strains (45). Over two-thirds of the ECOR collection had at least one large plasmid. In the SARA and SARB collections, which include subspecies I *Salmonella* strains, overall plasmid prevalence was moderately high, but there were significantly fewer plasmid-bearing strains than in ECOR ($p<0.01$, exact binomial test). Although the majority of SARA and SARB strains had large plasmids, large plasmid prevalence was also significantly lower than that of ECOR ($p<0.01$, exact binomial test). No significant differences were detected in overall or large plasmid prevalence between SARA and SARB.

In the non-subspecies I SARC strains, plasmid prevalence was dramatically lower both overall and for large plasmids compared to SARA and SARB ($p<0.01$, exact

binomial test). To further investigate this difference, we obtained 71 additional non-subspecies I strains, referred to as SGSC strains, that are part of the set from which SARC strains were chosen (11). About half of the SGSC strains had plasmids, and 45% had large plasmids. This is a significant increase in both overall and large plasmid prevalence compared to SARC ($p<0.01$, exact binomial test), showing that the 14 SARC strains under-represent the presence of plasmids in non-subspecies I *Salmonella*. There was no significant difference in plasmid prevalence between the SGSC strains and SARA or SARB, indicating that subspecies I and non-subspecies I *Salmonella* are not inherently different in plasmid prevalence.

Plasmid prevalence in ECOR phylogenetic subgroups

Within the ECOR collection, the four phylogenetic subgroups (A, B1, B2 and D) had different plasmid prevalences (Table 3.1). All strains of subgroups A and D had plasmids, whereas 9 subgroup B1 strains and 13 subgroup B2 strains did. Considering large plasmids, ECOR B2 had the lowest prevalence, whereas the majority of strains in the other subgroups had large plasmids. This difference in plasmid carriage between B2 and the other ECOR subgroups is further illustrated by comparing the total plasmid content (total kilobases of small and large plasmids) per strain. A nonparametric Kruskal-Wallis test showed a significant difference in the median values of the total plasmid content per strain among the four subgroups ($p<0.01$). Posthoc pairwise Mann-Whitney tests showed a significant difference between subgroup B2 and subgroups A and D. All other pairwise comparisons showed no significant differences, demonstrating that the strains in ECOR subgroup B2 differ from most of the ECOR collection in plasmid carriage.

Number and size of plasmids in ECOR and SAR collections

Strains with multiple plasmids were more common in *E. coli* than *Salmonella*. Most ECOR strains (63.5%) had multiple plasmids, compared to only about a fifth of SARA and SARB strains (23.5% and 22.5%, respectively). Although only two SARC strains (14%) had multiple plasmids, the proportion of multi-plasmid SGSC strains (21%) more closely resembled that of the subspecies I *Salmonella* collections.

ECOR strains had up to 8 plasmids per strain, averaging 2.4 ± 1.9. By contrast, SARA and SARB strains had up to 4 plasmids per strain, and both collections averaged approximately 1.0 ± 1.0 plasmids. Averages were slightly lower in SARC (0.5 ± 0.9) and the SGSC strains (0.8 ± 0.9), which both had up to 3 plasmids per strain. Because RFLP analysis was done for all <20 kb bands in each plasmid profile, we were able to include only supercoiled plasmids in the profile data and exclude other forms of small plasmids such as multimers.

The distributions of plasmid sizes for the ECOR and SAR collections were roughly bimodal (Figure 3.1). One mode of small plasmids occurred at ≤10 kb, and the other mode occurred between 80 and 110 kb. Very large plasmids up to 181 kb were detected in ECOR, and three plasmids >200 kb were observed in the SAR collections.

Throughout the Results and Discussion, the strain name will also be used as the plasmid name for strains with one large plasmid. For strains with multiple large plasmids, the plasmids are numbered according to size, starting with the largest. This number is then added to the strain name after a hyphen to designate an individual plasmid in a multi-plasmid strain (for example, the largest plasmid in ECOR3 is designated pECOR3-1).

Plasmid replicon typing

Plasmid replicon typing is a classification scheme in which backbone genes encoding replication and partitioning systems are used to define plasmid families. It is based on plasmid incompatibility, in which two plasmids with similar replication or partitioning systems cannot stably coexist in the same cell without selection (42). Using this property, plasmids can be grouped into incompatibility families, also known as replicon families. Approximately 26 replicon families are recognized for plasmids of Enterobacteriaceae. Building on previous methods of incompatibility testing (22) and probe hybridization (19), a PCR-based replicon typing method was developed recently that uses qualitative PCR to detect the replication or partitioning genes defining the majority of currently described enterobacterial plasmid replicon families (16).

We used this method to obtain plasmid replicon family profiles for all ECOR and SAR strains with large (≥30 kb) plasmids (n=129). The method includes primer pairs recognizing 18 enterobacterial plasmid replicons. We omitted the IncP, IncK, and IncB/O primer pairs because of difficulties with false negatives and nonspecific amplification. The resulting set of primers recognized 15 plasmid replicons (Table 3.2). Strains that did not produce a PCR amplicon with these primer pairs are denoted as "untypable". We used total plasmid DNA, rather than total genomic DNA, as template, however with either template, the typing method generates a plasmid replicon family profile for the strain, but does not differentiate individual plasmids within a multi-plasmid strain. Thus, some strains will have positive results for more than one replicon (Table 3.2).

Prevalence of plasmid replicon families in ECOR and SAR collections

For ECOR strains with large plasmids (n=50), IncF was the predominant replicon family detected, with 66% of strains producing an IncF amplicon. The IncF primer pair recognizes most of the replicons in the heterogeneous IncF group (16). Three other primer pairs target specific replicons of the IncFI subgroup, namely RepFIA, RepFIB, and RepFIC. In the ECOR collection, these three replicons were only detected in strains that were also positive for IncF. RepFIB occurred in 19 strains with large plasmids, whereas RepFIA and RepFIC occurred in only three and one strains, respectively. Combinations of the three replicons were rare. Two strains produced amplicons for both RepFIA and RepFIB, whereas one strain produced amplicons for both RepFIB and RepFIC. After IncF, IncI1 was the second most prevalent replicon family in ECOR, detected in seven strains with large plasmids (14%). The only other families detected were IncY, IncN and IncX, which occurred in four, two and one strains, respectively. Thirteen strains with large plasmids (26%) produced no PCR amplicons and were classified as "untypable".

In contrast to ECOR, SAR strains with large plasmids (n=76) were mostly negative for the IncF and three IncFI amplicons. Instead, the IncFII amplicon was predominant, detected in 31.5% of SAR strains. In SARA, 15 of the 37 strains with large plasmids (40.5%) produced an IncFII amplicon, whereas in the broader SARB collection, nine of the 36 strains with large plasmids produced IncFII amplicons (25%). The IncFII primer pair recognizes the RepFIIA replicon of *Salmonella* virulence plasmids (16). Although these plasmids belong to the larger, heterogeneous IncF replicon family, it was previously observed that RepFIIA replicons were not detected by the IncF primer pair

(16), which explains the lack of an IncF amplicon for these SAR strains. The prevalence of RepFIIA in the SAR collections may reflect the presence of virulence plasmids within these subspecies I *Salmonella* strains. RepFIIA was not detected in the non-subspecies I SARC strains, indicating that these strains do not harbor virulence plasmids.

As for ECOR, IncI1 was the second most prevalent replicon family in the SAR collections, occurring about equally in SARA and SARB. The IncA/C and IncN families were also detected in both SARA and SARB. IncHI2 was detected in SARA, but not SARB. The IncHI family includes >180 kb plasmids from *Serratia marcescens* and *S. enterica* serovar Typhi, which are known to encode multiple antibiotic and metal resistances (26). The two SARA strains producing IncHI2 amplicons each carried one >200 kb plasmid, consistent with the reported large sizes of members of this replicon family. Forty-seven percent of the SARB strains were "untypable", as opposed to 21.5% of the SARA strains, suggesting that the broader SARB collection contains more diversity in plasmid backbones than SARA. Only one non-subspecies I SARC strain was typable using the PCR-based method.

Four plasmid replicon families (IncHI1, IncL/M, IncW, and IncT) were not detected in the ECOR or SAR strains with large plasmids. During development of the protocol, two confirmed IncL/M plasmids did not produce the appropriate amplicon with the L/M primer pair (16). Because of the potential for false negatives with this primer pair, the ECOR and SAR collections may contain L/M plasmids that were not detected by this method.

Distribution of plasmid replicon families in ECOR and SAR collections

To identify patterns in the occurrence of plasmid replicon families, we overlaid the replicon typing results on a dendrogram of ECOR constructed from chromosomal multilocus enzyme electrophoresis (MLEE) (28) (Figure 3.2). ECOR subgroup A had the highest diversity in replicon families, with IncF, IncI1, IncY, IncX and IncN represented as well as seven strains with untypable plasmids. In contrast, the other three ECOR subgroups had primarily IncF and IncI1 replicons.

Within ECOR subgroup A, there was a striking absence of IncF plasmids in two clades of strains on the dendrogram (Figure 3.2). In these two clades, 10 strains lacked large IncF plasmids, whereas in the rest of the ECOR collection, only 8 strains had large plasmids but no IncF amplicon. There are no obvious differences in host or region of isolation for these strains compared to the rest of ECOR, indicating that perhaps the lack of IncF plasmids is due to a feature of their chromosomes, which may affect IncF plasmid transfer or maintenance. The branch topology of ECOR subgroup A has been observed to change when using different phylogenetic methods (35, 57), which disrupts the clades of strains identified by MLEE. However, since the assignment of these strains to subgroup A is preserved regardless of phylogenetic method, this concentrated lack of IncF plasmids within ECOR subgroup A is still significant.

Although ECOR 31, 43, 37 and 42 were originally assigned to a fifth ECOR subgroup (E), subsequent work indicated that these strains do not constitute a distinctive subgroup. Even so, it is interesting to note that IncY was detected in three of these four strains, and the only other IncY plasmid occurred in subgroup A (ECOR13). The

prevalence of IncY plasmids in strains outside of the recognized ECOR subgroups may indicate a barrier to IncY plasmid transfer or maintenance within the subgroups.

In SARA, RepFIIA only occurred in strains of serovar Typhimurium. In SARB, strains from serovars Choleraesuis, Dublin, Enteritidis, Paratyphi C and Typhimurium had plasmids with RepFIIA. The prevalence of the other replicon families was too low to assess any pattern in distribution, with the exception of IncI1. This family was detected in all five serovars in SARA and in nine of the 37 serovars in SARB, suggesting that IncI1 plasmids are relatively unrestricted in their distribution among *Salmonella* serovars.

Identification of multi-replicon plasmids

Combining plasmid profile and replicon typing data, we identified three plasmids that may encode more than one replication system. The plasmid profiles of the host strains showed a single plasmid, and the replicon typing results were positive for two plasmid replicon families. ECOR28 had a 127 kb plasmid and produced amplicons for IncF, RepFIA, RepFIB and IncI1. SARA10 had a 105 kb plasmid and produced amplicons for IncFII and IncI1. SARB11 had a 96 kb plasmid and produced amplicons for IncA/C and IncI1. The IncI1 replicon is present on all three putative multi-replicon plasmids, whereas the other replicon(s) are different.

Comparison of PCR-based replicon typing with probe hybridization method

Our results for RepFIB with the PCR-based replicon typing method are in good agreement with two previous studies in which a RepFIB probe was used to screen the ECOR and SARA collections by hybridization (7, 9). Seventeen ECOR strains were positive for RepFIB with both methods. Two strains produced RepFIB amplicons by

PCR but were not positive by hybridization. Conversely, three strains were positive for RepFIB by hybridization, but we did not detect a RepFIB amplicon.

For SARA, two strains were positive for RepFIB with both methods. One strain was positive by PCR but not hybridization, and one strain was positive by hybridization but not PCR. Overall, the comparison of these two methods shows good general agreement with some discrepancies that may result from the use of different RepFIB sequences for probe and primer design.

Plasmid restriction fragment length polymorphism (RFLP) analysis

To assess the diversity and distribution of large (≥30 kb) plasmids in the ECOR and SAR collections, we compared RFLP patterns obtained by excising individual plasmids from plasmid profile gels and digesting the agarose-embedded plasmid DNA. RFLP analysis has advantages over hybridization-based approaches, such as microarrays, because it is not limited to pre-defined gene targets. RFLP patterns are also more likely to show changes when rearrangements and duplications have occurred, which are important aspects of natural plasmid variation. In-gel-slice digestion enables analysis of plasmids isolated directly from their natural hosts, regardless of the presence of multiple plasmids. This is an advantage over previous studies of plasmid diversity using RFLP, which disregarded plasmids from multi-plasmid strains or relied on conjugation or transformation to transfer plasmids into laboratory hosts.

We used two stages of RFLP pattern comparison and cluster analysis. The first stage served as an initial screen for potentially similar plasmids. Large ECOR and SAR plasmids (n=164) were digested with *AccI*, the RFLP patterns were compared, and those clustering above an arbitrary threshold were identified. These plasmids were digested

with a different enzyme, and the RFLP patterns were compared to determine whether the plasmids were identical, similar or dissimilar.

Initial RFLP cluster analysis of plasmids

In-gel-slice digestion using *AccI* generated clear, reproducible RFLP patterns for 158 large ECOR and SAR plasmids. *AccI* has a degenerate recognition sequence (GTMKAC), which slightly reduces the chance that a single nucleotide change will affect the restriction fragment pattern. The average number of fragments in the range used for cluster analysis (2.5 – 20 kb) was 11 ± 4, which was sufficient for pattern comparison. We were unable to obtain *AccI* RFLP patterns for four large ECOR plasmids and two large SARA plasmids, therefore these plasmids were not included in the RFLP analysis.

