Published in final edited form as:

Microbiol Spectr. 2014 ; 2(1): . doi:10.1128/microbiolspec.MGM2-0022-2013.

Distributive Conjugal Transfer: New Insights into Horizontal Gene Transfer and Genetic Exchange in Mycobacteria

Keith M. Derbyshire and Todd A. Gray

Division of Genetics, Wadsworth Center, New York State Department of Health Albany, NY 12201, USA and Department of Biomedical Sciences, University at Albany, Albany, NY 12201, USA

Abstract

The last decade has seen an explosion in the application of genomic tools across all biological disciplines. This is also true for mycobacteria, where whole genome sequences are now available for pathogens and non-pathogens alike. Genomes within the *Mycobacterium tuberculosis* Complex (MTBC) bear the hallmarks of horizontal gene transfer (HGT). Conjugation is the form of HGT with the highest potential capacity and evolutionary influence. Donor and recipient strains of *Mycobacterium smegmatis* actively conjugate upon co-culturing in biofilms and on solid media. Whole genome sequencing of the transconjugant progeny demonstrated the incredible scale and range of genomic variation that conjugation generates. Transconjugant genomes are complex mosaics of the parental strains. Some transconjugant genomes are up to one-quarter donor-derived, distributed over 30 segments. Transferred segments range from ~50 bp to ~225,000 bp in length, and are exchanged with their recipient orthologs all around the genome. This unpredictable genome-wide infusion of DNA sequences is called Distributive Conjugal Transfer (DCT), to distinguish it from traditional *oriT*-based conjugation. The mosaicism generated in a single transfer event resembles that seen from meiotic recombination in sexually reproducing organisms, and contrasts with traditional models of HGT. This similarity allowed the application of a GWAS-like approach to map the donor genes that confer a donor mating identity phenotype. The mating identity genes map to the *esx1* locus, expanding the central role of ESX-1 function in conjugation. The potential for DCT to instantaneously blend genomes will affect how we view mycobacterial evolution, and provide new tools for the facile manipulation of mycobacterial genomes.

Eukaryotes use meiotic recombination to blend parental genomes and generate genome-wide variation between progeny. By contrast, bacteria divide asexually to generate two clones of the original. The genomic variation present in asexual populations is modest, and limited to those relatively rare events of spontaneous mutation and transposition. Significant variation must await exceedingly rare events wherein DNA from another organism is sporadically introduced into the bacterial genome. The acquisition of foreign DNA by a recipient bacterium provides the genetic diversity needed to facilitate evolution. This process is called horizontal gene transfer (HGT, or sometimes LGT for lateral gene transfer) and is mediated by three distinct mechanisms: conjugation, transformation and transduction (1, 2). In this chapter we will discuss in detail HGT mediated by conjugation in mycobacteria, its potential

evolutionary impact, and its application as a tool to mycobacterial genetics. Transduction is discussed in Chapter X on mycobacteriophages. Transformation (by electroporation) is a vital laboratory tool, but transformation *per se* is unlikely to play a major role in HGT because mycobacterial species are not known to be naturally competent.

HGT can occur between cells of the same species, different species, and even across kingdoms, thus blurring species boundaries. In the pre-genomic era, the impact of HGT was mostly restricted to our knowledge of the movement of mobile elements such as plasmids and phages. However, as the number of bacterial whole genome sequences increases, it is becoming clear that HGT has significantly influenced the genomes of extant bacteria (3–7). This is also true for the members of the *Mycobacterium tuberculosis* complex (MTBC: *M. africanum*, *M. bovis*, *M. canettii*, *M. microti*, *M. pinnipedii*, and *M. tuberculosis*), whose genomes have been extensively studied in an attempt to type and track TB outbreaks. HGT was not originally considered a major factor in the MTBC, because preliminary sequence comparisons among them indicated extremely low diversity and lacked any clear evidence of HGT (8–11). Consequently, it was postulated that the MTBC is essentially a clonal outgrowth of a particularly successful subtype of a progenitor species (10–12). However, more recent sequence analyses have provided convincing evidence of recombination occurring between isolates of *M. canettii*, the most divergent member, and probable progenitor species of the MTBC (13, 14). These sequence analyses, combined with the characterization of an entirely novel mycobacterial conjugation system have forced a re-thinking of the clonal paradigm and the role of HGT in shaping mycobacterial genomes (14, 15). In this chapter we will discuss the latest results indicating that HGT, specifically conjugation, is active amongst mycobacterial species, and that it likely has had--and will continue to have--a large impact on mycobacterial evolution.

CLASSICAL *oriT*-MEDIATED CONJUGATION - A BRIEF OVERVIEW

Plasmid Transfer

Bacterial conjugation is a naturally occurring process that involves the unidirectional transfer of DNA from a donor to a recipient, and requires cell-cell contact (16, 17). Conjugation generally involves the transfer of plasmids, or integrative and conjugative elements (ICE). ICEs are found in the chromosome, but excise to form a plasmid-like circle before transferring into a recipient and reintegrating into the chromosome (18). Most of the early characterization of conjugative plasmids and the transfer process was completed in *Escherichia coli*, but the mechanism of transfer is essentially the same for both plasmids and ICEs, regardless of bacterial species (16, 19).

The first step in conjugation is establishment of cell-cell contact between the donor and recipient cells, or mating-pair formation (mpf). In Gram-negative species, mpf is facilitated by the plasmid-encoded pilus that is assembled by a type IV secretion system (T4SS) (20, 21). Gram-positive species also encode many of the T4SS proteins, but they are mainly localized to the cytoplasm or membrane, and are therefore proposed to assemble as the conduit of DNA transfer and not to assemble pilus structures (21, 22). Instead, Gram-positive mating-pairs are likely formed by other mechanisms including surface exposed adhesins, which cause aggregation of donor and recipient Enterococci cells (23, 24). Once

mpf is established, DNA processing enzymes mediate transfer of a single-strand of DNA from the donor to the recipient (Fig. 1A). A key protein is a relaxase, which induces a strand-specific nick within a sequence called the origin of transfer, *oriT* (16). The relaxase, bound to the 5' end of the nicked DNA, mediates transfer of the DNA through the pore and into the recipient. On completion of plasmid transfer, the relaxase recognizes the 3' end of *oriT* and recircularizes the single strand by a reversal of the nicking process. Complementary DNA synthesis in the recipient results in a transconjugant containing a copy of the conjugative plasmid and, thus, is now a donor. This replicative process facilitates the rapid dissemination of conjugative plasmids through bacterial populations.

Chromosomal Transfer Mediated by Integrated Plasmids

Despite this plasmid-centric view of conjugation, the first description of conjugation was the transfer of *E. coli* chromosomal DNA by Lederberg and Tatum (25). Unaware that plasmids existed and promoted transfer, Lederberg and Tatum selected for transfer of chromosomal markers. Fortunately, one of the strains they were using had an F plasmid integrated into the chromosome, and thus chromosomal DNA was transferred, as it contained the *cis*-acting *oriT*. The mechanism of transfer by these so-called Hfr strains (high frequency recombination) is identical to that of plasmid transfer, but the outcome is very different (Fig. 1B). As the chromosome is now conceptually equivalent to a *very* large plasmid, transfer requires a commensurately extended time, and depends on nick-free DNA to be pulled along by the relaxase-*oriT* complex. Therefore, transfer is often incomplete, either due to physical dissociation of mating pairs, or because of nicks in the strand of DNA being transferred. Consequently only the segment of the donor chromosome adjacent to *oriT* is reliably transferred. Since partial transfer prevents reconstitution of *oriT* to circularize the chromosome, this linear segment must recombine into the recipient chromosome by homologous recombination for stable inheritance. Thus, chromosomal transfer contrasts with plasmid transfer: (i) host recombination functions are required to integrate transferred DNA into the chromosome; (ii) *oriT* is not regenerated in the recipient, and therefore transconjugants do not become donors, but remain as recipients; and, (iii) genes 3' proximal to *oriT* are transferred most frequently, and so the efficiency of transfer of chromosomal genes is location dependent. By taking advantage of these observations, Wollman *et al.* established the circularity of the *E. coli* genome, and mapped genes by minutes around the “100-minute” chromosome (26). As we will describe below, the outcomes and requirements of transfer in *M. smegmatis* substantially differ from *oriT*-initiated transfer, providing the first clues that mycobacterial conjugation occurs by an entirely new mechanism.

MYCOBACTERIAL CONJUGATION

Do Mycobacterial Plasmids Mediate Transfer?