Cluster analysis of *AccI* patterns was used to construct separate dendrograms for ECOR and SAR plasmids (Figures 3.3 and 3.4). We chose a threshold of 60% similarity to define clusters of similar plasmids. Based on this criterion, ECOR large plasmids (n=73) were highly variable. Fifty-one plasmids (70%) had unique *AccI* RFLP patterns, i.e. they did not cluster with any other pattern at or above 60% similarity. Twenty-two plasmids were in 10 clusters at or above 60% similarity. All 10 clusters had two or three plasmids only, highlighting the lack of any widely distributed RFLP pattern in the ECOR collection.

Of the SAR large plasmids analyzed by *AccI* digestion (n=85), 35 (41%) had a unique RFLP pattern, whereas 50 (59%) were in 15 clusters at or above 60% similarity. Three clusters had four or more plasmids each, which may reflect three distinct widely distributed plasmids. One of these clusters had 17 plasmids, all from serovar Typhimurium strains, and included the serovar-specific virulence plasmid pSLT from

strain LT2 (SARA2). This suggests that the other 16 plasmids are highly similar to the virulence plasmid. Most of the other 12 SAR clusters consisted of pairs of plasmids.

Although most SAR plasmids produced a suitable number of *AccI* fragments for comparison, 19 SAR plasmids had so many *AccI* fragments (average = 16 ± 2.5) that RFLP pattern comparison was not accurate. We included all 19 plasmids in the second stage of analysis to obtain more accurate comparisons (see **RFLP analysis with different enzyme** below), even though the main node for this group was below the 60% similarity threshold (Figure 3.4).

In addition to the separate cluster analyses, we also compared all ECOR and SAR *AccI* plasmid RFLP patterns to determine whether any plasmids occurred in both *E. coli* and *Salmonella*. In the resulting dendrogram, 34 of the 158 large plasmids (21.5%) were in 14 clusters of ECOR and SAR plasmids at or above 60% similarity. Most clusters had two or three plasmids, but two clusters had four plasmids each. All 34 plasmids were digested with a different enzyme to obtain more precise comparisons (see below).

We checked the accuracy of the *AccI* cluster analysis using seven SARA/SARB replicates (i.e., individual strains that were included in both collections). In this first stage of analysis, we expected the plasmid replicates to cluster with each other at or above 60% similarity. The plasmids of five SARA/SARB replicates did cluster at or above this threshold, but the plasmids from the other two replicates clustered in the group with too many *AccI* fragments. Therefore, they were included in the second stage of analysis (see below), even though they did not cluster at or above 60% similarity as expected.

RFLP analysis with different enzyme

Plasmids from clusters identified in the initial *AccI* analysis were digested with a different enzyme to obtain more precise comparisons. For most plasmids, we used *BglII*, which has a conserved recognition sequence (AGATCT). This enzyme generated an average of 8 ± 3.75 fragments in the range used for cluster analysis (2.5 – 20 kb). Four SAR plasmids did not have enough *BglII* fragments for accurate comparison, so we used *NcoI* (CCATGG) or *HindIII* (AAGCTT).

Separate dendrograms were constructed for ECOR and SAR plasmids for each enzyme and then examined to determine whether clusters identified in the *AccI* analysis also occurred in this analysis. For each cluster, the plasmids' relationships were classified by percent similarity shown on the dendrogram. Plasmids with RFLP patterns clustering at 100% similarity were classified as identical, those clustering between 90-100% similarity were highly similar, and those clustering between 50-90% were similar. If plasmids clustered together in the *AccI* analysis but did not cluster at or above 50% similarity when analyzed with a different enzyme, they were classified as not similar.

For all 22 ECOR plasmids analyzed, *BglII* digestion produced clear RFLP patterns. Only four of the 10 clusters identified in the initial analysis clustered at or above 50% similarity (Table 3.3). Two clusters, each with two plasmids, grouped at 100% similarity, indicating identical plasmids. One cluster of two plasmids was classified as highly similar, and one cluster of three plasmids was classified as similar. The remaining six clusters from the initial analysis, which included 13 plasmids, were classified as not similar. In summary, of the 73 large ECOR plasmids compared, four (5.5%) were identical to another plasmid, two (3%) were highly similar, 3 (4%) were similar, and 64

(87.5%) were not similar. Plasmids in the ECOR collection are clearly highly variable, and no widely distributed plasmid was identified.

Fifty-three SAR plasmids were digested with *BglII*, including all 19 plasmids that had too many *AccI* fragments for accurate comparison in the initial analysis. For these plasmids, *BglII* produced fewer fragments (average = 8.5 ± 4), enabling accurate pattern comparison. Four SAR plasmids could not be digested with *BglII*, therefore we used *NcoI* (for pSARB6 and pSARB69) or *HindIII* (for pSARC13-2 and pSARC14).

Of the 15 SAR clusters identified in the initial *AccI* analysis, nine occurred at or above 50% similarity after analysis with a different enzyme (Table 3.3). Two clusters, one of two plasmids and another of three plasmids, were classified as identical. Both of these clusters were serovar-specific. Seven clusters were classified as similar. This included the large, serovar Typhimurium-specific cluster of plasmids seen in the *AccI* analysis (Figure 3.5). Most of these plasmids were either identical or highly similar to the pSLT virulence plasmid from SARA2. Three outliers clustered with the larger group at 58% similarity, therefore the whole group is classified as similar. The RFLP patterns of these plasmids show minimal variation, primarily small shifts in the sizes of restriction fragments (Figure 3.5). Based on this comparison, all of these plasmids can be considered pSLT-like, and thus the majority of the Typhimurium strains in the SAR collections harbor pSLT or a pSLT-like plasmid. We were not able to obtain a second stage digest of pSARA6 because of difficulty isolating the plasmid, possibly due to the presence of a lytic phage in the strain (data not shown). However, based on the similarity of its *AccI* RFLP pattern, we consider this plasmid to be pSLT-like. The pSLT-like RFLP pattern was the only one found in more than four strains in the ECOR and SAR collections.

Another cluster of similar plasmids consisted of a pair of plasmids isolated from the two SARC representatives of subspecies VI. pSARC13-2 is about 30 kb smaller than pSARC14 (Table 3.3). The RFLP pattern for the larger plasmid has additional fragments compared to the smaller plasmid (data not shown), indicating that these plasmids evolved from the same lineage by gain or loss of DNA. None of the SARC plasmids clustered with any subspecies I *Salmonella* plasmids.

Apart from the pSLT-like cluster and the subspecies VI-specific cluster, the other five clusters of similar SAR plasmids were mixed-serovar, suggesting horizontal transfer between serovars. One cluster had four plasmids: pSARA33 and pSARB25, which clustered at 80% similarity, and pSARB19 and pSARB20, which clustered at 73% similarity. These two pairs were joined by a node at 57% similarity, indicating that this cluster may represent a widespread plasmid lineage undergoing divergence. The other four mixed-serovar clusters had only two or three plasmids.

One cluster identified in the initial *AccI* analysis, which included the 30 kb plasmids pSARB3 and pSARB4, could not be digested with any of the enzymes tested and was therefore omitted from this stage of analysis.

We checked the accuracy of the second stage cluster analysis using the seven SARA/SARB replicates, which were all digested with *BglII*. Five replicates clustered at 100% similarity, and the other two clustered at 98% similarity. Overall, the second stage of analysis clustered the seven SARA/SARB replicates with high accuracy. Because these replicates are the same plasmid represented twice, we included only the SARA plasmids in the summary in Table 3.3.

In summary, of the 75 large SAR plasmids compared, 33 plasmids (44%) were grouped in 9 clusters. Forty-two plasmids (56%) were not similar to any other plasmid in the SAR collections. Two clusters had four or more plasmids, suggesting the possibility of two distinct widespread plasmid lineages in *Salmonella*.

In the initial *AccI* analysis, comparing all ECOR and SAR RFLP patterns led to the identification of 14 clusters of ECOR and SAR plasmids at or above 60% similarity. All 34 plasmids from these clusters were digested with *BglII*. In a dendrogram of all ECOR and SAR *BglII* RFLP patterns, none of the clusters identified in the initial analysis occurred at or above 50% similarity, indicating that the plasmids are not similar. Digestion with a different enzyme did not provide evidence of particular plasmids occurring in both the ECOR and SAR collections.

In silico *analysis of* E. coli *and* Salmonella *plasmid sequences*

As of January 2009, there were 43 large (≥30 kb) *Escherichia coli* plasmids and 26 large *Salmonella* plasmids available as complete sequences in NCBI's GenBank. This does not include plasmids for which no host information was reported. The *Salmonella* plasmids were all isolated from subspecies I strains, comprising 10 serovars and one unidentified serovar. Serovars with the most plasmid sequences were Dublin (n=6), Choleraesuis (n=5), Kentucky (n=3), Typhi (n=3), and Typhimurium (n=3). We simulated plasmid replicon typing and RFLP analysis with *AccI in silico* for these 69 plasmid sequences to compare the natural variation observed in the ECOR and SAR reference collections with the coverage of *E. coli* and *Salmonella* plasmids in the NCBI sequence database.

In silico replicon typing

For *E. coli* plasmid sequences, IncF was the predominant replicon family (Table 3.2), occurring in 58% of the plasmids. Although IncI1 was the second most prevalent family in ECOR, only three IncI1 *E. coli* plasmids have been sequenced. No plasmid sequences belonged to IncY or IncX, which were detected in four and one ECOR strains, respectively. Ten *E. coli* plasmid sequences (23%) were untypable, suggesting that some of the plasmid replicon diversity observed in ECOR may be represented in currently available sequences. For *Salmonella* plasmid sequences, IncFII was predominant, as expected from the presence of multiple virulence plasmids in the database. IncI1 was also under-represented in *Salmonella* plasmids in GenBank compared to its prevalence in the SAR collections. Five plasmid sequences belonged to two replicon families (IncHI1 and IncW) that were not detected in the SAR collections. Seven *Salmonella* plasmid sequences (27%) were untypable. This proportion is lower than that of the SAR collections, suggesting that additional sequence coverage may be needed to assess the extent of diversity in *Salmonella* plasmid replicons.

Only one multi-replicon plasmid was identified in GenBank: the *E. coli* pCoo cointegrate plasmid possessing IncF and IncI1 replicons (24). This was surprising considering that we identified three putative multi-replicon plasmids in the ECOR and SAR collections, even though we were limited to strains with a single plasmid in the plasmid profile. Multi-replicon plasmids in *E. coli* and *Salmonella* appear to be under-represented in the available sequences. Because of their potential for generating novel combinations of accessory genes, such as antibiotic resistances and virulence factors, more multi-replicon plasmids should be sequenced.

In silico RFLP analysis

Cluster analysis of predicted *AccI* patterns identified seven *E. coli* plasmid sequences (16%) in three clusters at or above the 60% similarity threshold (Table 3.4). One cluster included three variants of the pO157 virulence plasmid found in enterohemorrhagic *E. coli* O157:H7 strains (14, 37). pSFO157, a more divergent variant of pO157, was not included in this cluster due to the loss of some pO157 genes and the acquisition of other genes leading to a 30 kb size difference (13). The remaining two clusters of *E. coli* plasmid sequences were pairs of plasmids with minor, artificially introduced differences in gene content. pMAR7 was derived from pMAR2 by insertion of a 5 kb transposon as a selective marker (12, 33), which resulted in 3 differences between the plasmids' RFLP patterns. pLEW517 was isolated from its natural host and a transconjugant (56). The versions differ by a 1,342 bp region overlapping repetitive elements, which resulted in 3 differences between the plasmids' RFLP patterns. Apart from these three clusters, thirty-six *E. coli* plasmid sequences (84%) had a unique predicted *AccI* pattern, which is very similar to the proportion of ECOR plasmids with unique patterns (87.5%).

Four *Salmonella* plasmid sequences (15.5%) were in two clusters at or above 60% similarity. One cluster at 100% similarity included two variants of the serovar Choleraesuis virulence plasmid, which differ at only 92 nucleotide positions (58). The other cluster most likely represents variants of a serovar Dublin virulence plasmid (18). Twenty-two *Salmonella* plasmid sequences (84.5%) had a unique RFLP pattern. This differs from the SAR collections, most likely due to the high representation of the pSLT

virulence plasmid in SARA. One version of pSLT was included in the sequence database, and it did not cluster with any other plasmid.

As observed in the ECOR and SAR collections, cluster analysis of all 69 plasmid sequences did not identify any clusters with *E. coli* and *Salmonella* plasmids at or above 60% similarity.

DISCUSSION

Plasmid profiles of ECOR, SARA, SARB and SARC showed that plasmids are a sizeable mobile genetic element pool in both *E. coli* and *Salmonella*. Large plasmids, which frequently encode transfer operons and multiple accessory elements, are a considerable part of this mobile genetic element pool. The majority of strains in ECOR, SARA and SARB had large plasmids. Most of the nonsubspecies I strains of SARC lacked plasmids, but almost half of the additional set of nonsubspecies I strains we analyzed had large plasmids. The prevalence of large plasmids in these collections, which include strains isolated between the mid-1960s and 1989, is much higher than that reported for the Murray collection of enteric bacteria isolated before 1954. In the Murray collection, only 25% of *E. coli* strains and 22% of *Salmonella* strains had large conjugative plasmids (32). The increase in large plasmids observed in the ECOR and SAR collections may be due partially to selective pressure from antibiotics, since previous studies found antibiotic resistance in none of the Murray collection strains but in 20-25% of ECOR strains (31, 39). However, this documented increase in antibiotic resistance does not completely account for the dramatic difference in large plasmids between the Murray collection and the ECOR and SAR collections. Future plasmid

sequencing efforts should focus on the cryptic plasmids of these collections to identify their phenotypes and elucidate their ecological roles.