Rather strangely, mycobacteria seem to lack the sheer numbers and classes of plasmids found in almost all other bacteria. While many environmental species contain plasmids (27), many mycobacteria lack plasmids altogether, including members of the MTBC and *M. smegmatis*. This paucity of plasmids in mycobacteria explains why all natural antibiotic resistance is chromosomally mediated in *M. tuberculosis*, unlike most other pathogens. We speculate that this is a consequence of the unusual nature of the mycobacterial envelope

preventing plasmid transfer, and the unique requirements for plasmid replication and maintenance in mycobacteria.

In general, mycobacterial plasmids have only been partially characterized, while the main focus has been their development as vectors (28, 29). Descriptions of plasmids are mostly limited to the environmental species, such as members of the *M. avium*, *intracellulare*, *scrofulaceum* (MAIS) complex and *M. fortuitum* (27, 28). Hybridization studies indicate that some of these plasmids are related despite being found in different species and isolates, which suggests that the plasmids have moved between species in the environment (30). However, the technical problems associated with manipulating members of the MAIS complex and other environmental mycobacteria have made it difficult to experimentally address the mechanism of dissemination. Most of these plasmids can be transformed and maintained in other mycobacterial species, such as *M. smegmatis* and *M. tuberculosis*, indicating their limited host range is not an inability to replicate, but a consequence of either a physical inability to spread or because their hosts do not occupy the same environmental niches (28, 31, 32).

A few of these characterized mycobacterial plasmids have been sequenced and these analyses have not identified obvious transfer functions (31, 33, 34). *M. ulcerans* contains a large plasmid, pMUM1001 (>174 Kb), which encodes polyketide synthases that produce the macrolide toxin necessary for its pathogenesis (35). Its large size has prevented functional characterization, but sequence analysis again indicates that pMUM1001 does not encode a classical conjugation system (36). Together, the lack of defined transfer-associated genes on these plasmids suggests that either they are not transferred between cells, plasmid transfer occurs by a novel mechanism, or that they are indirectly mobilized between environmental species via chromosomal transfer, as observed in *M. smegmatis* (37).

Sequence analyses of two mycobacterial plasmids indicates they encode genes related to those found on classical *oriT*-like plasmids, but these plasmids have not been shown to be mobile in mycobacteria. The sequence of the plasmid pVT2 from *M. avium* revealed that it encodes a relaxase, with homology to the relaxases of F, R100 and R388, including conservation of the active site sequence motifs (31). However, pVT2 mediated conjugation was not experimentally demonstrated and the plasmid lacks other conjugation-like genes. One possible scenario is that pVT2 is a mobilizable plasmid, encoding its own relaxase and *oriT*, but requiring a conjugative plasmid, or equivalent, to establish mating-pair formation. More recently, a plasmid was identified in an epidemic strain of *M. abscessus*, which was sequenced and shown to belong to the broad-host-range IncP plasmids found in many Gram-negative bacteria (38). The 56 kb plasmid, pMAB01, contained all the genes needed for mating-pair formation and DNA transfer. Importantly, when introduced into *E. coli*, pMAB01 could transfer between strains of *E. coli*, albeit at low frequencies. While this study demonstrated that a large broad-host range plasmid can get into a mycobacterial cell (probably by conjugation, also see (39), the restriction of pMAB01 to a single epidemic strain suggests that it is unable to mediate transfer efficiently between mycobacteria, i.e. mycobacteria are a dead end for *oriT*-mediated plasmid transfer systems. We speculate that this is because of the novel composition and architecture of the mycobacterial membrane, which prevents assembly of a functional T4SS for pilus synthesis and DNA export. This

membrane barrier may also prevent acquisition of broad-host-range plasmids, which are normally capable of transfer into an extraordinarily diverse range of bacterial species.

Chromosomal Transfer in Mycobacteria

Similar to its discovery in *E. coli*, the first definitive proof of conjugation in mycobacteria was that of chromosomal DNA transfer in *M. smegmatis* (40). However, as we will describe, the mechanism of transfer is fundamentally different, requiring a complete rethinking of the conjugation process and its impact on genome dynamics and mycobacterial evolution.

The Basics...—The first accounts of a conjugation-like process between isolates of *M. smegmatis* were described in the early 1970s (41, 42). However, these initial observations required modern molecular genetic tools to convincingly demonstrate that recombinants were generated by conjugal transfer of chromosomal DNA and did not involve plasmids (40). To monitor DNA transfer, cassettes encoding antibiotic resistance were integrated into the chromosomes of *M. smegmatis* isolates, the differentially marked strains were co-incubated, and transconjugants identified by selecting for recombinant cells resistant to both antibiotics (Fig. 2). This simple assay confirmed that DNA transfer in *M. smegmatis* satisfies all criteria of conjugation: (i) transfer required co-incubation of both parent cells and that both parental types were viable, (ii) there are distinct donor and recipient strains, (iii) transformation was excluded because recombinants were isolated in the presence of DNaseI, (iv) transduction was also ruled out because transfer did not occur in liquid medium and phage could not be detected in culture filtrates, and (v) cell fusion was ruled out because an episomal plasmid introduced into the donor strain was not transferred in transconjugants even when a plasmid-encoded antibiotic marker was selected (40). One pair of isolates, derivatives of Jucho and P73, while normally acting exclusively as recipients with other donors, could exchange antibiotic markers bidirectionally. This unusual bidirectional transfer could occur if these two strains are capable of switching mating-types, i.e. they contain genetic information for both donor and recipient functions, but only one function is expressed in any one cell. It is possible this switch is relevant to the conversion of transconjugants to donors, which will be discussed later.

Of considerable note to the mycobacterial research community was that mc²155, the widely used laboratory strain, was characterized as a conjugal donor. This had the reciprocal dual benefit of using tools developed in mc²155 to study the conjugation mechanism, while allowing the development of conjugation as a tool to study mc²155.

Initiation of Mycobacterial DNA Transfer Does Not Occur From a Unique Site

—*oriT* regions are small (~200 bp) *cis*-acting sequences that include all the necessary sites for the relaxase and accessory proteins to bind, nick and mediate DNA transfer (16). They are easily isolated as they confer mobility on otherwise non-mobilizable plasmids, when the necessary transfer proteins are supplied in *trans* (43, 44). The pAL5000-based mycobacterial plasmid is not mobilized, even when a plasmid antibiotic marker is selected, indicating it lacks the *cis*-acting sequences needed to mediate chromosomal transfer into the recipient cell. Wang *et al.* (2003) took advantage of this observation to isolate chromosomal *cis*-acting mobilization sites by constructing a library of chromosomal donor DNA in a

pAL5000 vector and selecting for its transfer into the recipient. By analogy with other transfer systems, it was anticipated that a single, *cis*-acting *oriT*-containing DNA segment would be identified in vector derivatives that successfully transferred. Surprisingly, multiple, non-overlapping segments of donor DNA were identified in plasmid transconjugants, indicating that there are many *cis*-acting sequences in the chromosome. The *cis*-acting sequences were termed *bom* regions, for basis of mobilization, to distinguish them from the smaller, singular, better characterized, *oriT*.

oriT provides two functions: it initiates transfer in the donor and re-circularizes plasmid DNA in the recipient to complete transfer (Fig. 1A). In *M. smegmatis*, a study to define functionally comparable *cis*-acting sequences revealed that efficient plasmid transfer required at least 5 kb of chromosomal DNA and often resulted in acquisition of recipient chromosome SNPs in the transferred plasmid DNA (37). This simple observation provided key insights for a new model, which proposed that *bom* might initiate transfer, but re-circularization of a mycobacterial plasmid required homologous recombination functions in the recipient in the form of gap repair, using the recipient chromosome as a template (Fig. 3A). A gap-repair model encompasses all of the experimental findings: transfer requires a minimum size of *bom* (for homology to promote strand invasion and gap repair), the requirement of RecA in the recipient, and the acquisition of recipient SNPs in the plasmid DNA. In further support of this gap-repair model, plasmids were engineered to contain two neighboring *bom* regions, normally spaced 5 kb apart on the chromosome. Transconjugants containing these two-*bom* plasmids were shown to have acquired the entire intervening segment of recipient chromosomal DNA (Fig. 3B; (37, 45). The recovered *bom* sequences were not enriched for elements that might function as a binding site for an enzyme, or have any discernable structural features, similar to *oriT*. In addition, although *bom* initiates plasmid transfer it is unclear whether initiation involves a nick or a double-stranded break, resulting in single-stranded or double-stranded DNA transfer respectively.

Extrapolating these plasmid findings to a chromosomal context suggests that cleavage at multiple chromosomal *bom* sites could result in transfer of discrete segments of DNA that are then integrated into the recipient chromosome by homologous recombination. In addition, the presence of numerous *bom* sites is consistent with location-independent DNA transfer, in contrast to that observed in Hfr strains (46). Importantly, these mechanistic distinctions are consistent with two very different evolutionary paths to conjugal DNA transfer. Chromosomal transfer by Hfr in *E. coli* is directed by *cis* and *trans* functionality from a plasmid—something that has evolved for self-propagation—simply because incidental integration embedded the plasmid in the chromosome. In contrast, mycobacterial conjugal DNA transfer is apparently plasmid-independent, implying that the necessary constituent conjugation proteins and elements have co-evolved for the express purpose of transferring chromosomal genes between cells.