Comparison of both overall and large plasmid prevalences showed that plasmids were significantly more frequent in the ECOR collection than in the SAR collections. The high plasmid prevalence observed in ECOR is consistent with recent analyses of naturally occurring *E. coli*, such as a study of isolates from wild Australian mammals in which 100% of *E. coli* strains had plasmids (50). It is possible that this strong trend for natural *E. coli* to harbor plasmids is a result of mutations in the *E. coli* chromosome that reduce the fitness cost of plasmid carriage. This phenomenon has been demonstrated in the laboratory in long-term studies of plasmid-host coevolution (5, 20), but its occurrence in nature has not been explored.

Among the four ECOR phylogenetic subgroups, both large plasmid prevalence and total plasmid content per strain were lower in subgroup B2. This was surprising given the association of this subgroup with extraintestinal infections (6, 23). The virulence loci of *E. coli* B2 strains are primarily located on the chromosome rather than plasmids, but it is striking that these strains do not harbor large plasmids carrying antibiotic resistances. The B2 strains representing this subgroup in ECOR may not reflect plasmid prevalence in other natural populations of *E. coli* B2 strains. However, if limited plasmid carriage is a feature of B2 strains, the decoupling of virulence and plasmid carriage in *E. coli* B2 contrasts with *Salmonella*, in which virulent strains often harbor plasmids encoding antibiotic resistances and virulence genes such as the *spv* operon. In addition, studies have implicated horizontal gene transfer in the spread of pathogenicity islands and virulence genes from *E. coli* B2 strains to other subgroups (4, 6). If plasmids

are not common to subgroup B2, bacteriophage may be more important in the transfer of virulence loci between *E. coli* lineages.

The few large plasmids we detected in ECOR B2 strains may differ in accessory element content from other ECOR plasmids. Using data from a previous survey of six IS elements within ECOR (47), 67% of ECOR B2 strains with large plasmids did not have plasmid-borne copies of the IS elements (IS*1*, IS*2*, IS*3*, IS*4*, IS*5* and IS*30*), whereas only 21-33% of large-plasmid-bearing strains from the other ECOR subgroups lacked plasmid-borne copies. This suggests that, in addition to the differences in plasmid carriage between subgroup B2 and the other ECOR subgroups, B2 strains may have a more genetically isolated pool of plasmids. This hypothesis is further supported by our RFLP analysis, which showed that none of the B2 large plasmids were similar by cluster analysis to any other ECOR or SAR plasmid (Table 3.3).

Within the SAR collections, there was no difference in plasmid carriage between the two subspecies I *Salmonella* collections SARA and SARB, indicating that the five serovars of SARA are not a different plasmid host environment than the larger group of subspecies I *Salmonella* serovars in SARB. Comparisons of SARA and SARB with the non-subspecies I SARC strains showed a dramatic difference in plasmid carriage, but analysis of additional non-subspecies I strains revealed that plasmid prevalence is under-represented in SARC. Despite their different host ranges and disease spectrums, non-subspecies I and subspecies I *Salmonella* are similar in plasmid carriage, both overall and for large plasmids. This observation contrasts with recent statements that plasmids are absent in subspecies of *Salmonella* not associated with disease (46). Very little information is available regarding non-subspecies I *Salmonella* plasmids, therefore the

ecological roles of these plasmids are unclear and the extent of shared gene content among plasmids of different subspecies is unknown.

We used PCR-based replicon typing to identify plasmid replicon families in all strains with large plasmids. The heterogeneous IncF family and IncI1 were prevalent in both *E. coli* and *Salmonella*. These two families were also prevalent in a survey of recent commensal and pathogenic *E. coli* isolates of avian and human origin (34). IncF and IncI1 appear to be highly successful plasmid replicons within enteric bacteria. Interestingly, strains with untypable plasmids, which did not produce amplicons for any of the 15 primer pairs, were also prevalent in ECOR and SAR, comprising almost half of SARB strains with large plasmids. These plasmids may belong to one of the replicon families currently described but not included in our analysis, to a replicon family that has diverged sufficiently from a known family that the primers do not anneal to the target, or to an uncharacterized replicon family. The diversity in plasmid replicons observed in the ECOR and SAR collections is an important element of natural plasmid variation, since it involves divergence in the core components that encode plasmid-specific functions.

Within ECOR, almost two-thirds of the strains with large plasmids belonged to IncF. Of the specific IncFI replicons screened, RepFIB was the most prevalent. Only one occurrence of RepFIC was detected in ECOR. The RepFIC primer pair was designed to specifically target the FIC replicon of the F fertility plasmid (16). Its absence (with one exception) in ECOR suggests that although many of the plasmids in this collection belong to the IncF family, they are divergent from the F plasmid itself. This supports conclusions from a previous study of ECOR that identified F-like plasmids based on hybridization to four F plasmid genes, including RepFIB (9). The absence of the FIC

replicon reported here for ECOR emphasizes that these plasmids are F-like, but not identical to the F plasmid.

In the SAR collections, the IncF plasmids belonged almost entirely to the IncFII family. The IncFII primer pair was designed to detect the RepFIIA replicon of *Salmonella* virulence plasmids (16), which explains the lack of IncFII amplicons in ECOR. Within SARA, RepFIIA was only detected in serovar Typhimurium strains, which often carry the serovar-specific virulence plasmid pSLT. In fact, with one exception, the distribution of RepFIIA in SARA correlates exactly with plasmids belonging to the pSLT-like cluster identified by RFLP analysis. None of the other four serovars in SARA, which are not known to harbor pSLT-like plasmids, had RepFIIA. Within SARB, distribution of the RepFIIA replicon was highly correlated with previously determined distribution of the *spv* operon (8), which contains the virulence loci characteristic of pSLT-like virulence plasmids. Serovars Choleraesuis, Dublin, Enteritidis, Paratyphi C and Typhimurium had plasmids with RepFIIA, and all five of these serovars previously showed positive hybridization to *spv* (8). The RepFIIA primer pair used here appears to very strictly identify pSLT-like *Salmonella* virulence plasmids. It is possible that other SAR strains carry virulence plasmids that are not in the pSLT lineage, and therefore are not detected with the IncFII primer pair.

The IncI1 plasmid replicon family had relatively high prevalence in both *E. coli* and subspecies I *Salmonella* strains. Variation in RFLP patterns of IncI1 plasmids showed that this is not due to a single widespread IncI1 plasmid. Rather, it appears that the IncI1 plasmid backbone is an efficient vehicle for interspecies transfer. It may also contain recombinogenic loci, based on the observation that all three putative multi-

replicon plasmids had an IncI1 replicon. Plasmids from the IncI1 family are frequently associated with antibiotic resistances, particularly extended spectrum cephalosporinases of both the CTX-M and CMY types (30, 38). Over the past decade, resistance to extended spectrum cephalosporins has increased significantly (15, 52), and this may be due in part to the presence of these genes on IncI1 plasmids. Our analysis of ECOR and SAR suggests that IncI1 plasmids were already transferring between *E. coli* and *Salmonella* before the rise of resistance to extended spectrum cephalosporins. In a possible scenario, IncI1 plasmids acquired CTX-M and CMY type beta-lactamase genes, which subsequently benefited from the plasmids' existing adaptations to multiple hosts and spread throughout bacterial populations. In turn, the resistance genes may have provided an additional selective advantage for the maintenance of these plasmids in bacterial communities.

The observed prevalence of particular plasmid replicon families in the ECOR and SAR collections could be the result of a few widely distributed plasmids. Alternatively, plasmids within these collections could have highly variable genomes despite the similarity of the backbone genes used to classify them into replicon families. We compared RFLP patterns of large plasmids from ECOR and SAR to assess plasmid distribution and diversity. The analysis was done in two stages: an initial screen to identify potentially similar plasmids and an additional analysis to more accurately determine plasmid similarity. Based on this analysis, large plasmids of ECOR, SARA, SARB and SARC were highly variable. The majority of large plasmids were not similar to any other plasmid in the collections by RFLP analysis. Only one plasmid RFLP pattern was detected in more than four strains, all of which belonged to *S. enterica* serovar

Typhimurium. All other clusters were composed of four or fewer plasmids, demonstrating the overall lack of widely distributed plasmids.

Plasmids from the ECOR collection were more variable than those of the SAR collections. Only four clusters of similar *E. coli* plasmids were identified. Two were composed of pairs of identical plasmids harbored by strains from different ECOR subgroups. The different genotypes of the host strains combined with the identity of the plasmids suggest that recent horizontal transmission had occurred. A third cluster of two highly similar plasmids can be attributed to the isolation of two ECOR strains with identical MLEE profiles from the same host animal. It is likely that these two plasmids are descendants of the same lineage, however, minor differences in the RFLP patterns indicate that some plasmid evolution has occurred. This pair of plasmids may be a snapshot of the early stages of plasmid divergence in an otherwise clonal bacterial strain.

The fourth ECOR plasmid cluster included three plasmids: two from subgroup D and one from subgroup A. In a previous study, these plasmids were consistently grouped together by nucleotide sequence analyses of four plasmid genes (9). The plasmid RFLP comparison reported here suggests that these plasmids share extensive similarity along their genomes. Apart from the four clusters, all other ECOR plasmids appear unique by RFLP pattern cluster analysis. No widely distributed plasmid RFLP pattern was identified in ECOR, indicating that the prevalence of IncF and IncI1 replicon families is not due to the widespread distribution of a particular plasmid, but rather the occurrence of common plasmid backbones. It is interesting that the ECOR collection had significantly more multi-plasmid strains than the SAR collections. The presence of multiple plasmids within a single strain provides an opportunity for the exchange of accessory elements between

plasmids, and thereby for rapid plasmid evolution. The higher levels of plasmid variation in ECOR compared to SAR may be due in part to the higher frequency of multi-plasmid ECOR strains.

In contrast to the ECOR collection, 33 SAR large plasmids were grouped into nine clusters by RFLP analysis. Three serovar-specific clusters accounted for all identical and highly similar plasmids: one cluster of two plasmids from sv Muenchen, another of three plasmids from sv Heidelberg and a large cluster from sv Typhimurium. The three identical plasmids from sv Heidelberg were carried by strains isolated from different hosts in different U.S. states (Table 3.3). Although serovar Heidelberg is one of the most common serovars associated with *Salmonella* outbreaks (17), no serovar-specific virulence plasmid has been reported. If this plasmid is a virulence plasmid, the lack of a RepFIIA replicon suggests that it is not in the pSLT-like lineage. A previous study of serovar Heidelberg strains isolated recently from a range of hosts identified a large plasmid in all strains (1), which looks potentially similar by visual comparison of the *BglII* RFLP pattern with our data. It is possible that the serovar Heidelberg plasmid identified here is a widespread serovar-specific plasmid that has persisted in populations of serovar Heidelberg over time. Sequencing of this plasmid is essential to determine whether it encodes virulence loci or other ecologically significant phenotypes.

The large sv Typhimurium cluster included 13 SARA plasmids that clustered with the pSLT plasmid of SARA2 (also known as Typhimurium strain LT2). This cluster of pSLT-like plasmids was the only cluster with four or more plasmids in the ECOR and SAR collections. Contrary to the high variation observed for other ECOR and SAR plasmids, the pSLT-like plasmids showed very little variation in RFLP pattern despite

occurring in multiple strains of different genotypes isolated from a wide range of hosts and geographic locations (Figure 3.5 and Table 3.3). Comparative genomics and microarray studies of *Salmonella* virulence plasmids have shown that some pSLT genes are found on other plasmids (29, 44). However, it seems that once pSLT became serovar-specific, plasmid evolution was constrained in some way. It is unclear what led to this plasmid niche specialization or why the evolutionary dynamics of this virulence plasmid are different from other plasmids.

In addition to the three serovar-specific clusters, there were also six clusters of similar *Salmonella* plasmids. Five included plasmids from different subspecies I serovars, demonstrating possible horizontal transmission of plasmids. Variations in RFLP patterns indicate that although the plasmids may be descendants of the same lineage, plasmid evolution has occurred. None of the large plasmids from non-subspecies I *Salmonella* strains showed similarity to any subspecies I plasmids. The two subspecies VI strains in SARC had large plasmids that were similar, despite a size difference of ~30 kb (Table 3.3). This pair may represent a widespread plasmid lineage in subspecies VI. The two descendants detected in SARC appear to have diverged slightly from each other by gain or loss of genes.

The high variability in plasmid RFLP patterns illustrates the high degree of natural plasmid variation in large plasmids of the Enterobacteriaceae, even within plasmid replicon families. Other studies, which focused on enterobacterial plasmids conferring particular antibiotic resistances or from geographically restricted areas (21, 49), have also documented high variation in plasmid RFLP patterns. It is clear that plasmid evolution is a significant process shaping the mobile genetic element pool. One

question that remains is how the plasmid variation observed here leads to differences in plasmid phenotype, and therefore the ecological and physiological effects of particular plasmids. It is possible that localized nucleotide changes at the enzyme recognition site have resulted in variable RFLP patterns that do not actually reflect differences in gene content. However, evidence from the *in silico* RFLP analysis, especially the pair of Choleraesuis virulence plasmids that had diverged at 92 nucleotide positions but were 100% similar by RFLP, indicates that variation in RFLP patterns reflects real differences in plasmid gene content.

The data presented here support the conclusion that plasmids are not static entities, but rather appear to undergo rapid evolution. This idea has not yet been adequately addressed in evolutionary and ecological models of plasmid dynamics. The evolutionary dynamics of plasmids are intertwined with those of their host bacteria, but they are clearly also subject to a unique set of evolutionary pressures. Accessory element content, which is an important driver of plasmid variation, dramatically affects plasmid phenotype, but may not be a static feature of plasmid genomes. As a result, tracking a plasmid through a bacterial population based on components of the backbone may miss important changes in the adaptive and ecological effects of the plasmid arising from gain and loss of accessory DNA.

In addition to informing models of plasmid dynamics, this data is also an important contribution towards incorporating mobile genetic elements into population studies of these reference collections of Enterobacteriaceae. The information on plasmids reported here complements previous studies of other mobile genetic elements such as bacteriophage (48). Plasmids, phage and other mobile genetic elements constitute a

flexible gene arsenal that can have dramatic effects on bacterial evolution and adaptation, and therefore they are an important component of population-based studies.

TABLE 3.1. Plasmid prevalence in reference collections of natural *E. coli* and *Salmonella*.

Collection (# strains analyzed)		% of strains with plasmids			
		Any size		Large (≥30 kb)	
ECOR (71)	A (25)	89[a]	100	70.5[a]	80
	B1 (15)		60		53.5
	B2 (15)		87		40
	D (12)		100		100
SARA (72)		61[b]		51.5[b]	
SARB (71)		62[b]		51[b]	
SARC (14)		28.5[c]		21.5[c]	
SGSC (71)		51[b]		45[b]	

Data is also shown for the four phylogenetic subgroups of ECOR: A1, B1. B2 and D. Italicized letters indicate statistically significant differences ($p<0.01$) by exact binomial test.

TABLE 3.2. Plasmid replicon families in reference collections of natural *E. coli* and *Salmonella* and in NCBI GenBank.

Replicon family[a]		ECOR	SARA	SARB	SARC	NCBI[b]	
						E. coli	*Salmonella*
	F	33	5	0	1	25	1
FI	RepFIA	3	3	0	0	2	1
FI	RepFIB	19	3	0	0	8	2
FI	RepFIC	1	0	0	0	0	0
	FII	0	15	9	0	0	7
	I1	7	10	9	0	3	3
	Y	4	0	0	0	0	0
	A/C	0	2	2	0	0	1
	HI2	0	2	0	0	1	0
	N	2	1	1	0	3	1
	X	1	0	0	0	0	0
	HI1	0	0	0	0	0	3
	L/M	0	0	0	0	0	0
	W	0	0	0	0	0	2
	T	0	0	0	0	0	0
	Untypable	13	8	17	2	10	7

[a] IncP, IncK, and IncB/O primer pairs were omitted due to difficulties with false negatives and/or nonspecific amplification.

[b] *In silico* replicon typing results for 43 *E. coli* and 26 *Salmonella* plasmids available as complete sequences in GenBank as of January 2009.

TABLE 3.3. Summary of plasmid clusters by RFLP analysis.

RFLP type	Clusters[a, b]		Similarity by RFLP	Replicon families in strain	Plasmid profile of strain	Strain origin[d] Host, location, year	Subgroup or serovar
	Plasmid[c]	Size					
E1	pECOR43	94	100%	F	94, 12, 7, 5, 4	Human, Sweden, -	other[e]
	pECOR10	93		F	93, 9, 7, 5, 4	Human, U.S. (NY), -	A
E2	pECOR72-1	92	100%	F	92, 78, 6	Human, Sweden, -	B1
	pECOR24-1	90		F, FIB	90, 74, 6, 3, 3, 2, 2	Human, Sweden, -	A
E3	pECOR20-1	119	93%	Untypable	119, 34, 6, 3, 3	Steer, Bali, -	A
	pECOR21	119		Untypable	119, 8, 5, 3		A
E4	pECOR11	105	83%	F, FIB	105, 7, 7, 2	Human, Sweden, -	A
	pECOR49	128	64%	F, FIB	128, 2	Human, Sweden, -	D
	pECOR50	132		F, FIB	132, 6, 5, 3, 2	Human, Sweden, -	D
S1	pSARA65	131	100%	F, FIA, FIB	131, 10	Chicken, U.S. (FL), 1987	Muenchen
	pSARA66	143		F, FIA, FIB	143, 25, 10, 5	Human, U.S. (MA), -	Muenchen
S2	pSARA30	41	100%	Untypable	41	Chicken, U.S. (PA), 1987	Heidelberg
	pSARA31	41		Untypable	41	Swine, U.S. (MD), 1987	Heidelberg
	pSARA32	41		Untypable	41	Dog, U.S. (TX), 1986	Heidelberg
S3	pSARA4	91	100% / 98% / 58%	FIIA	91, 9, 6	Rabbit, U.S. (IN), 1986	Typhimurium
	pSARA11	95		FIIA	95	-, Thailand, -	Typhimurium
	pSARA12	102		FIIA	102, 4	Horse, U.S. (LA), 1987	Typhimurium
	pSARA1	108		FIIA	108	Human, Mexico, -	Typhimurium
	pSARA2	105		FIIA	105	-, Lab strain LT2, -	Typhimurium
	pSARA9-2	96		FIIA, I1	124, 96	Parrot, U.S. (CA), 1987	Typhimurium
	pSARA13	99		FIIA	99, 4	-, France, -	Typhimurium
	pSARA19	93		FIIA	93, 5, 3	Human, Mexico, -	Typhimurium
	pSARA20	94		FIIA, F, A/C	94, 4	-, France, -	Typhimurium
	pSARA21	87		FIIA	87	Heron, U.S. (OR), -	Typhimurium
	pSARA5-1	93	100% / 72%	FIIA	93, 65, 10	-, Mongolia, -	Typhimurium
	pSARA14	100		FIIA	100	-, Panama, -	Typhimurium
	pSARA15	99		FIIA	99	Dog, U.S. (TX), 1987	Typhimurium

S4	pSARA33-2 pSARB25-1 pSARB19 pSARB20-1	92 96 103 100	80% 73% 57%	A/C, I1 I1 I1 I1	137, 92, 3 96, 32 103, 4 100, 24	Human, Mexico, -	Heidelberg Indiana Enteritidis Emek
S5	pSARC13-2 pSARC14	73 41	75%	Untypable Untypable	123, 73, 23 41		subsp. VI subsp. VI
S6	pSARA9-1 pSARA10 pSARA50	124 105 95	58% 51%	FIIA, I1 FIIA, I1 I1	124, 96 105 95	Parrot, U.S. (CA), 1987 Opossum, U.S. (CA), 1987 Food, Middle East, 1976	Typhimurium Typhimurium Paratyphi B
S7	pSARB6 pSARB69	51 59	57%	FIIA FIIA	51 59		Choleraesuis Typhisuis
S8	pSARA22 pSARB11	88 96	50%	I1 A/C, I1	88 96	Human, U.S. (MA), -	Saintpaul Derby
S9	pSARB8 pSARB70	83 83	50%	I1 I1	83 83		Decatur Typhisuis

[a] Plasmids from the SARB strains of the SARA/SARB counterparts are not included.

[b] The 102 kb plasmid from SARA6 and the 30 kb and 32 kb plasmids from SARB3 and SARB4, respectively, did not produce clear second stage digests and therefore are not included.

[c] Plasmids are identified by host strain name. If a strain has multiple large plasmids, they are numbered from largest to smallest, and this number is added to the strain name after a hyphen to identify individual plasmids.

[d] - information not available.

[e] Strain was not assigned to any of the four recognized ECOR phylogenetic subgroups.

TABLE 3.4. Summary of sequenced plasmid clusters by *in silico* RFLP analysis.

Clusters			Similarity by RFLP	Replicon family	Source
Plasmid	Accession number	Size (kb)			
pMAR7	NC_010862	102	76%	F	*E. coli*
pMAR2	NC_011603	98		F	*E. coli*
pO157	NC_011350	95	76%	F	*E. coli* O157:H7 strain EC4115
pO157	NC_002128	93	62%	F	*E. coli* O157:H7 strain Sakai
pO157	NC_007414	92		F	*E. coli* O157:H7 strain EDL933
pLEW517	NC_009131	65	66%	N	*E. coli* (transconjugant)
pLEW517	NC_009132	64		N	*E. coli* (natural host)
pKDSC50	NC_002638	50	100%	FII	*S. enterica* sv Choleraesuis
pSCV50	NC_006855	50		FII	*S. enterica* sv Choleraesuis
pCT02021853_74	NC_011204	75	80%	FII	*S. enterica* sv Dublin
pOU1115	NC_010422	75		FII	*S. enterica* sv Dublin

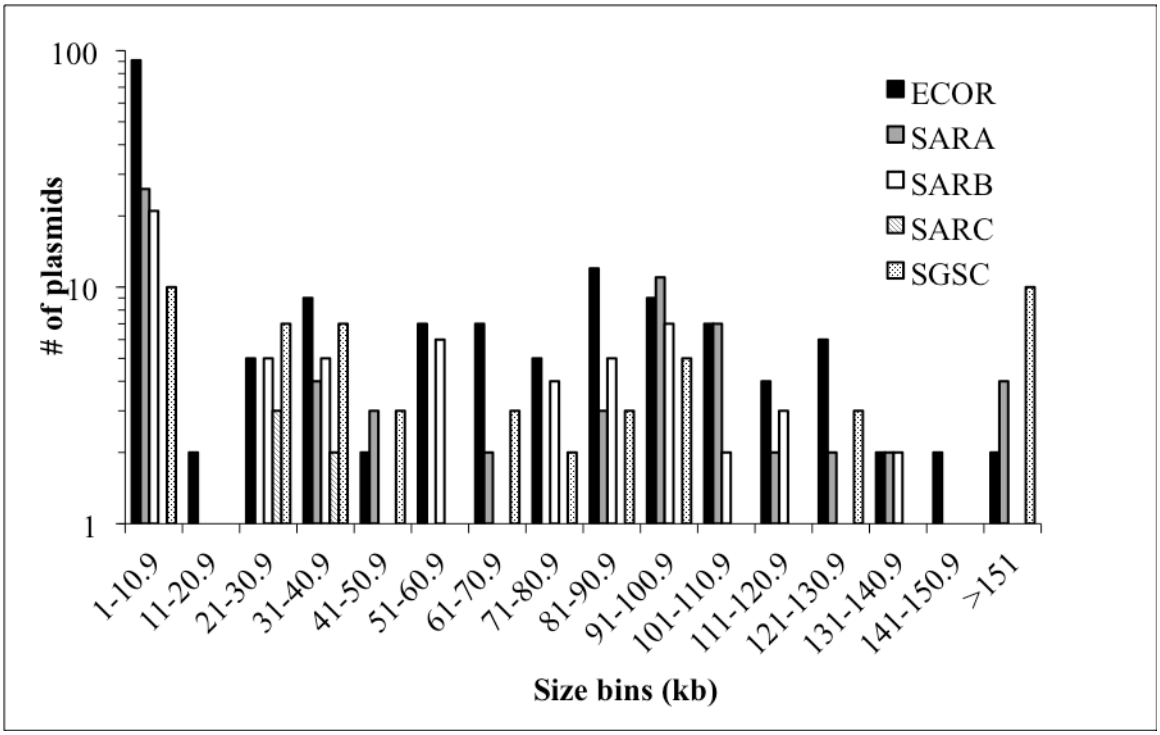
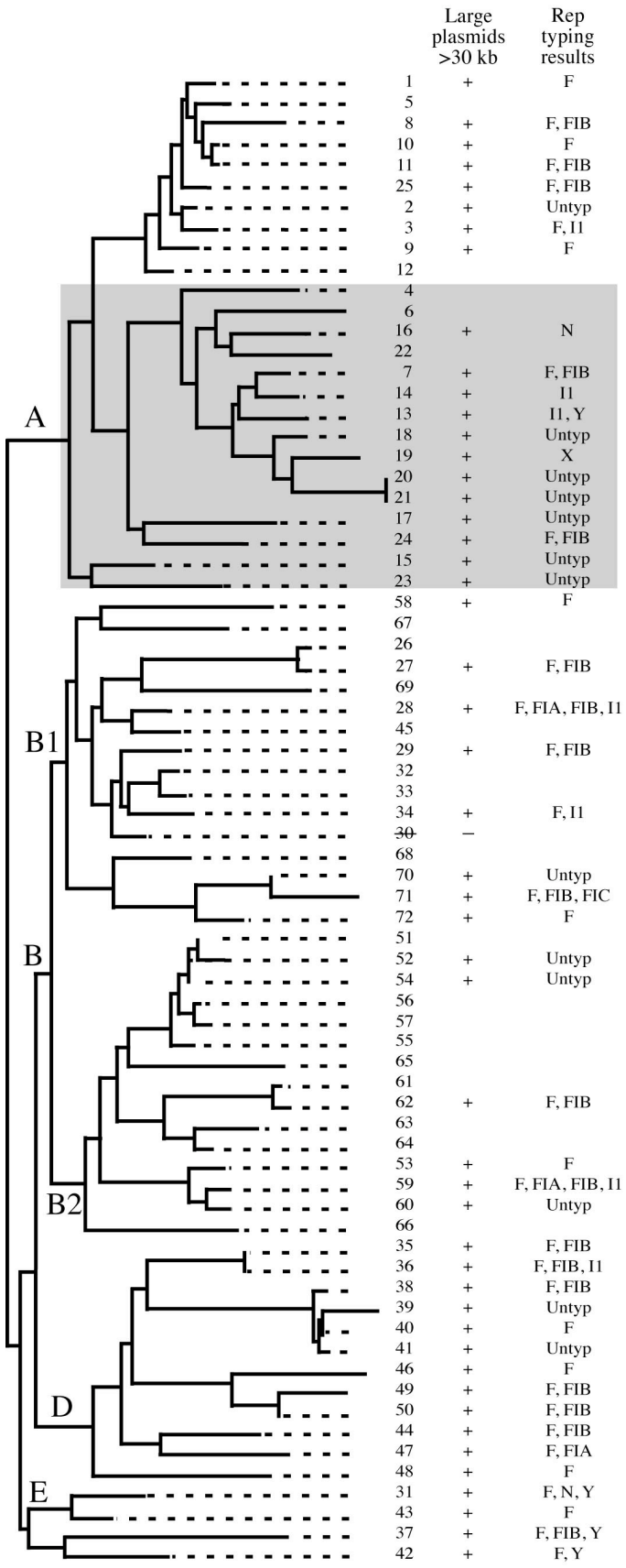


FIGURE 3.1. Distribution of ECOR and SAR plasmids by size. For strains with multiple plasmids <20 kb, RFLP analysis identified the supercoiled forms of small plasmids. To ensure sizing accuracy, plasmid sizes were estimated by comparison to three standard curves corresponding to three different size ranges. The y axis is a logarithmic scale.