Genetic Requirements of Mycobacterial Conjugation—Bioinformatic searches have failed to identify any orthologs of transfer genes in *M. smegmatis* genomes. In hindsight this is not entirely surprising, given both the unique aspects of mycobacterial transfer and the structure of the mycobacterial cell envelope (47). Further underscoring the unique properties of mycobacterial conjugation is the unanticipated role of the ESX-1

secretion apparatus in DNA transfer. Transposon mutagenesis screens demonstrated very different roles for the ESX-1 secretion apparatus in the donor and recipient. In the donor, ESX-1 functions suppress transfer: *esxI* donor mutants are hyper-conjugative (48). Paradoxically, in the recipient, ESX-1 functions are essential for transfer: *esxI* recipient mutants are non-conjugative (49). This reciprocal *esxI* conjugation phenotype suggests a level of complexity well beyond a simple model of DNA passing through a secretory apparatus.

ESX-1 is the flagship representative of type VII secretion systems. There are five apparently non-redundant paralogous *esx* loci in *M. tuberculosis*, with *esxI* mutants having an attenuated phenotype (discussed in more detail in Chapter X; (50, 51)). Briefly, ESX-1 is a complex secretion machine with its own specific set of substrates, which include the dominant antigens EsxA and EsxB (previously called Esat6 and Cfp10). The majority of the proteins required for secretion are encoded from a single, multi-gene locus called *esxI*. The genetic organization and the proteins encoded by *esxI* are highly conserved among mycobacteria, including *M. smegmatis*. ESX-1 mutants of *M. smegmatis*, like those of *M. tuberculosis*, fail to secrete EsxA and EsxB and other ESX-1-dependent proteins (49, 52). Thus, a role for ESX-1 in conjugation suggests an overarching function for ESX-1 in regulating extracellular interactions; in *M. smegmatis*, ESX-1 regulates interaction between mating mycobacterial cells, while in *M. tuberculosis*, ESX-1 regulates interactions with the host.

There are many unanswered questions concerning the mechanism and functions of Type VII secretion systems, but several are relevant to this chapter. How do the roles of ESX-1 differ between *M. smegmatis* and *M. tuberculosis*? And, within *M. smegmatis*, how do the roles differ between the donor and recipient strains? Given the high level of conservation of the *esxI*-encoded proteins, the most likely model is that the structural apparatus and mechanisms of secretion are the same, but some of the proteins that are secreted differ between species and strains. Thus, *M. tuberculosis* would secrete effector proteins to mediate pathogenesis, while the donor and recipient *M. smegmatis* strains would use the same delivery system to secrete different sets of proteins to regulate or mediate DNA transfer. Whether ESX-1 mediates these disparate functions through secretion of diffusible factors, or by decorating or modifying the mycobacterial cell, wall are pressing questions.

Could a Paradoxical Riddle Hold the Key?—Paradoxes are only paradoxes until the underlying reason is known. The opposing effects of *esxI* mutation in donor and recipient *M. smegmatis* strains defies most models of orthologous function, in which orthologous genes and their encoded proteins will have similar—not opposing—functions. Solving this paradox will likely go hand-in-hand with understanding the fundamentals of mycobacterial conjugation, but critical pieces of the puzzle are still missing.

In the donor, an active ESX-1 apparatus suppresses DNA transfer suggesting that the ESX-1 secreted proteins are inhibitors of conjugation (48). One possibility is that ESX-1-secreted proteins coat the surface of the donor and act as a physical barrier to prevent intimate cell contact and the transfer of DNA. An alternative hypothesis is that the secreted proteins are signaling proteins, or quorum sensors, that suppress activation of DNA transfer until suitable

recipient cells are present. There is precedent for transfer being activated by secreted peptides in the Gram-positive *Enterococcus*, in which peptides secreted by both donor and recipient regulate the induction and suppression of DNA transfer (24).

In the recipient, ESX-1 secreted proteins may be receptors on the cell surface that promote donor-recipient contact (49). Alternatively, as for the donor, the secreted recipient proteins may actively signal the donor to initiate transfer. A third scenario is that the ESX-1 apparatus itself is responsible for DNA uptake into the recipient. However, this model would imply that the donor ESX-1 apparatus is fundamentally different – donors cannot take up DNA – and that its presence negatively interferes with DNA transfer into the recipient. Regardless of model, DNA transfer provides a simple, sensitive and robust assay for ESX-1 function for the molecular genetic dissection of ESX-1 and its many, apparently disparate, roles in mycobacterial biology.

DISTRIBUTIVE CONJUGAL TRANSFER

Transconjugant genomes are mosaic

Recent revelations on mycobacterial DNA transfer have come from whole genome sequencing of transconjugant progeny (15). Whole genome comparison between transconjugant progeny and their parents is feasible because parental strains differ significantly at the nucleotide level (~1 SNP per 56 nts), thus allowing discrimination between the parental origin (donor or recipient) of genomic DNA in transconjugants. The most striking feature of such a comparison is that the progeny genomes are mosaic blends of the parental genomes, in striking contrast to that predicted for *oriT*-mediated transfer (Fig. 4). Thus, while all transconjugants acquire a segment of donor DNA encoding the selected marker (for example, *Km^r*), segments of DNA not selected are also transferred. In fact, on average, 12 additional segments of donor DNA are co-inherited for every selected *Km^r* segment, and these bonus segments are distributed around the genome with no obvious regional biases. The sizes of the donor segments vary dramatically, ranging from 59 bp to 226 kb, with an average size of 44.2 kb, and a mean of 13 tracts, totaling 575 kb of transferred DNA per genome (data from 22 transconjugants). To more accurately reflect the products of mycobacterial conjugation, and to distinguish it from the classical *oriT*-mediated transfer, this process was termed Distributive Conjugal Transfer (DCT).

Surprisingly, whole genome sequencing has not been reported for Hfr-transconjugants, preventing a detailed genomic comparison of the two conjugation systems. However, current Hfr models predict that a single segment of donor DNA would be integrated into the recipient chromosome corresponding to donor DNA located 3' of *oriT* (Fig. 1B). DCT is remarkable compared to other modes of HGT for two main reasons. First, DCT creates unprecedented genome-wide mosaicism within individual transconjugants, and, second, the variation it generates in a single step approaches that seen in sexual reproduction.

Models for DCT

Based on the mosaicism observed in progeny genomes, we speculate that random chromosomal DNA fragments are generated in the donor, some of which are co-transferred into a recipient cell where they replace recipient sequences through homologous

recombination (Fig. 5, left). The trigger for chromosome fragmentation is not clear, but the model is consistent with multiple donor *bom* break-sites initiating transfer of individual segments (37).

An alternative to the model described above is that a single, large DNA molecule is transferred into the recipient, which is processed into smaller segments before their integration into the recipient chromosome by homologous recombination (Fig. 5, right). This scenario seems less likely as it predicts some progeny would contain exceedingly large chunks of donor DNA (3–4 Mb) integrated into the chromosome. These large chunk events would have resulted from recombination close to the ends of the transferred molecule, sites that are known to load recombination machines and promote homologous recombination (53, 54). Based on the expected recombination frequencies it would seem more parsimonious for these large chunks to be integrated before fragmentation occurred into smaller segments. This large chunk scenario is also less consistent with previous observations, which indicated that the donor chromosome contained multiple initiation sites (37).

The majority of the integration events observed in transconjugants can be explained by homologous recombination promoting double crossover events at either end of the transferred DNA segment (Fig. 6, left). RecA is required for transfer in the recipient and likely works with the RecA-dependent AdnAB recombination system described by Glickman and colleagues (55). Mycobacteria encode a RecBCD complex, however these enzymes perform a different role in mycobacteria (RecA-independent single-strand annealing) than in *E. coli* and, therefore, are unlikely to mediate conjugational recombination as they do in *E. coli* (56, 57) and Chapter X). However, regions of micro-complexity that contain extremely short alternating tracts of donor and recipient DNA demonstrate that other mechanisms are also at work (Fig. 6, right). These tracts indicate that the recombinant products likely arose from resolution of heteroduplexes between homeologous sequences, rather than simple exchange events. In other bacteria, the mismatch repair proteins recognize such heteroduplex DNA and prevent recombination (58). However, mycobacteria lack mismatch repair genes (59) and, thus likely rely on alternative recombination mechanisms to resolve heteroduplexes, offering one explanation for the observed tracts of blended micro-complexity. Untangling the processes that create these tracts will require a defined genetic approach, identifying the required recombination functions known to be present in mycobacteria (NHEJ, RecBCD, AdnAB, RecO), and equally likely to reveal yet other recombination surprises (57, 60). Regardless of the mechanism, the net effect of micro-complexity is to generate a localized composite blend of nucleotide substitutions. From an evolutionary standpoint, these provide localized subtle diversity, which, for example, could modify the activity or interaction specificity of an enzyme.