FIGURE 3.2. Distribution of plasmid replicon families in ECOR. The presence of large (≥30 kb) plasmids and the plasmid replicon typing results were overlaid on a dendrogram of the ECOR strains constructed using MLEE profiles (adapted from reference 28). ECOR30 was omitted from the analysis. The two clades of subgroup A strains lacking large IncF plasmids are highlighted in grey.



	Large plasmids >30 kb	Rep typing results
1	+	F
5		
8	+	F, FIB
10	+	F
11	+	F, FIB
25	+	F, FIB
2	+	Untyp
3	+	F, I1
9	+	F
12		
4		
6		
16	+	N
22		
7	+	F, FIB
14	+	I1
13	+	I1, Y
18	+	Untyp
19	+	X
20	+	Untyp
21	+	Untyp
17	+	Untyp
24	+	F, FIB
15	+	Untyp
23	+	Untyp
58	+	F
67		
26		
27	+	F, FIB
69		
28	+	F, FIA, FIB, I1
45		
29	+	F, FIB
32		
33		
34	+	F, I1
30	–	
68		
70	+	Untyp
71	+	F, FIB, FIC
72	+	F
51		
52	+	Untyp
54	+	Untyp
56		
57		
55		
65		
61		
62	+	F, FIB
63		
64		
53	+	F
59	+	F, FIA, FIB, I1
60	+	Untyp
66		
35	+	F, FIB
36	+	F, FIB, I1
38	+	F, FIB
39	+	Untyp
40	+	F
41	+	Untyp
46	+	F
49	+	F, FIB
50	+	F, FIB
44	+	F, FIB
47	+	F, FIA
48	+	F
31	+	F, N, Y
43	+	F
37	+	F, FIB, Y
42	+	F, Y

FIGURE 3.3. Dendrogram of ECOR large plasmid *AccI* RFLP patterns. Scale indicates % similarity. The vertical line denotes 60% similarity. Plasmids are named after ECOR host strains. For strains with multiple large plasmids, the plasmids are numbered from largest to smallest, and this number is added to the plasmid name after a hyphen.

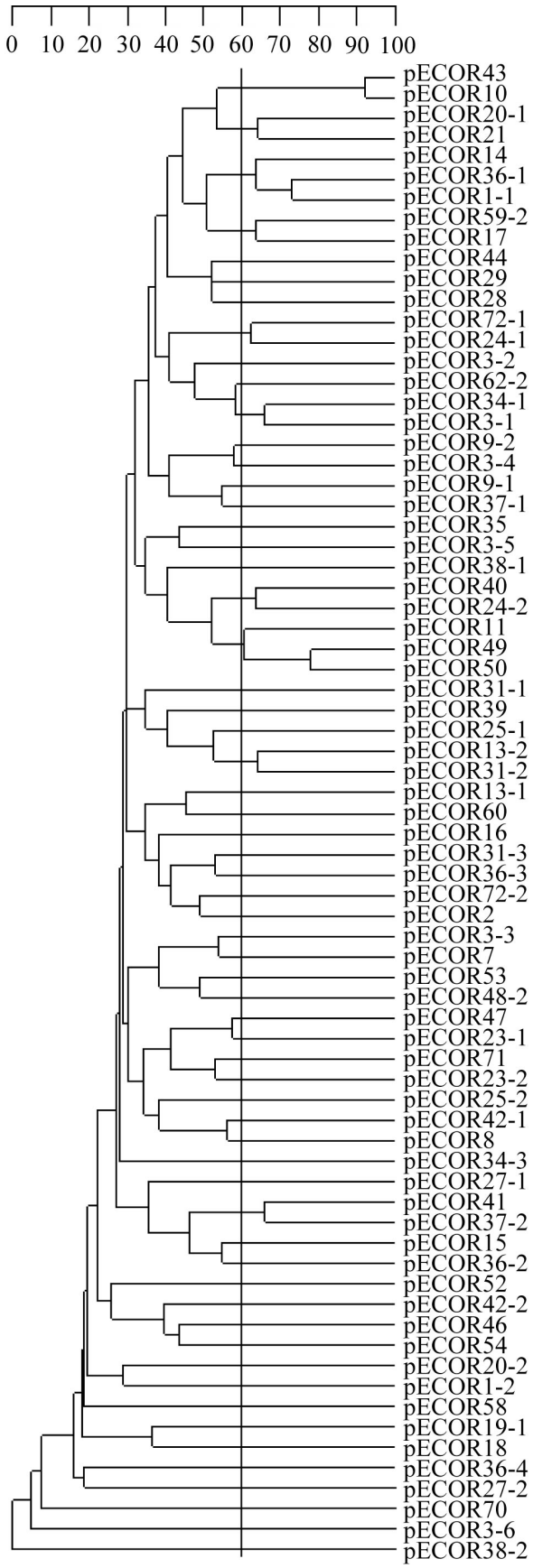
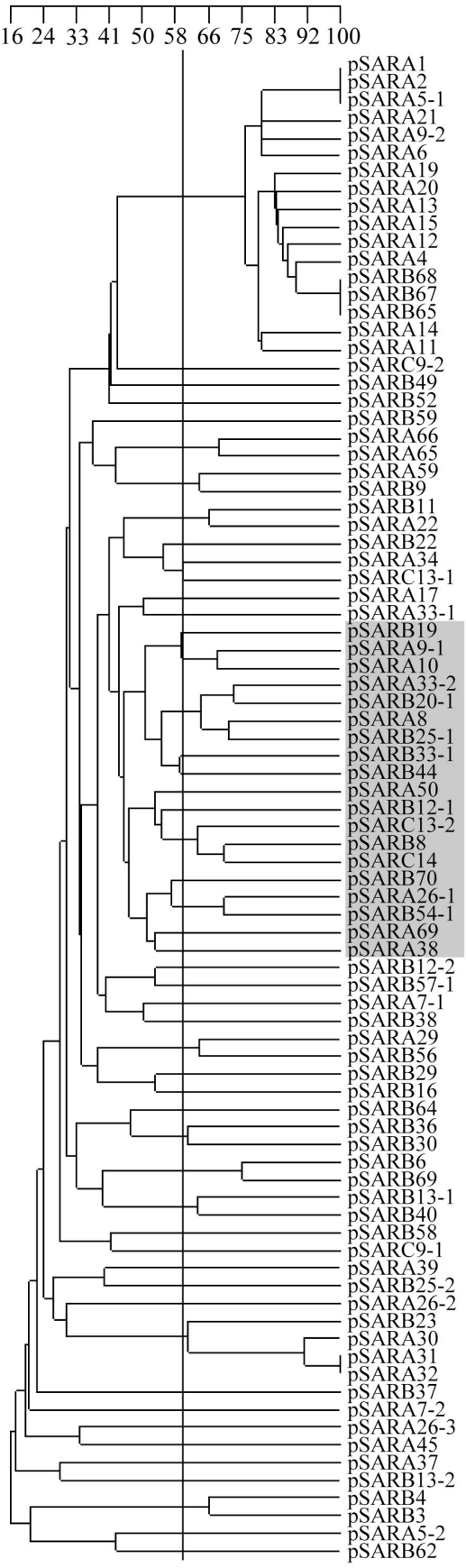

0 10 20 30 40 50 60 70 80 90 100
pECOR43
pECOR10
pECOR20-1
pECOR21
pECOR14
pECOR36-1
pECOR1-1
pECOR59-2
pECOR17
pECOR44
pECOR29
pECOR28
pECOR72-1
pECOR24-1
pECOR3-2
pECOR62-2
pECOR34-1
pECOR3-1
pECOR9-2
pECOR3-4
pECOR9-1
pECOR37-1
pECOR35
pECOR3-5
pECOR38-1
pECOR40
pECOR24-2
pECOR11
pECOR49
pECOR50
pECOR31-1
pECOR39
pECOR25-1
pECOR13-2
pECOR31-2
pECOR13-1
pECOR60
pECOR16
pECOR31-3
pECOR36-3
pECOR72-2
pECOR2
pECOR3-3
pECOR7
pECOR53
pECOR48-2
pECOR47
pECOR23-1
pECOR71
pECOR23-2
pECOR25-2
pECOR42-1
pECOR8
pECOR34-3
pECOR27-1
pECOR41
pECOR37-2
pECOR15
pECOR36-2
pECOR52
pECOR42-2
pECOR46
pECOR54
pECOR20-2
pECOR1-2
pECOR58
pECOR19-1
pECOR18
pECOR36-4
pECOR27-2
pECOR70
pECOR3-6
pECOR38-2

FIGURE 3.4. Dendrogram of SAR large plasmid *AccI* RFLP patterns. Scale indicates % similarity. The vertical line denotes 60% similarity. Plasmids are named after SAR host strains. For strains with multiple large plasmids, the plasmids are numbered from largest to smallest, and this number is added to the plasmid name after a hyphen. The highlighted plasmids are the group of 19 plasmids for which *AccI* digestion generated too many fragments (see text).


16 24 33 41 50 58 66 75 83 92 100
pSARA1
pSARA2
pSARA5-1
pSARA21
pSARA9-2
pSARA6
pSARA19
pSARA20
pSARA13
pSARA15
pSARA12
pSARA4
pSARB68
pSARB67
pSARB65
pSARA14
pSARA11
pSARC9-2
pSARB49
pSARB52
pSARB59
pSARA66
pSARA65
pSARA59
pSARB9
pSARB11
pSARA22
pSARB22
pSARA34
pSARC13-1
pSARA17
pSARA33-1
pSARB19
pSARA9-1
pSARA10
pSARA33-2
pSARB20-1
pSARA8
pSARB25-1
pSARB33-1
pSARB44
pSARA50
pSARB12-1
pSARC13-2
pSARB8
pSARC14
pSARB70
pSARA26-1
pSARB54-1
pSARA69
pSARA38
pSARB12-2
pSARB57-1
pSARA7-1
pSARB38
pSARA29
pSARB56
pSARB29
pSARB16
pSARB64
pSARB36
pSARB30
pSARB6
pSARB69
pSARB13-1
pSARB40
pSARB58
pSARC9-1
pSARA39
pSARB25-2
pSARA26-2
pSARB23
pSARA30
pSARA31
pSARA32
pSARB37
pSARA7-2
pSARA26-3
pSARA45
pSARA37
pSARB13-2
pSARB4
pSARB3
pSARA5-2
pSARB62

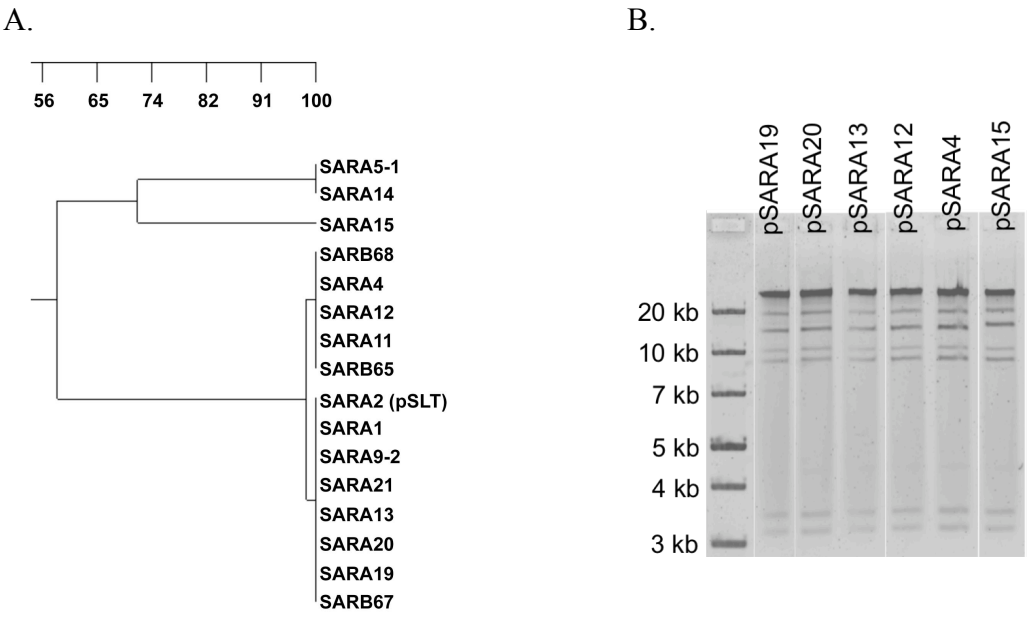


FIGURE 3.5. Cluster analysis of pSLT-like plasmids. A) Dendrogram of *BglII* RFLP patterns of the pSLT-like cluster. pSARA2 is the pSLT plasmid. The three SARB plasmids are replicates of SARA plasmids due to the inclusion of the same strain in both SARA and SARB. The scale indicates % similarity. We had difficulty isolating pSARA6 for BglII analysis, so this plasmid is not included in this dendrogram. B) Gel image of *BglII* RFLP patterns of selected pSLT-like plasmids.

REFERENCES

1. **Andrysiak, A. K., A. B. Olson, D. M. Tracz, K. Dore, R. Irwin, L. K. Ng, M. W. Gilmour, and C. I. Program.** 2008. Genetic characterization of clinical and agri-food isolates of multi drug resistant *Salmonella enterica* serovar Heidelberg from Canada. BMC Microbiology **8:**-.
2. **Baumler, A. J., R. M. Tsolis, T. A. Ficht, and L. G. Adams.** 1998. Evolution of host adaptation in *Salmonella enterica*. Infection and Immunity **66:**4579-4587.
3. **Beltran, P., S. A. Plock, N. H. Smith, T. S. Whittam, D. C. Old, and R. K. Selander.** 1991. Reference collection of strains of the *Salmonella*-Typhimurium complex from natural populations. Journal of General Microbiology **137:**601-606.
4. **Bingen, E., B. Picard, N. Brahimi, S. Mathy, P. Desjardins, J. Elion, and E. Denamur.** 1998. Phylogenetic analysis of *Escherichia coli* strains causing neonatal meningitis suggests horizontal gene transfer from a predominant pool of highly virulent B2 group strains. Journal of Infectious Diseases **177:**642-650.
5. **Bouma, J. E., and R. E. Lenski.** 1988. Evolution of a bacteria plasmid association. Nature **335:**351-352.
6. **Boyd, E. F., and D. L. Hartl.** 1998. Chromosomal regions specific to pathogenic isolates of *Escherichia coli* have a phylogenetically clustered distribution. Journal of Bacteriology **180:**1159-1165.
7. **Boyd, E. F., and D. L. Hartl.** 1997. Recent horizontal transmission of plasmids between natural populations of *Escherichia coli* and *Salmonella enterica*. Journal of Bacteriology **179:**1622-1627.

8. **Boyd, E. F., and D. L. Hartl.** 1998. *Salmonella* virulence plasmid: Modular acquisition of the *spv* virulence region by an F-plasmid in *Salmonella enterica* subspecies I and insertion into the chromosome of subspecies II, IIIa, IV and VII isolates. Genetics **149:**1183-1190.

9. **Boyd, E. F., C. W. Hill, S. M. Rich, and D. L. Hartl.** 1996. Mosaic structure of plasmids from natural populations of *Escherichia coli*. Genetics **143:**1091-1100.

10. **Boyd, E. F., F. S. Wang, P. Beltran, S. A. Plock, K. Nelson, and R. K. Selander.** 1993. *Salmonella* reference collection-B (Sarb) - strains of 37 serovars of subspecies I. Journal of General Microbiology **139:**1125-1132.

11. **Boyd, E. F., F. S. Wang, T. S. Whittam, and R. K. Selander.** 1996. Molecular genetic relationships of the salmonellae. Applied and Environmental Microbiology **62:**804-808.

12. **Brinkley, C., V. Burland, R. Keller, D. J. Rose, A. T. Boutin, S. A. Klink, F. R. Blattner, and J. B. Kaper.** 2006. Nucleotide sequence analysis of the enteropathogenic *Escherichia coli* adherence factor plasmid pMAR7. Infection and Immunity **74:**5408-5413.

13. **Brunder, W., H. Karch, and H. Schmidt.** 2006. Complete sequence of the large virulence plasmid pSFO157 of the sorbitol-fermenting enterohemorrhagic *Escherichia coli* O157 : H- strain 3072/96. International Journal of Medical Microbiology **296:**467-474.

14. **Burland, V., Y. Shao, N. T. Perna, G. Plunkett, H. J. Sofia, and F. R. Blattner.** 1998. The complete DNA sequence and analysis of the large virulence plasmid of *Escherichia coli* O157 : H7. Nucleic Acids Research **26:**4196-4204.

15. **Canton, R., and T. M. Coque.** 2006. The CTX-M beta-lactamase pandemic. Current Opinion in Microbiology **9:**466-475.

16. **Carattoli, A., A. Bertini, L. Villa, V. Falbo, K. L. Hopkins, and E. J. Threlfall.** 2005. Identification of plasmids by PCR-based replicon typing. Journal of Microbiological Methods **63:**219-228.

17. **CDC.** 2008. *Salmonella* surveillance: annual summary, 2006. U.S. Department of Heath and Human Services, Atlanta, GA.

18. **Chu, C. S., Y. Feng, A. C. Chien, S. N. Hu, C. H. Chu, and C. H. Chiu.** 2008. Evolution of genes on the *Salmonella* virulence plasmid phylogeny revealed from sequencing of the virulence plasmids of *S. enterica* serotype Dublin and comparative analysis. Genomics **92:**339-343.

19. **Couturier, M., F. Bex, P. L. Bergquist, and W. K. Maas.** 1988. Identification and classification of bacterial plasmids. Microbiological Reviews **52:**375-395.

20. **Dahlberg, C., and L. Chao.** 2003. Amelioration of the cost of conjugative plasmid carriage in Eschericha coli K12. Genetics **165:**1641-1649.

21. **Daniels, J. B., D. R. Call, and T. E. Besser.** 2007. Molecular epidemiology of *bla*(CMY-2) plasmids carried by *Salmonella enterica* and *Escherichia coli* isolates from cattle in the pacific northwest. Applied and Environmental Microbiology **73:**8005-8011.

22. **Datta, N., and R. W. Hedges.** 1971. Compatibility groups among Fi-R factors. Nature **234:**222-&.

23. **Escobar-Paramo, P., O. Clermont, A. B. Blanc-Potard, H. Bui, C. Le Bouguenec, and E. Denamur.** 2004. A specific genetic background is required

for acquisition and expression of virulence factors in *Escherichia coli*. Molecular Biology and Evolution **21:**1085-1094.

24. **Froehlich, B., J. Parkhill, M. Sanders, M. A. Quail, and J. R. Scott.** 2005. The pCoo plasmid of enterotoxigenic *Escherichia coli* is a mosaic cointegrate. Journal of Bacteriology **187:**6509-6516.

25. **Frost, L. S., R. Leplae, A. O. Summers, and A. Toussaint.** 2005. Mobile genetic elements: The agents of open source evolution. Nature Reviews Microbiology **3:**722-732.

26. **Gilmour, M. W., N. R. Thomson, M. Sanders, J. Parkhill, and D. E. Taylor.** 2004. The complete nucleotide sequence of the resistance plasmid R478: defining the backbone components of incompatibility group H conjugative plasmids through comparative genomics. Plasmid **52:**182-202.

27. **Groisman, E. A., and H. Ochman.** 1997. How *Salmonella* became a pathogen. Trends in Microbiology **5:**343-349.

28. **Herzer, P. J., S. Inouye, M. Inouye, and T. S. Whittam.** 1990. Phylogenetic distribution of branched RNA-linked multicopy single-stranded-DNA among natural isolates of *Escherichia coli*. Journal of Bacteriology **172:**6175-6181.

29. **Hong, S. F., C. H. Chiu, C. S. Chu, Y. Feng, and J. T. Ou.** 2008. Complete nucleotide sequence of a virulence plasmid of *Salmonella enterica* serovar Dublin and its phylogenetic relationship to the virulence plasmids of serovars Choleraesuis, Enteritidis and Typhimurium. Fems Microbiology Letters **282:**39-43.

30. **Hopkins, K. L., E. Liebana, L. Villa, M. Batchelor, E. J. Threlfall, and A. Carattoli.** 2006. Replicon typing of plasmids carrying CTX-M or CMY beta-lactamases circulating among *Salmonella* and *Escherichia coli* isolates. Antimicrobial Agents and Chemotherapy **50:**3203-3206.

31. **Houndt, T., and H. Ochman.** 2000. Long-term shifts in patterns of antibiotic resistance in enteric bacteria. Applied and Environmental Microbiology **66:**5406-5409.

32. **Hughes, V. M., and N. Datta.** 1983. Conjugative plasmids in bacteria of the pre-antibiotic era. Nature **302:**725-726.

33. **Iguchi, A., N. R. Thomson, Y. Ogura, D. Saunders, T. Ooka, I. R. Henderson, D. Harris, M. Asadulghani, K. Kurokawa, P. Dean, B. Kenny, M. A. Quail, S. Thurston, G. Dougan, T. Hayashi, J. Parkhill, and G. Frankel.** 2009. Complete genome sequence and comparative genome analysis of enteropathogenic *Escherichia coli* O127:H6 strain E2348/69. Journal of Bacteriology **191:**347-354.

34. **Johnson, T. J., Y. M. Wannemuehler, S. J. Johnson, C. M. Logue, D. G. White, C. Doetkott, and L. K. Nolan.** 2007. Plasmid replicon typing of commensal and pathogenic *Escherichia coli* isolates. Applied and Environmental Microbiology **73:**1976-1983.

35. **Lecointre, G., L. Rachdi, P. Darlu, and E. Denamur.** 1998. *Escherichia coli* molecular phylogeny using the incongruence length difference test. Molecular Biology and Evolution **15:**1685-1695.

36. **LeMinor, L., and M. Y. Popoff.** 1987. Designation of *Salmonella enterica* Sp-Nov, Nom Rev, as the type and only species of the genus *Salmonella*. International Journal of Systematic Bacteriology **37:**465-468.

37. **Makino, K., K. Ishii, T. Yasunaga, M. Hattori, K. Yokoyama, C. H. Yutsudo, Y. Kubota, Y. Yamaichi, T. Iida, K. Yamamoto, T. Honda, C.-G. Han, E. Ohtsubo, M. Kasamatsu, T. Hayashi, S. Kuhara, and H. Shinagawa.** 1998. Complete nucleotide sequences of 93-kb and 3.3-kb plasmids of an enterohemorrhagic *Escherichia coli* O157:H7 derived from Sakai outbreak. DNA Res **5:**1-9.

38. **Marcade, G., C. Deschamps, A. Boyd, V. Gautier, B. Picard, C. Branger, E. Denamur, and G. Arlet.** 2009. Replicon typing of plasmids in *Escherichia coli* producing extended-spectrum beta-lactamases. Journal of Antimicrobial Chemotherapy **63:**67-71.

39. **Mazel, D., B. Dychinco, V. A. Webb, and J. Davies.** 2000. Antibiotic resistance in the ECOR collection: Integrons and identification of a novel *aad* gene. Antimicrobial Agents and Chemotherapy **44:**1568-1574.

40. **McQuiston, J. R., S. Herrera-Leon, B. C. Wertheim, J. Doyle, P. I. Fields, R. V. Tauxe, and J. M. Logsdon.** 2008. Molecular phylogeny of the salmonellae: Relationships among *Salmonella* species and subspecies determined from four housekeeping genes and evidence of lateral gene transfer events. Journal of Bacteriology **190:**7060-7067.

41. **Mrazek, J., and S. H. Xie.** 2006. Pattern locator: a new tool for finding local sequence patterns in genomic DNA sequences. Bioinformatics **22:**3099-3100.

42. **Novick, R. P.** 1987. Plasmid incompatibility. Microbiological Reviews **51:**381-395.

43. **Ochman, H., and R. K. Selander.** 1984. Standard reference strains of *Escherichia coli* from natural populations. Journal of Bacteriology **157:**690-693.

44. **Porwollik, S., E. F. Boyd, C. Choy, P. Cheng, L. Florea, E. Proctor, and M. McClelland.** 2004. Characterization of *Salmonella enterica* subspecies I genovars by use of microarrays. Journal of Bacteriology **186:**5883-5898.

45. **Riley, M. A., and D. M. Gordon.** 1992. A survey of Col plasmids in natural isolates of *Escherichia coli* and an investigation into the stability of Col plasmid lineages. Journal of General Microbiology **138:**1345-1352.

46. **Rychlik, I., D. Gregorova, and H. Hradecka.** 2006. Distribution and function of plasmids in *Salmonella enterica*. Veterinary Microbiology **112:**1-10.

47. **Sawyer, S. A., D. E. Dykhuizen, R. F. Dubose, L. Green, T. Mutangaduramhlanga, D. F. Wolczyk, and D. L. Hartl.** 1987. Distribution and abundance of insertion sequences among natural isolates of *Escherichia coli*. Genetics **115:**51-63.