Unanswered questions

For all that is known about the products of DCT, there is still much to be discovered about the process. Whether single-stranded or double-stranded DNA is transferred is unknown. While *oriT*-mediated transfer is via a single-stranded intermediate, conjugation in

Streptomyces is of double-stranded DNA (61, and below). The absence of known mycobacterial relaxases suggests that DCT might also be double-stranded, and obviates the requirement for complementary-strand synthesis in the recipient before recombination.

An intriguing possibility is that DCT is similar to the conjugation system in *Streptomyces*, a fellow actinomycete. *Streptomyces* conjugation is mediated by a single protein, TraB, which is plasmid encoded. No orthologs of the T4SS or pilus protein are encoded from the plasmids or chromosome, most likely because mating-pair formation involves hyphal fusion (22). TraB is an ATPase, resembling the DNA translocators FtsK/SpoEIII in its overall structural organization (62). In further functional analogy to FtsK, which binds 8 bp KOPS sequences to propel DNA through the septal pore (63), TraB binds a series of 8 bp repeats (TRS; TraB recognition sequences). TRS are found on transferable plasmids associated with TraB and at sites around the chromosome, and are thought to provide the specificity and directionality to transfer (62, 64). TraB localizes at the hyphal tips, where it is proposed to form a channel for translocation of DNA (64, 65). It is possible that a chromosomal ortholog of TraB performs a similar role in mycobacteria. If this protein also carried out essential replication functions, it might explain the inability to isolate transfer-defective donors (48). The dual functionality might also allow coordination of transfer and chromosome replication, such that transfer occurs after a round of replication ensuring that DCT is not a donor suicide process. The lack of a defined mycobacterial *oriT*-like sequence might also be explained by its TraB ortholog recognizing cryptic TRS-like repeats. However, whether DCT is mediated by a TraB ortholog, or by an entirely different process, will require more definitive experimentation, especially as there are so many differences between mycobacteria and *Streptomyces* (perhaps most significantly, in membrane organization and growth characteristics).

DCT AS AN ENGINE FOR MYCOBACTERIAL EVOLUTION

DCT Dramatically Shortens the Evolutionary Clock

The analysis of transconjugant genomes not only provides insights on the transfer mechanism, but also suggests that mycobacterial DCT can create genome diversity almost instantly. Genomic changes that typically accompany evolution of bacteria are assumed to be a serial accrual of HGT and spontaneous mutation events that occur over many generations and great spans of time (Fig. 7). By contrast, a single-step DCT event between two *M. smegmatis* cells generates a transconjugant that is a mosaic blend of the parental genomes, and not merely an incrementally altered derivative (Figs. 4 and 7). Thus, theoretically, DCT could dramatically increase the evolutionary rate estimates for mycobacteria.

Is DCT Still Active Among Native Mycobacterial Species?

Determining whether DCT is still active is extremely difficult to address. Naturally occurring HGT events can only be inferred from anecdotal sequence analyses of extant parental and transconjugant lineages. As the HGT event may have happened long ago, the participating parental strains may be extinct or may not have been sequenced as yet, and so the nearest sequenced relative in its lineage must suffice as a stand-in for comparisons.

Subsequent changes to either the parental or transconjugant genomes, including possible secondary HGT events, can complicate interpretations. Thus, while genome analysis is the most effective method to detect HGT, definitive proof that DCT is active in other species will require using sequenced, marked, parental strains and selecting for recombination events *in vitro*, as demonstrated for *M. smegmatis*. However, this also requires that the mating type of the species is known. For example, one would predict that all *M. tuberculosis* isolates are the same mating type, either donor or recipient, because of its clonal nature. Therefore, DCT is unlikely to occur between *M. tuberculosis* isolates. But, this does not preclude *M. tuberculosis* exchanging genetic information with other mycobacteria of the opposite mating type. Below, we describe three documented examples of mosaicism in mycobacterial genomes that could have been generated by DCT, although we caution hard conclusions because of the above caveats.

(i) DCT provides a plausible mechanism for the genome mosaicism observed in *M. canettii*—While DCT certainly is active in isolates of *M. smegmatis*, a bigger

question is whether it occurs in other mycobacterial species and especially, from a public health perspective, the MTBC. Early sequencing studies of *M. tuberculosis* isolates found extremely low genetic variation, suggesting that *M. tuberculosis* does not undergo HGT, are evolutionary young, and resulted from a recent clonal expansion (10–12). However, there is convincing evidence for HGT among *M. canettii*, and other smooth-colony MTBC strains: genome comparisons between *M. canettii* isolates have identified large numbers of SNPs (See Chapter by Brosch). Notably, clusters of SNPs are shared between different isolates, indicating that pairs of isolates have undergone recombination events via HGT (13). As a result of these pairwise recombination events, the *M. canettii* isolates display genome-wide mosaicism similar to that observed following DCT in *M. smegmatis* (13, 14). It was proposed that *M. canettii* strains are extant members of a genetically diverse MTBC progenitor species, *M. prototuberculosis*, whose members underwent frequent HGT, and that *M. tuberculosis* emerged from this pool of diverse species, by acquisition of enhanced virulence mechanisms (13).

The unspecified HGT process underlying the *M. canettii* genome mosaicism is presumed to result from a series of sequential transfer events. However, DCT involving the ancestral *M. prototuberculosis* offers a plausible and parsimonious explanation for the remarkably similar mosaicism observed among the extant *M. canettii*. We could envision that DCT in *M. prototuberculosis* rapidly incorporated the necessary blend of parental genotypes that drove the emergence of a pathogenic, rough-colony morphology species, like *M. tuberculosis*, allowing their subsequent clonal expansion.

(ii) Has DCT Occurred Between *M. canettii* and *M. tuberculosis*?—While the HGT events proposed for *M. prototuberculosis* are thought to have occurred over 35,000 years ago, there are indicators that gene flux still occurs amongst the *M. canettii* isolates and *M. tuberculosis*. Sequence comparison of the *M. canettii* isolates with *M. tuberculosis* confirms that their genomes are highly conserved and syntenic, but the two species are significantly more divergent and are easily distinguished by numerous SNPs. However, there are shared blocks of sequence identity embedded in regions of nucleotide diversity

detected between *M. tuberculosis* and individual *M. canettii* isolates (14, 66). These regions are entirely consistent with relatively recent recombination events that have occurred between the two species since their divergence. The strikingly similar pattern of mosaicism observed in our experimental system (Fig. 6), and among extant members of the MTBC, suggests that the latter may have been created by the same mechanism: DCT.

(iii) Genome Mosaicism Observed in *M. avium* could also result from DCT—A study on the glycopeptidolipid (GPL) biosynthetic pathways of four different *M. avium* strains, belonging to different serotypes led to the sequencing of their *gpl* gene clusters (67). The region was shown to be highly mosaic amongst the four strains, with multiple SNPs clearly defining shared regions and breakpoints. The lack of insertion sequences and the extended sequence continuity beyond the mosaic region strongly indicated acquisition by some form of HGT. Again, the products are remarkably similar to those generated by DCT.

We suspect that these observations of mosaicism represent the tip of the iceberg that will gradually be exposed as more mycobacterial species are sequenced and their genomes compared. While the pathogenic MTBC species are frequently sequenced for public health reasons, their extreme similarity undermines their value in identifying HGT events.

DCT Adds up to Reproductive Success

It is important to keep in mind that, in spite of the many superficial similarities between DCT and the mammalian version of sex, DCT is not reproduction *per se*. Sexual reproduction generally brings together two parents whose haploid genomes combine to create genetically blended progeny. A basic mathematical representation would be $1+1=3$, where each of the parents and the created single offspring total three individuals; a net population increase of one. Since the viability/fate of the donor bacterium is unknown in mycobacterial mating, DCT might either result in no net increase ($1+1=2$) or, in the case of donor suicide, a net decrease ($1+1=1$). Even though the short-term math may not tally, it is likely that in the long term, participants of DCT should eventually be present in greater numbers, as an occasional transconjugant may have a competitive advantage and divide more frequently.

The nearly boundless parameters of DNA transferred by DCT—large and small, isolated tracts or clustered micro-complexity—quickly create so many combinatorial possibilities that it becomes clear that no two transconjugants can be identical. A typical laboratory mating experiment generates $>10^4$ antibiotic selectable recombinants. If DCT transfers 10% of the donor genome on average (15), the corollary is that 90% of the transconjugants go undetected in this simple assay. Thus, of the $>10^5$ total transconjugants generated in an experiment, all will be different from each other, and from the parental strains.

DCT Complements Spontaneous Mutation

Mycobacteria undoubtedly experience spontaneous mutation, a ubiquitous source of *de novo* genetic diversity. However, only a very small fraction of mutations will be advantageous, and if multiple sequence changes are required for a competitive advantage, the beneficial mutations must occur sequentially in the same lineage to become fixed in the population.

This would likely require a very protracted time scale. DCT can expedite this process by actively mixing variants available within a community, and letting competition select for the best combinations. Importantly, not only does DCT introduce variation instantly, but it also brings in variant sequences that have already been vetted. They were functional in the context of the donor genome and therefore should have a reduced chance of carrying debilitating mutations (Fig. 6, right). While this consideration might elevate the transconjugant viability rate somewhat, their overall success will ultimately depend on how well the altered genes interact with partner genes or with pathways remaining in the cell, and how well the new genome interacts with the environment. This is analogous to trying on new shoes for a specific occasion. The existing pair might suffice, but another pair might be a better match for your clothing (loafers vs sneakers) or the environment (sandals vs boots) or the activity (cleats vs flippers). The rack of shoes from which to choose is limited only by the variation of the neighbors that are willing to donate them (DCT donors), and that the shoes fit (homology to recombine into the recipient). Whereas a random mutation might give rise to a nice fresh set of shoelaces, most mutations will more likely put a hole in the shoe or leave it stuck in the mud. The lack of mis-match repair systems in mycobacteria allows homeologous recombination and, thus, further enhances the repertoire of mutations a recipient can acquire from distant relatives.

DCT Can Result in Indels

Significant evolutionary changes require entirely new activities that cannot be created by incremental changes to a finite genome, and must be acquired through HGT. Classic horizontal gene transfer imports new genes (having no orthologue in the recipient) that are then integrated into the genome to result in Indels (inserts/deletions). Indels initiate potentially major evolutionary leaps, as these might include entire operons that encode complete signaling pathways, biochemical activities, or structural components, which could open up opportunities to populate a new environment. The extremely large sizes of some transferred segments (250 kb) in DCT allow it to introduce, or remove, non-homologous segments by bridging to regions of homology (Fig. 6, left). Indeed, insertions of up to ~50 kb were identified in *M. smegmatis* transconjugants, representing new donor DNA sequences that were originally absent in the recipient chromosome (15).

DCT AS A GENETIC MAPPING TOOL

esx1 Encodes a Mating Identity Switch

The mosaic genomes generated in a single-step by DCT look remarkably like the products of meiotic recombination in that homologous recombination indiscriminately swaps segments of DNA from both parents to create a new unique chromosome having tracts from both parents. Single sperm sequencing has shown that spermatogenesis creates an average of 23 crossovers per genome (68), comparable to exchanges by DCT in a mycobacterial genome. This level of genome mixing has been exploited to map genetic traits through association studies in mammals. With the exception of monozygotic twins (and inbred mouse lines), individuals have different combinations of sequence variants in their genomes. In spite of their genome-wide variation, some individuals may share a gene that gives rise to a distinct phenotype. Genome-wide association studies (GWAS) look for variants--usually

SNPs--that are over-represented in the group exhibiting the trait of interest relative to a group that does not have the phenotype. Therefore, SNPs tightly linked with a gene that controls a trait will be highly enriched in the affected group. The overt similarities between meiotic and DCT-generated mosaicism enabled a similar approach to be applied to mycobacteria. A mycobacterial GWAS can theoretically map any genetic trait that differs between the parental strains. This trait could be as simple as colony morphology or color, or as complex as a biochemical pathway, as long as there is a measurable phenotype. One practical application might be to map donor genes encoding drug resistance, upon transfer of resistance into the drug-susceptible recipient. Whole genome sequencing of drug-resistant transconjugants should identify a segment of DNA in common containing the responsible gene(s).

Most relevantly for this Chapter, this GWAS-DCT methodology was used to map a locus that conferred *mid* mating identity (*mid*) to *M. smegmatis* transconjugants (15). Unlike *oriT*-mediated Hfr transfer, a sub-set of mycobacterial transconjugants become donors (46), and therefore likely acquired a donor-conferring locus. Comparison of the genome sequence of just 10 donor-proficient transconjugants identified a single region in common, which encompassed the *mc*²155 *esxI* locus (Fig. 4), and indicating that *esxI* encodes a switch in mating identity from recipient to donor in these transconjugants (15). Despite the previous links between ESX-1 and transfer, this result was not anticipated, as transposon insertions or deletion of *esxI* genes do not result in a switch in mating phenotype. The ability to rapidly associate genes with phenotypes via this GWAS approach provides a simple and effective tool for mycobacterial geneticists. Importantly, DCT can be used to map genes by exchange of function, as opposed to traditional methods that employ loss-of-function mapping. The ease of sequencing bacterial genomes, combined with the dramatic drop in costs, makes this a cost-effective process.

Where are the *mid* Genes in *esxI*?

In mammalian genetic studies, fine mapping of a genetic determinant can be achieved by performing successive backcrosses to genetically purify a locus in a recipient background. A similar approach was employed using DCT to introgress the *mid*-locus in the recipient genome. F1 donors were crossed with the original recipient to generate (N1) progeny that were screened for donor proficiency. N1 donors were then used in a second backcross and the process repeated. Six serial backcrosses resulted in a purifying selection of the donor-conferring locus (and the *Km^r* genes used to select for transfer of donor DNA) such that donor-proficient transconjugants contained as little as 1.5% of donor DNA embedded in a recipient background (Fig. 8). Just as informative are those lines that were reiteratively scored as donors (transferring the donor *mid* region of the *esxI* locus at each generation), up until the last generation when some transconjugants, now scored as recipients, were also sequenced. In this last generation, the recipient-proficient transconjugants invariably retained all or most of the *esxI* genes of the recipient, as donor-specific SNPs throughout this region were not transferred (Fig. 8).

By combining the genome data of both donor-proficient backcross progeny (having donor *mid* genes that were sufficient to confer the donor phenotype), and those that are recipient-

proficient (having donor genes that were insufficient to confer the donor phenotype), the *mid* locus was mapped within *esx1* to six genes, *Ms0069-0071* and *Ms0076-0078* (Fig. 8). None of the encoded proteins have been functionally annotated, and so their putative role(s) in determining mating identity is currently unknown. The region spanning genes 0069–0071 is the most divergent part of the *esx1* locus between the donor and recipient. Orthologs of some these genes are present in sequenced environmental mycobacterial species, and to a lesser extent in the MTBC. More accurate predictions of potentially active DCT participants await a more definitive mapping and characterization of the pivotal *mid* activities in the existing *M. smegmatis* DCT model strains, or directed empirical testing of new strains for DCT.

How Do the Mid Proteins Confer Mating Identity?

The genetic mapping data and transposon mutagenesis data provide compelling evidence for a role for ESX-1 in DNA transfer. But what is that role and how can it be addressed? Transposon insertions and targeted deletions in recipient genes that map within *esx1*, but outside *mid*, abolish both DNA transfer and ESX-1 secretion, suggesting that there is a functional requirement for the apparatus *per se*. This requirement for ESX-1 likely reflects the need to secrete some of the Mid proteins to define mating identity. In support of this we, and others, have already shown Ms0076 (EspB) is secreted (69, 70), and have evidence that at least one other Mid protein, Ms0077, is secreted (J. Krywy, A. Collins, T.A. Gray & K.M. Derbyshire, unpublished data). Alternatively, Mid proteins could modify other ESX-1 substrates to exert an effect. Either way, any future mechanistic model of DCT is likely to begin with the Mid proteins, as they seem to dictate how a shared apparatus can have opposing activities in conjugation. Conceptually, communication between opposite mating types might be a good checkpoint activity to coordinate mating pair formation and donor genome mobilization to the waiting recipient.

SUMMARY AND FUTURE PROSPECTS FOR DCT

The work described in this chapter demonstrates the unique nature of conjugal DNA transfer in mycobacteria. Our knowledge has expanded dramatically since the chapter on conjugation in the first edition of this book. Perhaps the biggest advance is the realization of DCT's potential impact on all mycobacteria. In 2000, *M. tuberculosis* was considered clonal, evolving only by random genetic drift and selection, untouched by HGT. Mycobacterial conjugation was assumed to be similar to *E. coli* Hfr transfer, and restricted to *M. smegmatis* and, perhaps, other environmental mycobacteria.