48. **Schicklmaier, P., E. Moser, T. Wieland, W. Rabsch, and H. Schmieger.** 1998. A comparative study on the frequency of prophages among natural isolates of *Salmonella* and *Escherichia coli* with emphasis on generalized transducers. Antonie Van Leeuwenhoek International Journal of General and Molecular Microbiology **73:**49-54.

49. **Sherley, M., D. M. Gordon, and P. J. Collignon.** 2004. Evolution of multi-resistance plasmids in Australian clinical isolates of *Escherichia coli*. Microbiology **150:**1539-1546.

50. **Sherley, M., D. M. Gordon, and P. J. Collignon.** 2003. Species differences in plasmid carriage in the Enterobacteriaceae. Plasmid **49:**79-85.

51. **Sorensen, S. J., M. Bailey, L. H. Hansen, N. Kroer, and S. Wuertz.** 2005. Studying plasmid horizontal transfer *in situ*: A critical review. Nature Reviews Microbiology **3:**700-710.

52. **Su, L. H., C. Chu, A. Cloeckaert, and C. H. Chiu.** 2008. An epidemic of plasmids? Dissemination of extended-spectrum cephalosporinases among *Salmonella* and other Enterobacteriaceae. Fems Immunology and Medical Microbiology **52:**155-168.

53. **Thomas, C. M.** 2000. Paradigms of plasmid organization. Molecular Microbiology **37:**485-491.

54. **Tindall, B. J., P. A. D. Grimont, G. M. Garrity, and J. P. Euzeby.** 2005. Nomenclature and taxonomy of the genus *Salmonella*. International Journal of Systematic and Evolutionary Microbiology **55:**521-524.

55. **Top, E. M., Y. Moenne-Loccoz, T. Pembroke, and C. M. Thomas.** 2000. Phenotypic traits conferred by plasmids, p. 249-285. *In* C. M. Thomas (ed.), The horizontal gene pool: bacterial plasmids and gene spread. Harwood Academic Publishers.

56. **Williams, L. E., C. Detter, K. Barry, A. Lapidus, and A. O. Summers.** 2006. Facile recovery of individual high-molecular-weight, low-copy-number natural

plasmids for genomic sequencing. Applied and Environmental Microbiology **72:**4899-4906.

57. **Wirth, T., D. Falush, R. T. Lan, F. Colles, P. Mensa, L. H. Wieler, H. Karch, P. R. Reeves, M. C. J. Maiden, H. Ochman, and M. Achtman.** 2006. Sex and virulence in *Escherichia coli*: an evolutionary perspective. Molecular Microbiology **60:**1136-1151.

58. **Yu, H., J. B. Wang, J. H. Ye, P. Tang, C. S. Chu, S. N. Hu, and C. H. Chiu.** 2006. Complete nucleotide sequence of pSCV50, the virulence plasmid of *Salmonella enterica* serovar Choleraesuis SC-B67. Plasmid **55:**145-151.

CHAPTER 4

CONCLUSIONS

Since their discovery in the 1950s, plasmids have proved fascinating research subjects. From the initial report of chromosomal DNA transfer by the F plasmid to the recent dramatic increase in bacterial antibiotic resistance, it is unquestionably clear that plasmids play significant roles in the adaptation and evolution of their host bacteria. Plasmid-borne accessory genes, which encompass a diverse assortment of antibiotic and metal resistances, virulence factors and metabolic functions, constitute a flexible gene arsenal providing host bacteria with extraordinary adaptive potential. The impact of plasmids is wide-ranging, from the virulence plasmids that confer pathogenicity on *Bacillus anthracis* to the pSym plasmids that facilitate the symbiotic relationship between rhizobia and leguminous plants.

Despite their important roles in adaptation and evolution, plasmids have primarily been studied either from a mechanistic perspective, for instance as a model system for chromosome segregation (6), or from a biotechnological perspective, such as in the development of cloning vectors for sequencing and gene expression. Less attention is paid to the ecological and evolutionary dynamics of plasmids. Questions remain regarding the distribution and diversity of plasmids in natural bacterial communities. How variable are these components of the horizontal gene pool? What boundaries limit their spread? How do they evolve? The overall objective of my dissertation research was to investigate natural plasmid variation in an attempt to address these questions and

contribute to the body of knowledge regarding plasmid ecology and evolution. To achieve this objective, I used approaches at two different scales: a sequence-based approach described in Chapter 2 and a population-based approach described in Chapter 3. In the following chapter, I will present an update on the sequence analysis of Chapter 2, discuss the findings of both chapters in the broad context of plasmid ecological and evolutionary dynamics, and propose directions for future research.

UPDATE OF SEQUENCE ANALYSIS FROM CHAPTER TWO

Chapter 2 described the adaptation of a magnetic bead-based BAC preparation method for the isolation of large and small plasmids from Gram-negative and -positive bacteria. The method was used to prepare plasmid DNA from natural strains of *Escherichia coli*, *Staphylococcus* and *Corynebacterium* for whole genome shotgun sequencing, which yielded four novel plasmid sequences. Of these sequences, the *E. coli* plasmid pLEW517 had the most coverage in terms of hits to sequences in the NCBI databases at the time of analysis. pLEW517 had very high nucleotide sequence identity in replication, maintenance and transfer regions to the IncN plasmid R46 from *Salmonella*, but the accessory content of these two plasmids differed.

Since the original analysis presented in Chapter 2, five more plasmid sequences were deposited that also have high nucleotide sequence identity to the backbone of R46. The plasmids were isolated from *Klebsiella pneumoniae* (n=3), *Salmonella enterica* serovar Dublin (n=1) and *E. coli* (n=1). BLASTN analysis comparing the plasmids to R46 divided them into two groups based on very slight divergence in the backbone genes. All three *Klebsiella* plasmids were 97-98% identical to the R46 backbone, whereas the two *E. coli* plasmids, including pLEW517, and the *Salmonella* plasmid had ≥99%

nucleotide identity to the R46 backbone. It is tempting to speculate that these sequences have captured the early stages of "plasmid speciation" or the establishment of two distinct branches of this IncN lineage. Perhaps an ancestor of these IncN plasmids was transferred to a *Klebsiella* strain and then circulated among *Klebsiella*, undergoing evolution apart from the plasmids found in *E. coli* and *Salmonella*. But it is equally likely that this IncN plasmid backbone has been transferred frequently among enterobacteria, and additional sequencing may uncover more divergent IncN plasmid backbones in *E. coli* and *Salmonella*.

The accessory content of the five new IncN plasmids is highly variable despite their nearly identical backbones, which is consistent with examples from the IncW and IncHI plasmid families (7, 16). Overall, five of the seven IncN plasmids encode a Class 1 integron, based on the presence of *intI1*. However, with the exception of the aminoglycoside resistance gene *aadA1*, the integron cassettes are different, emphasizing the remarkable phenotypic variability afforded by the integron platform. Other similarities in accessory content can be attributed to separate acquisition events rather than inheritance as a conserved region of the ancestral IncN plasmid. In particular, a *Salmonella* plasmid and a *Klebsiella* plasmid encoded tetracycline resistance genes also found on R46, but they were only 73% identical to the R46 genes. Therefore, they were probably acquired in separate recombination events. Similarly, a *Klebsiella* plasmid encoded an arsenic resistance operon that was 99% identical to that of R46. Because this operon is absent in all other IncN plasmids, it was most likely acquired in separate events. Analysis of the new IncN plasmid sequences confirmed the conclusion from the

original analysis that although backbone genes are conserved within this plasmid lineage, accessory elements are not.

In addition to providing a more nuanced picture of IncN plasmid evolution, this set of plasmids offers an example of plasmids' astonishing peripatetic nature. The host strains of these IncN plasmids were isolated from a wide range of geographic locations and hosts; for example, one of the *E. coli* host strains was isolated from a Spanish pig (8), the other was isolated from a laboratory primate during a longitudinal mercury exposure study (20) and two of the *Klebsiella* host strains were clinical isolates from hospitals in New York City (9). The travels of this IncN plasmid backbone reinforce the concept that a single mobile genetic element may be accessible to a wide range of bacterial communities in ecosystems that we may mistakenly regard as separate and clearly bounded.

Overall, the additional IncN plasmid sequences deposited since the analysis reported in Chapter 2 show that this plasmid lineage is evolving in two major ways: by divergence of backbone genes due to point mutations and small indels and by gain and loss of accessory elements, which dramatically impacts each plasmid's phenotype. Based on BLASTN analysis of *repA* from plasmid R46, the seven plasmids analyzed here are the only IncN representatives in the sequence databases. Two other plasmids, isolated from *Yersinia pestis* and *Escherichia fergusonii*, have high similarity to half of *repA*, but very little similarity to any other IncN backbone genes. It will be interesting to see if future sequencing efforts identify more divergent IncN plasmid backbones.

The other three plasmids discussed in Chapter 2 were isolated from Gram-positive bacteria obtained during a study of poultry litter. A striking feature of these plasmids was

the lack of similarity to other nucleotide or protein sequences for 40-60% of the plasmid genomes. It was hypothesized that additional sequencing, especially of environmental Gram-positive bacteria, would fill in some of this missing information. Since the original analysis, 37 *Staphylococcus* and 4 *Corynebacterium* plasmid sequences were added to GenBank. Despite these additions, only 12 large (>30 kb) *Staphylococcus* and 4 large *Corynebacterium* plasmids are currently available. Clearly, there are still gaps in the plasmid sequence space. Even so, these new sequences and the development of tools to identify protein families and domains may fill in some of the missing information regarding the three Gram-positive plasmids.

In the original analysis, a number of ORFs were either similar to hypothetical proteins, or they were ORFans, meaning that they had no significant similarity to any other protein sequence in the database (4). To update this analysis, each of these ORFs was compared by BLASTP to the non-redundant NCBI protein database as of January 2009. The outcomes did very little to change the annotations of the ORFs (Table 4.1). Three ORFs from the *Staphylococcus* plasmid that were previously similar to hypothetical proteins were assigned a function based on information from protein domains and COG (Cluster of Orthologous Groups) features. These included an oxidoreductase, a beta-lactamase and an ABC-type transport protein. One ORF from the larger *Corynebacterium* plasmid pLEW279a and two ORFs from the smaller *Corynebacterium* plasmid pLEW279b that previously matched hypothetical proteins were also assigned functions. These included an integrase, a methyltransferase, and a cation transport ATPase.

For the *Staphylococcus* plasmid and the larger *Corynebacterium* plasmid, there was no change in the number of ORFans. However, two ORFans from the smaller *Corynebacterium* plasmid showed similarity to recent additions to the sequence databases: a hypothetical protein from an *Actinomyces* chromosome and a hypothetical protein from a *Corynebacterium* plasmid. ORFans are one of the enduring puzzles of genomic analysis, and debate continues regarding their origins (3, 18). It has been suggested that most ORFans do not encode functional proteins (15), however, recent experimental studies have validated a few ORFans as legitimate protein-coding genes (2). In the case of the three plasmids analyzed here, some of the ORFans are <50 amino acids in length and may not encode functional proteins, but many are 100 amino acids or more. Future experimental studies will undoubtedly validate some of these ORFans as protein-coding genes and hopefully shed some light on their functions.

Previous analyses have suggested that 10-15% of the ORFs on a bacterial chromosome are ORFans (12); an equivalent percentage of plasmid ORFs (13%) had no significant similarity to any other protein sequences in an analysis of plasmids within the ACLAME database (13). The percentage of ORFans reported here for the *Staphylococcus* plasmid and the smaller *Corynebacterium* plasmid agree with those numbers. In contrast, 28% of the ORFs from the larger *Corynebacterium* plasmid are ORFans. This percentage is much closer to the 30% reported for viral genomes (5, 21). It is curious that such disparate ORFan distributions are found on plasmids from the same *Corynebacterium* isolate. In this case, it seems likely that the poor coverage of *Corynebacterium* plasmids in the current sequence databases may be partially responsible for the lack of similarity. This missing information highlights a persistent gap in our knowledge of natural plasmid

variation. Additional sequencing of plasmids from non-clinical sources and under-represented bacterial genera is essential for understanding the complexity of the horizontal gene pool.

BROAD IMPLICATIONS OF DISSERTATION RESEARCH

In contrast to the sequence-level approach of Chapter 2, Chapter 3 describes a population-level analysis of natural plasmid variation. Restriction fragment length polymorphism (RFLP) analysis of individual large plasmids from *E. coli* and *Salmonella* reference collections revealed a high level of variation, even within the common replicon families IncF and IncI1. With the exception of the pSLT virulence plasmid detected in multiple strains of *Salmonella enterica* serovar Typhimurium, there were no widely distributed plasmid RFLP patterns, demonstrating that conservation of entire plasmid sequences, including backbone and accessory genes, is rare. Combined with the variation in IncN plasmid accessory content first observed in Chapter 2 and further described above, these data emphasize that natural plasmid variation is extensive. Clearly, most plasmids do not persist in nature as static entities. Instead, these members of the horizontal gene pool are evolving in response to various selective pressures.