Now, thirteen years later, DCT can be almost envisaged as a fourth mechanism of HGT, quite distinct from *oriT*-mediated conjugation. DCT requires a functional ESX-1 system, previously considered a virulence determinant. The essential role for ESX-1 in DCT has provided an ideal genetic model system to dissect ESX-1 functions. Genome sequence analyses have provided strong, albeit indirect, evidence for extensive HGT among the MTBC, as well as other mycobacteria. The mosaicism observed in some of these genomes is entirely consistent with it occurring by DCT, suggesting DCT is more prevalent among mycobacteria than previously appreciated. Together, these observations indicate that the MTBC is capable of recombination and that HGT has played a role in its evolution. We

anticipate additional examples from genome sequencing will further reinforce the ability of these species to diversify via HGT, in addition to genetic drift.

What might we expect in the next edition of this book? There are likely to be many advances, especially in defining the transfer genes, their protein structure and the mechanism of transfer. However, we speculate that the more likely global impact will be on the application of DCT to genetically engineer strains, and on understanding gene flow in mycobacteria and related species. Identification of the genes required for mating identity and those that mediate transfer will allow construction of defined mating pairs in all mycobacteria. Thus, DCT could be used to introduce defined mutations into the slow-growing pathogens from the genetically amenable fast-growing species, or to create attenuated, hybrid strains specifically designed to enhance live vaccine development. The similarity of DCT to meiotic recombination certainly opens the genetic doors, and in this era of genome-scale science, we predict DCT will continue to be a major contributor to mycobacterial genetics.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge input from all members of their laboratory over the years and the generous support of the NIAID.

References

1. Frost LS, Leplae R, Summers AO, Toussaint A. Mobile genetic elements: the agents of open source evolution. *Nat Rev Microbiol.* 2005; 3:722–732. [PubMed: 16138100]
2. Thomas CM, Nielsen KM. Mechanisms of, barriers to, horizontal gene transfer between bacteria. *Nat Rev Microbiol.* 2005; 3:711–721. [PubMed: 16138099]
3. Gogarten JP, Townsend JP. Horizontal gene transfer, genome innovation and evolution. *Nat Rev Microbiol.* 2005; 3:679–687. [PubMed: 16138096]
4. McDaniel LD, Young E, Delaney J, Ruhnau F, Ritchie KB, Paul JH. High frequency of horizontal gene transfer in the oceans. *Science.* 2010; 330:50. [PubMed: 20929803]
5. Nakamura Y, Itoh T, Matsuda H, Gojobori T. Biased biological functions of horizontally transferred genes in prokaryotic genomes. *Nat Genet.* 2004; 36:760–766. [PubMed: 15208628]
6. Ochman H, Lawrence JG, Groisman EA. Lateral gene transfer and the nature of bacterial innovation. *Nature.* 2000; 405:299–304. [PubMed: 10830951]
7. Wiedenbeck J, Cohan FM. Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches. *FEMS Microbiol Rev.* 2011; 35:957–976. [PubMed: 21711367]
8. Smith NH, Dale J, Inwald J, Palmer S, Gordon SV, Hewinson RG, Smith JM. The population structure of *Mycobacterium bovis* in Great Britain: clonal expansion. *Proc Natl Acad Sci U S A.* 2003; 100:15271–15275. [PubMed: 14657373]
9. Smith NH, Gordon SV, de la Rua-Domenech R, Clifton-Hadley RS, Hewinson RG. Bottlenecks and broomsticks: the molecular evolution of *Mycobacterium bovis*. *Nat Rev Microbiol.* 2006; 4:670–681. [PubMed: 16912712]
10. Sreevatsan S, Pan X, Stockbauer KE, Connell ND, Kreiswirth BN, Whittam TS, Musser JM. Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. *Proc Natl Acad Sci U S A.* 1997; 94:9869–9874. [PubMed: 9275218]
11. Supply P, Warren RM, Banuls AL, Lesjean S, Van Der Spuy GD, Lewis LA, Tibayrenc M, Van Helden PD, Locht C. Linkage disequilibrium between minisatellite loci supports clonal evolution

- of *Mycobacterium tuberculosis* in a high tuberculosis incidence area. *Mol Microbiol.* 2003; 47:529–538. [PubMed: 12519202]
12. Brosch R. A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. *Proceedings of the National Academy of Sciences.* 2002; 99:3684–3689.
 13. Guttierrez MC, Brisse S, Brosch R, Fabre M, Omais B, Marmiesse M, Supply P, Vincent V. Ancient origin and gene mosaicism of the progenitor of *Mycobacterium tuberculosis*. *PLOS pathogens.* 2005; 1:1–7.
 14. Supply P, Marceau M, Mangenot S, Roche D, Rouanet C, Khanna V, Majlessi L, Criscuolo A, Tap J, Pawlik A, Fiette L, Orgeur M, Fabre M, Parmentier C, Frigui W, Simeone R, Boritsch EC, Debrie AS, Willery E, Walker D, Quail MA, Ma L, Bouchier C, Salvignol G, Sayes F, Cascioferro A, Seemann T, Barbe V, Loch C, Gutierrez MC, Leclerc C, Bentley SD, Stinear TP, Brisse S, Medigue C, Parkhill J, Cruveiller S, Brosch R. Genomic analysis of smooth tubercle bacilli provides insights into ancestry and pathoadaptation of *Mycobacterium tuberculosis*. *Nat Genet.* 2013; 45:172–179. [PubMed: 23291586]
 15. Gray TA, Krywy JA, Harold J, Palumbo MJ, Derbyshire KM. Distributive Conjugal Transfer in *Mycobacteria* Generates Progeny with Meiotic-like Genome-Wide Mosaicism, Allowing Mapping of a Mating Identity Locus. *PLoS Biol.* 2013; 11
 16. de la Cruz F, Frost LS, Meyer RJ, Zechner EL. Conjugative DNA metabolism in Gram-negative bacteria. *FEMS Microbiol Rev.* 2010; 34:18–40. [PubMed: 19919603]
 17. Firth, N.; Ippen-Ihler, K.; Skurray, RA. Structure and Function of the F Factor and Mechanism of Conjugation. In: Neidhardt, FC., editor. *Escherichia coli and Salmonella Cellular and Molecular Biology.* 2nd ed. Vol. 2. Washington, DC: American Society for Microbiology; 1996. p. 2377–2401.
 18. Wozniak RA, Waldor MK. Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow. *Nat Rev Microbiol.* 2010; 8:552–563. [PubMed: 20601965]
 19. Guglielmini J, Quintais L, Garcillan-Barcia MP, de la Cruz F, Rocha EP. The repertoire of ICE in prokaryotes underscores the unity, diversity, and ubiquity of conjugation. *PLoS Genet.* 2011; 7:e1002222. [PubMed: 21876676]
 20. Christie PJ, Atmakuri K, Krishnamoorthy V, Jakubowski S, Cascales E. Biogenesis, architecture and function of bacterial TypeIV secretion systems. *Annual Review of Microbiology.* 2005; 59:451–485.
 21. Bhatti M, Laverde Gomez JA, Christie PJ. The expanding bacterial type IV secretion lexicon. *Res Microbiol.* 2013
 22. Grohmann E, Muth G, Espinosa M. Conjugative Plasmid Transfer in Gram- Positive Bacteria. *Microbiol Mol Biol Rev.* 2003; 67:277–301. [PubMed: 12794193]
 23. Alvarez-Martinez CE, Christie PJ. Biological diversity of prokaryotic type IV secretion systems. *Microbiol Mol Biol Rev.* 2009; 73:775–808. [PubMed: 19946141]
 24. Dunny GM. The peptide pheromone-inducible conjugation system of *Enterococcus faecalis* plasmid pCF10: cell-cell signalling, gene transfer, complexity and evolution. *Philos Trans R Soc Lond B Biol Sci.* 2007; 362:1185–1193. [PubMed: 17360276]
 25. Lederberg J, Tatum EL. Gene recombination in *E. coli*. *Nature.* 1946; 158:558. [PubMed: 21001945]
 26. Wollman EL, Jacob F, Hayes W. Conjugation and genetic recombination in *Escherichia coli* K-12. *Cold Spring Harb Symp Quant Biol.* 1956; 21:141–162. [PubMed: 13433587]
 27. Crawford, JT.; Falkinham, JO. Plasmids of the *Mycobacterium avium* complex. In: McFadden, J., editor. *Molecular Biology of the Mycobacteria.* San Diego: Academic Press Inc.; 1990. p. 97–120.
 28. Movahedzadeh F, Bitter W. Ins and outs of mycobacterial plasmids. *Methods Mol Biol.* 2009; 465:217–228. [PubMed: 20560075]
 29. Pashley, C.; Stoker, NG. Plasmids in mycobacteria. Washington DC: American Society for microbiology; 2000.
 30. Jucker MT, Falkingham JO. Epidemiology of infection by nontuberculous mycobacteria. *Am Rev Respir Dis.* 1990; 142:858–862. [PubMed: 2221593]

31. Kirby C, Waring A, Griffin TJ, Falkinham JO 3rd, Grindley ND, Derbyshire KM. Cryptic plasmids of *Mycobacterium avium*: Tn552 to the rescue. *Mol Microbiol*. 2002; 43:173–186. [PubMed: 11849545]
32. Picardeau M, Le Dante C, Vincent V. Analysis of the internal replication region of a mycobacterial linear plasmid. *Microbiology*. 2000; 146:305–313. [PubMed: 10708369]
33. Le Dantec C, Winter N, Gicquel B, Vincent V, Picardeau M. Genomic sequence and transcriptional analysis of a 23-kilobase mycobacterial linear plasmid: evidence for horizontal transfer and identification of plasmid maintenance systems. *J Bacteriol*. 2001; 183:2151–2164. [PubMed: 11222621]
34. Rauzier J, Moniz-Pereira J, Gicquel-Sanzey B. Complete nucleotide sequence of pAI5000, a plasmid from *Mycobacterium fortuitum*. *Gene*. 1988; 71:315–321. [PubMed: 3224826]
35. Stinear TP, Mve-Obiang A, Small PL, Frigui W, Pryor MJ, Brosch R, Jenkin GA, Johnson PD, Davies JK, Lee RE, Adusumilli S, Garnier T, Haydock SF, Leadlay PF, Cole ST. Giant plasmid-encoded polyketide synthases produce the macrolide toxin of *Mycobacterium ulcerans*. *Proc Natl Acad Sci U S A*. 2004; 101:1345–1349. [PubMed: 14736915]
36. Stinear TP, Pryor MJ, Porter JL, Cole ST. Functional analysis and annotation of the virulence plasmid pMUM001 from *Mycobacterium ulcerans*. *Microbiology*. 2005; 151:683–692. [PubMed: 15758215]
37. Wang J, Parsons LM, Derbyshire KM. Unconventional conjugal DNA transfer in mycobacteria. *Nat Genet*. 2003; 34:80–84. [PubMed: 12669068]
38. Leao SC, Matsumoto CK, Carneiro A, Ramos RT, Nogueira CL, Lima JD Jr, Lima KV, Lopes ML, Schneider H, Azevedo VA, da Costa da Silva A. The detection and sequencing of a broad-host-range conjugative IncP-1beta plasmid in an epidemic strain of *Mycobacterium abscessus* subsp. *bolletii*. *PLoS One*. 2013; 8:e60746. [PubMed: 23565273]
39. Gormley EP, Davies J. Transfer of plasmid RSF1010 by conjugation from *Escherichia coli* to *Streptomyces lividans* and *Mycobacterium smegmatis*. *Journal of Bacteriology*. 1991; 173:6705–6708. [PubMed: 1657866]
40. Parsons LM, Jankowski CS, Derbyshire KM. Conjugal transfer of chromosomal DNA in *Mycobacterium smegmatis*. *Mol. Micro*. 1998; 28:571–582.
41. Mizuguchi Y, Suga K, Tokunaga T. Multiple mating types of *Mycobacterium smegmatis*. *Japanese Journal of Microbiology*. 1976; 20:435–443. [PubMed: 994349]
42. Mizuguchi Y, Tokunaga T. Recombination between *Mycobacterium smegmatis* strains Jucho and *Lacticola*. *Japanese Journal of Microbiology*. 1971; 15:359–366. [PubMed: 5315708]
43. Derbyshire KM, Willetts NS. Mobilization of the non-conjugative plasmid RSF1010: a genetic analysis of its origin of transfer [published erratum appears in *Mol Gen Genet* 1987 Sep;209(2): 411]. *Molecular & General Genetics*. 1987; 206:154–160. [PubMed: 3033437]
44. Everett R, Willetts NS. Cloning, mutation and location of the origin of conjugal transfer. *EMBO Journal*. 1982; 1:747–753. [PubMed: 6329701]
45. Wang J, Derbyshire KM. Plasmid DNA transfer in *Mycobacterium smegmatis* involves novel DNA rearrangements in the recipient, which can be exploited for molecular genetic studies. *Mol Microbiol*. 2004; 53:1233–1241. [PubMed: 15306024]
46. Wang J, Karnati PK, Takacs CM, Kowalski JC, Derbyshire KM. Chromosomal DNA transfer in *Mycobacterium smegmatis* is mechanistically different from classical Hfr chromosomal DNA transfer. *Mol Microbiol*. 2005; 58:280–288. [PubMed: 16164565]
47. Daffe, M.; Reyat, JM. *The Mycobacterial Cell Envelope*. Washington, DC: ASM Press; 2008.
48. Flint JL, Kowalski JC, Karnati PK, Derbyshire KM. The RD1 virulence locus of *Mycobacterium tuberculosis* regulates DNA transfer in *Mycobacterium smegmatis*. *Proc Natl Acad Sci U S A*. 2004; 101:12598–12603. [PubMed: 15314236]
49. Coros A, Callahan B, Battaglioli E, Derbyshire KM. The specialized secretory apparatus ESX-1 is essential for DNA transfer in *Mycobacterium smegmatis*. *Mol Microbiol*. 2008; 69:794–808. [PubMed: 18554329]
50. Abdallah AM, Gey van Pittius NC, DiGiuseppe Champion PA, Cox J, Luirink J, Vandenbroucke-Grauls CMJE, Appelmek BJ, Bitter W. Type VII secretion — mycobacteria show the way. *Nature Reviews Microbiology*. 2007; 5:883–891.

51. Bitter W, Houben EN, Bottai D, Brodin P, Brown EJ, Cox J, Derbyshire KM, Fortune SM, Gao LY, Liu J, Gey van Pittius NC, Pym AS, Rubin EJ, Sherman DR, Cole ST, Brosch R. Systematic genetic nomenclature for type VII secretion systems. *PLoS Pathog.* 2009; 5:e1000507. [PubMed: 19876390]
52. Converse SE, Cox JS. A protein secretion pathway critical for *Mycobacterium tuberculosis* virulence is conserved and functional in *Mycobacterium smegmatis*. *J Bacteriol.* 2005; 187:1238–1245. [PubMed: 15687187]
53. Smith GR. Conjugational recombination in *E. coli*: myths and mechanisms. *Cell.* 1991; 64:19–27. [PubMed: 1986865]
54. Taylor AF, Smith GR. RecBCD enzyme is altered upon cutting DNA at a chi recombination hotspot. *Proc Natl Acad Sci U S A.* 1992; 89:5226–5230. [PubMed: 1535156]
55. Sinha KM, Unciuleac MC, Glickman MS, Shuman S. AdnAB: a new DSB-resecting motor-nuclease from mycobacteria. *Genes Dev.* 2009; 23:1423–1437. [PubMed: 19470566]
56. Gupta R, Barkan D, Redelman-Sidi G, Shuman S, Glickman MS. Mycobacteria exploit three genetically distinct DNA double-strand break repair pathways. *Mol Microbiol.* 2011; 79:316–330. [PubMed: 21219454]
57. Warner DF, Mizrahi V. Making ends meet in mycobacteria. *Mol Microbiol.* 2011; 79:283–287. [PubMed: 21219450]
58. Rayssiguier C, Thaler DS, Radman M. The barrier to recombination between *Escherichia coli* and *Salmonella typhimurium* is disrupted in mismatch-repair mutants. *Nature.* 1989; 342:396–401. [PubMed: 2555716]
59. Springer B, Sander P, Sedlacek L, Hardt WD, Mizrahi V, Schar P, Bottger EC. Lack of mismatch correction facilitates genome evolution in mycobacteria. *Mol Microbiol.* 2004; 53:1601–1609. [PubMed: 15341642]
60. Shuman S, Glickman MS. Bacterial DNA repair by non-homologous end joining. *Nat Rev Microbiol.* 2007; 5:852–861. [PubMed: 17938628]
61. Possoz C, Ribard C, Gagnat J, Pernodet JL, Guerineau M. The integrative element pSAM2 from *Streptomyces*: kinetics and mode of conjugal transfer. *Mol Microbiol.* 2001; 42:159–166. [PubMed: 11679075]
62. Vogelmann J, Ammelburg M, Finger C, Guezguez J, Linke D, Flotenmeyer M, Stierhof YD, Wohlleben W, Muth G. Conjugal plasmid transfer in *Streptomyces* resembles bacterial chromosome segregation by FtsK/SpoIIIE. *EMBO J.* 2011; 30:2246–2254. [PubMed: 21505418]
63. Lee JY, Finkelstein IJ, Crozat E, Sherratt DJ, Greene EC. Single-molecule imaging of DNA curtains reveals mechanisms of KOPS sequence targeting by the DNA translocase FtsK. *Proc Natl Acad Sci U S A.* 2012; 109:6531–6536. [PubMed: 22493241]
64. Reuther J, Gekeler C, Tiffert Y, Wohlleben W, Muth G. Unique conjugation mechanism in mycelial streptomycetes: a DNA-binding ATPase translocates unprocessed plasmid DNA at the hyphal tip. *Mol Microbiol.* 2006; 61:436–446. [PubMed: 16776656]
65. Sepulveda E, Vogelmann J, Muth G. A septal chromosome segregator protein evolved into a conjugative DNA-translocator protein. *Mob Genet Elements.* 2011; 1:225–229. [PubMed: 22479692]
66. Namouchi A, Didelot X, Schock U, Gicquel B, Rocha EP. After the bottleneck: Genome-wide diversification of the *Mycobacterium tuberculosis* complex by mutation, recombination, and natural selection. *Genome Res.* 2012; 22:721–734. [PubMed: 22377718]
67. Krzywinska E, Krzywinski J, Schorey JS. Naturally occurring horizontal gene transfer and homologous recombination in *Mycobacterium*. *Microbiology.* 2004; 150:1707–1712. [PubMed: 15184557]
68. Wang J, Fan HC, Behr B, Quake SR. Genome-wide single-cell analysis of recombination activity and de novo mutation rates in human sperm. *Cell.* 2012; 150:402–412. [PubMed: 22817899]
69. McLaughlin B, Chon JS, MacGurn JA, Carlsson F, Cheng TL, Cox JS, Brown EJ. A *Mycobacterium* ESX-1-Secreted Virulence Factor with Unique Requirements for Export. *PLOS pathogens.* 2007; 3:e105. [PubMed: 17676952]
70. Xu J, Laine O, Masciocchi M, Manoranjan J, Smith J, Du SJ, Edwards N, Zhu X, Fenselau C, Gao L-Y. A unique *Mycobacterium* ESX-1 protein co-secreted with CFP-10/ESAT-6 and is necessary