To define how plasmids are evolving in bacterial populations, it is important to define the "rules" of plasmid evolution; for instance, how do different selective pressures affect the evolution of plasmid backbones or the gain and loss of accessory elements? The conservation of the pSLT RFLP pattern reported in Chapter 3 suggests that stabilizing selection maintains a very steep adaptive peak and limits the genotypic variation of this plasmid. The pSLT genotype may occupy a fitness optimum for the host niche provided by serovar Typhimurium, therefore any divergence or mutation is selected

against. Although our sequence data for the IncN plasmids is limited, it is possible that stabilizing selection also influenced the high conservation of the IncN plasmid backbone in *E. coli* and *Salmonella* host strains. Upon transferring to *Klebsiella*, the backbone genotype may not have occupied the fitness optimum (due to the new host niche), and mutations occurred that increased the genotype's fitness and moved it towards a fitness optimum for the new environment.

From the above examples, it seems that stabilizing selection should act to limit the variation of plasmid backbones or even complete plasmid genomes. However, the majority of plasmids in the *E. coli* and *Salmonella* reference collections were highly variable by RFLP analysis. In addition, the comparison of IncN plasmids presented above, along with previous studies of other plasmid families, has convincingly shown that gain and loss of accessory elements is a common occurrence for most plasmid genomes. Therefore, how do fluctuations in accessory gene content affect the fitness of plasmid genotypes? Are there ideal combinations of backbones and accessory elements that are well-equipped for particular host environments, or is it an "everything is everywhere" affair? What are the evolutionary pressures that allow certain plasmid variants to persist in bacterial populations and force others to be lost? Such questions require further investigation to get at the heart of plasmid evolution in nature.

A plasmid-centric approach is necessary, since it appears that the rules of evolution defined for bacterial chromosomes may not be applicable to plasmids. For example, both strong and weak selective pressures are expected to result in genome contraction for bacterial chromosomes (12). If this "rule" also applies to plasmids, we would expect plasmids to shed non-adaptive accessory genes. Instead, some have

suggested that plasmids actually add accessory genes in response to strong selection (11). In support of this contention, the plasmid profiles in Chapter 3 showed that large (≥30 kb) plasmids, including a few of 200 kb or more, were abundant. Many plasmids encode multiple accessory genes, and most of these genes are probably not under simultaneous selection. Therefore, plasmids appear to flout at least one of the basic "rules" of bacterial chromosome evolution. Obviously, plasmids are a different beast than chromosomes. How does their evolution affect adaptation and evolution of the host bacteria and the balance of bacterial community dynamics? How does plasmid evolution compare to that of bacteriophages, their partners in horizontal gene transfer? These are key questions for further study.

As discussed in Chapter 1, a limited number of experimental evolution studies and a few comparative genomics analyses have focused on the concept of plasmid evolution, but it does not yet occupy a prominent place in plasmid biology, nor has it attracted the attention of evolutionary biologists. The theoretical considerations of plasmid population biology outlined in Chapter 1 mostly ignore the possibility that adaptive genes may be gained or lost by plasmids or that plasmids may evolve to reduce or eliminate fitness cost to the host. The existing models are an important and valuable starting point for analyzing plasmid persistence, but now that evidence has accumulated to support dynamic plasmid evolution, these processes must be incorporated into theories of plasmid population biology.

For example, we can imagine a scenario in which the frequent exchange of accessory elements among plasmids leads to plasmid persistence. A plasmid-bearing bacterial strain enters an ecological niche, whereupon the plasmid is transferred to a new

host strain. This strain already carries a plasmid that imposes a fitness cost on the host. The strain may even be at low frequency in the bacterial community due to its growth disadvantage compared to other strains. However, if the resident plasmid recombines with the newly transferred plasmid and acquires a transposon encoding an advantageous gene, the resulting variant of the resident plasmid now confers an adaptive advantage on the host. The host strain, and its plasmid, will now persist in the bacterial community. In this imaginary scenario, plasmid evolution directly leads to plasmid persistence.

Bergstrom and colleagues argued that any plasmid-borne advantageous genes should eventually be transferred to the chromosome, causing the plasmid to lose its adaptive benefit (1). But, the acquisition of adaptive genes is not a one-time event for plasmids. As selective pressures change and competition between bacterial strains fluctuates, the plasmid will continue to lose accessory genes and gain new ones from other sources. In this way, the flexibility of the plasmid genome may allow it to persist in the population.

The above scenario can be considered a gene-based variation of the "selective sweep" hypothesis, which proposes that plasmid transfer to a strain that is sweeping through the population can result in plasmid persistence. Instead of plasmids transferring to fitter strain genotypes, the constant flux of adaptive genes from other mobile genetic elements may allow a plasmid to persist due to the acquisition of accessory genes under selection. At the very least, this reshuffling of accessory genes may enable a plasmid to persist in a population long enough for compensatory evolution to occur and reduce the cost of plasmid carriage, as observed in experimental studies. Scenarios like this illustrate

how plasmid evolution may impact plasmid population dynamics and reinforce the need to integrate plasmid ecology and evolution into future studies of plasmid persistence.

Another key conclusion from the data presented in the dissertation is that investigating plasmid evolution with traditional phylogenetic methods is complicated and possibly misleading. Certainly, narrow datasets, such as the IncN plasmid backbones discussed above, can be analyzed using methods that assume descent from a common ancestor. But, just as horizontal gene transfer can obscure the analysis of bacterial chromosomes, the highly variable accessory content of plasmids complicates phylogenetic analysis of plasmid genomes. The extensive variation in RFLP patterns detected within common replicon families in *E. coli* and *Salmonella* indicate that the natural history of the entire plasmid cannot be accurately reconstructed from backbone genes alone. Alternative methods of representing plasmid evolution are essential. Leplae and colleagues proposed networks as an attractive alternative to tree-based methods; they have presented proteomic graphs depicting plasmids as nodes joined by edges corresponding to shared protein families (13). As the number of shared families required to join plasmid nodes was increased, groups of plasmids emerged that may have shared evolutionary histories. Approaches like this are important for integrating both backbone and accessory genes in the analysis and reconstruction of plasmid evolution.

In addition to the broad implications for plasmid evolution, the data presented here also highlight a gap in our understanding of the ecological roles of plasmids. Since their identification as vehicles for antibiotic resistance, a clinical perspective has dominated consideration of plasmids' impact on bacterial community dynamics. However, plasmids also affect symbiotic interactions between bacteria and their hosts and

provide genes that allow bacteria to use xenobiotic compounds (19). The *E. coli* and *Salmonella* reference collections used in Chapter 3 have been surveyed for plasmid-borne antibiotic resistances (10, 14), but very little information is available on other accessory functions, despite the abundance of large plasmids. The sequence analysis of the Gram-positive plasmids presented in Chapter 2 identified antibiotic and metal resistances, but also a two-component response regulator and kinase and a glycine betaine transporter. A more thorough characterization of these accessory functions is necessary to fully understand the plasmid ecology.

FUTURE RESEARCH

As mentioned above, increased sequence coverage of large plasmids is one essential focus for future research. Plasmid sequence data is sparse or missing for a number of bacterial and archaeal species as well as for some ecological niches. As a result, we have a very narrow picture of how plasmids affect bacterial community dynamics, and our understanding of plasmid genome evolution is lagging well behind that of chromosomal evolution. An ongoing project in Dr. Summers' laboratory is sequencing plasmids from a diverse assortment of bacteria, including the marine symbiont *Vibrio fischeri* and phytopathogenic *Pseudomonas* species. This and other such projects will contribute to the expansion of the plasmid sequence space.

Computational analysis of plasmid sequences offers limited ability to assign functions to plasmid ORFs. Experimental analysis will be necessary to determine the functions of plasmid proteins and further characterize their backbone systems and accessory content. Metagenomics also offers a particularly taxing challenge to plasmid biologists. It is difficult to confidently assemble a complete bacterial chromosome from

metagenomics data at this stage; the mosaic nature of plasmids presents an even more complicated task. However, metagenomics is a powerful tool for *in situ* analysis of bacterial communities, and it is important to develop methods and tools for gleaning plasmid information from these vast datasets (17).

Experimental evolution will be a powerful tool for defining the "rules" of plasmid evolution. As discussed above, there are a number of predictions for bacterial chromosome evolution, and an experimental evolution approach is ideal for testing their applicability to plasmids. Additionally, the gain and loss of accessory genes and the resulting effects on plasmid persistence are excellent questions to address using experimental evolution. However, it will be important to incorporate substantial ecological complexity in order to simulate plasmid evolution in natural bacterial communities.

TABLE 4.1. BLASTP analysis of Gram-positive plasmid sequences.

Plasmid	GeneMark ORF calls	ORFs with hit to known proteins (% of total ORF calls)		ORFs with hit to hypothetical proteins (% of total ORF calls)		ORFans[a] (% of total ORF calls)	
		2005	2009	2005	2009	2005	2009
Staphylococcus pLEW6932	56	35 (63)	38 (68)	13 (23)	10 (18)	8 (14)[b]	
Corynebacterium pLEW279a	43	28 (65)	29 (67)	3 (7)	2 (5)	12 (28)[b]	
Corynebacterium pLEW279b	27	16 (59)	18 (67)	7 (26)[b]		4 (15)	2 (7)

[a] ORFans are defined as ORFs with no significant hit by BLASTP to any other protein in the NCBI non-redundant protein sequence database.

[b] Merged cells indicate no change from 2005 to 2009.

REFERENCES

1. **Bergstrom, C. T., M. Lipsitch, and B. R. Levin.** 2000. Natural selection, infectious transfer and the existence conditions for bacterial plasmids. Genetics **155:**1505-19.

2. **Chin, K. H., S. K. Ruan, A. H. Wang, and S. H. Chou.** 2007. XC5848, an ORFan protein from Xanthomonas campestris, adopts a novel variant of Sm-like motif. Proteins **68:**1006-10.

3. **Daubin, V., and H. Ochman.** 2004. Bacterial genomes as new gene homes: the genealogy of ORFans in E. coli. Genome Res **14:**1036-42.

4. **Fischer, D., and D. Eisenberg.** 1999. Finding families for genomic ORFans. Bioinformatics **15:**759-62.

5. **Frost, L. S., R. Leplae, A. O. Summers, and A. Toussaint.** 2005. Mobile genetic elements: the agents of open source evolution. Nat Rev Microbiol **3:**722-32.

6. **Garner, E. C.** 2008. GE Prize essay. Understanding a minimal DNA-segregating machine. Science **322:**1486-7.

7. **Gilmour, M. W., N. R. Thomson, M. Sanders, J. Parkhill, and D. E. Taylor.** 2004. The complete nucleotide sequence of the resistance plasmid R478: defining the backbone components of incompatibility group H conjugative plasmids through comparative genomics. Plasmid **52:**182-202.

8. **Gonzalez-Zorn, B., A. Catalan, J. A. Escudero, L. Dominguez, T. Teshager, C. Porrero, and M. A. Moreno.** 2005. Genetic basis for dissemination of armA. J Antimicrob Chemother **56:**583-5.

9. **Gootz, T., F. Dib-Hajj, M. Lescoe, R. Huard, and P. Della Latta.** 2007. Multiresistant *Klebsiella pneumoniae* from NYC hospitals contain beta-lactamase KPC-2 or KPC-3 and porin changes. Int J Antimicrob Agents **29:**S107-S108.

10. **Houndt, T., and H. Ochman.** 2000. Long-term shifts in patterns of antibiotic resistance in enteric bacteria. Appl Environ Microbiol **66:**5406-9.

11. **Kado, C. I.** 1998. Origin and evolution of plasmids. Antonie Van Leeuwenhoek **73:**117-26.

12. **Koonin, E. V., and Y. I. Wolf.** 2008. Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. Nucleic Acids Res **36:**6688-719.

13. **Leplae, R., G. Lima-Mendez, and A. Toussaint.** 2006. A first global analysis of plasmid encoded proteins in the ACLAME database. FEMS Microbiol Rev **30:**980-94.

14. **Mazel, D., B. Dychinco, V. A. Webb, and J. Davies.** 2000. Antibiotic resistance in the ECOR collection: integrons and identification of a novel aad gene. Antimicrob Agents Chemother **44:**1568-74.

15. **Ochman, H.** 2002. Distinguishing the ORFs from the ELFs: short bacterial genes and the annotation of genomes. Trends Genet **18:**335-7.

16. **Revilla, C., M. P. Garcillan-Barcia, R. Fernandez-Lopez, N. R. Thomson, M. Sanders, M. Cheung, C. M. Thomas, and F. de la Cruz.** 2008. Different pathways to acquiring resistance genes illustrated by the recent evolution of IncW plasmids. Antimicrob Agents Chemother **52:**1472-80.

17. **Schluter, A., L. Krause, R. Szczepanowski, A. Goesmann, and A. Puhler.** 2008. Genetic diversity and composition of a plasmid metagenome from a wastewater treatment plant. J Biotechnol **136:**65-76.

18. **Siew, N., and D. Fischer.** 2003. Unravelling the ORFan Puzzle. Comp Funct Genomics **4:**432-41.

19. **Top, E. M., Y. Moenne-Loccoz, T. Pembroke, and C. M. Thomas.** 2000. Phenotypic traits conferred by plasmids, p. 249-285. *In* C. M. Thomas (ed.), The Horizontal Gene Pool: Bacterial Plasmids and Gene Spread. Harwood Academic Publishers, Amsterdam.

20. **Wireman, J., C. A. Liebert, T. Smith, and A. O. Summers.** 1997. Association of mercury resistance with antibiotic resistance in the gram-negative fecal bacteria of primates. Appl Environ Microbiol **63:**4494-503.

21. **Yin, Y., and D. Fischer.** 2008. Identification and investigation of ORFans in the viral world. BMC Genomics **9:**24.