for inhibiting phagosome maturation. *Molecular Microbiology*. 2007; 66:787–800. [PubMed: 17908204]

FIGURE 1.

oriT directs both plasmid and chromosomal transfer in traditional conjugation systems. (A) F episome plasmid transfer. The *oriT* is nicked and a single strand is guided into the engaged recipient by a plasmid-encoded relaxase. Upon transfer, relaxase catalyzes recircularization at *oriT*, and host polymerases synthesize a complementary strand. The recipient chromosome is unaltered, but the cell now exhibits a donor phenotype (blue). (B) Hfr strains have a plasmid integrated into the chromosome, shown as a single strand for simplicity. The integrated *oriT* functions as it would in the plasmid scenario above, except that transfer of the chromosome is usually incomplete. The linear chromosomal fragment must be incorporated into the recipient chromosome by homologous recombination for stable inheritance. Homologous recombination excludes the *oriT*, and, while the transconjugant chromosome now has some donor sequences, it retains a recipient phenotype (yellow). For an Hfr transconjugant to become a donor, the entire donor chromosome must be transferred to regenerate *oriT*.

Mated Strains	$Km^r + Sm^r$ Colonies	Transcojugants per Donor
Donor only	0	$<1 \times 10^{-9}$
Recipient only	0	$<1 \times 10^{-9}$
Donor x Recip	8×10^4	1×10^{-4}

FIGURE 2.

An outline of the mycobacterial mating procedure. Selectable markers for donor and recipient are chromosomally encoded. Efficient transfer requires prolonged incubation (overnight) on solid medium. Transfer does not take place in liquid cultures, probably reflecting the need to force cell-cell contact on the solid medium. The transfer frequency is 1 event per 10^4 donors. However, this only reflects successful transfer of Km^r and therefore is an underestimate of the total number of events.

FIGURE 3.

Donor chromosomal *bom* sites mobilize plasmids and mediate recircularization by gap-repair in the recipient. Episomal plasmids are not subject to conjugal transfer unless they carry chromosomal DNA segments functionally defined as *bom* (basis of mobility) sites. (A) Recovery of the transferred plasmids, and sequencing of the *bom* sites, revealed the presence of embedded recipient SNPs, suggesting a gap-repair mechanism. In this model, transfer would be initiated in the donor via a break in *bom*. Following transfer of the linear plasmid, the homologous region of the recipient chromosome would act as the template for gap-repair to seal the break and recircularize the plasmid. As a consequence, recipient SNPs would be incorporated into the plasmid DNA. (B) This model was confirmed by the use of two

adjacent *bom* sites, separated by a non-homologous sequence. The region of non-identity was replaced by the intervening recipient chromosomal sequence upon transfer. In this two-*bom* model a break could occur at both *boms*, to create a gap spanning hypothetical gene *b*. Alternatively, a break could occur at just one *bom* site (as shown) and then the non-homologous region resected by exonucleases to generate ends suitable for gap repair, capture of gene *b* and plasmid circularization.

FIGURE 4.

Transconjugant chromosomes are mosaic blends of the parental strains. Whole genome sequencing of transconjugants and alignment with the parental sequences used the presence or absence of parental SNPs to define tracts of transferred donor DNA. A Circos projection of the 7 Mb chromosome is shown as concentric circles, with the recipient genome on the outside (yellow), the donor on the inside (blue), and four transconjugants between. The segment containing the *Km^r* gene used for post-mating selection is indicated (green); the recipient antibiotic marker is episomally encoded. The outer three transconjugants have a

donor mating identity, and all have inherited the *esxI* locus from the donor strain (indicated at 0.1 Mb on the chromosome). The inner most transconjugant lacks the *esxI* locus and is a recipient strain. These data and additional Circos plots can be seen in (15).

FIGURE 5.

Models for chromosome fragmentation in DCT. DCT could follow one of two pathways depending on whether fragmentation of the transferred chromosome occurs in the donor cell before transfer (left), or following transfer in the recipient (right and indicated by arrows). The left hand model posits that multiple chromosome segments are co-transferred into the recipient, where they are recombined into the recipient chromosome to generate a mosaic pattern. Large chunk transfer (right) would be predicted to result in fewer large integrated

segments rather than many widely distributed donor segments. Current results, as seen in Fig. 4, are more consistent with the fragmentation-before-transfer model.

FIGURE 6.

DCT brings large and small changes to transconjugant genomes. Transferred donor segments can be contiguous blocks hundreds of kilobases in length, spanning hundreds of genes. These large blocks can exchange recipient for donor orthologues (depicted at left). The large segments may also contain additional genes not present in the original recipient sequence (insertion), or may lack some genes that had been present (deletion). At the opposite end of the size spectrum, donor segments of <100 bp are often found in clusters to generate microcomplexity, with the potential to fine-tune genes or functional elements (depicted at right). Comparison of *M. canettii* genomes has identified similar mosaicism, in which sequences of entire genes are identical between some isolates, while other regions contain short regions of exchanged SNPs consistent with the recombinant patterns observed with DCT (14).

FIGURE 7.

DCT generates instant diversity on a genome wide scale. Progeny from a bacterial cross are shown for *oriT*-mediated chromosomal transfer (left) and for DCT (right). In a single Hfr-cross, transconjugants can acquire segments of DNA proximal to *oriT*. Depending on the length of transfer, all progeny will have overlapping regions of the donor chromosome extending from *oriT*. By contrast, all DCT progeny are different as transfer initiates from sites all around the chromosome. In addition, DCT recombination results in both large segment exchanges and smaller regions of micro-heterogeneity creating further diversity.

The extreme diversity in the transconjugant population allows for rapid expansion under changing selective pressures.

FIGURE 8.

Mating identity is determined by genes within *esx1*. The *M. smegmatis* *esx1* locus spans 26 homologous genes in the donor (blue) and recipient (orange) parental strains (top). Genes *espE* and *MycP1* define the ends of the locus. *esxA* and *esxB* encode the primary secreted substrates EsxA and EsxB. Whole genome sequencing and genome-wide association mapping of F1 transconjugants that exhibited a donor phenotype revealed that all donor-proficient transconjugants had an *esx1* locus of donor origin (Fig. 4). Mapping of the mating identity (*mid*) locus—the donor genes associated with the donor conjugal phenotype—to *esx1* was confirmed by successively backcrossing donor-proficient transconjugants with the recipient, while maintaining the donor phenotype (transconjugant donor). Backcrossed transconjugants that showed a recipient conjugal phenotype (transconjugant recipient) all shared a region of their *esx1* locus that was composed of recipient-derived genes, further refining the *mid* locus to six genes (0069–0078) near the 3' end of *esx1*. White-filled gene symbols represent repetitive elements.