University of Alberta

Evolution and Assembly of a Plasmid DNA Conjugation Apparatus

by

Matthew Warren Gilmour 

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Bacteriology

Department of Medical Microbiology and Immunology

Edmonton, Alberta
Spring, 2004

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Library and Archives Canada

Published Heritage Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et Archives Canada

Direction du Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 0-612-96272-5
Our file Notre référence
ISBN: 0-612-96272-5

The author has granted a non-exclusive license allowing the Library and Archives Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou aturement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

ABSTRACT

Intercellular transfer of plasmid DNA molecules is achieved by a large array of proteins which assemble into both cytoplasmic and membrane-associated complexes. The incompatibility group H plasmids R478, R27, and pHCM1 are large conjugative genetic elements (≥180 kbp), are harboured within pathogenic bacteria, and encode resistance to multiple antibiotics and heavy metals. The evolution and function of the H plasmid transfer apparatus was investigated by comparative sequence analyses, fluorescent labeling of transfer protein complexes, and biochemical characterization of protein interactions.

The 274,762 bp sequence of R478 was determined and used to define the minimal set of core H plasmid determinants (including replication, partitioning, and conjugative transfer) by a comparative analysis with R27 and pHCM1 sequences. No resistance determinants are predicted to be present in the ancestor common to these plasmids, and a comparison with related transfer sequences suggested that the H-type transfer determinants encoding mating pair formation (Mpf) functions are of a different lineage than those determinants encoding DNA-processing (Dtr) and membrane-targeting (coupling) functions.

An assembly of R27-encoded proteins was visualized through an Mpf protein-green fluorescent protein fusion, which appeared as discrete membrane-associated fluorescent foci. Thirteen of the nineteen R27 transfer proteins were required for focus formation, and an individual focus possibly represents a subassembly comprised of some or all of these transfer proteins. These data support the notion that the transfer apparatus is a multi-component structure. The fluorescence assay also demonstrated that the NTP-binding motifs of the putative ATPase TrhC were necessary for conjugative transfer, but not for assembly of the fluorescent subcomplex, and that the temperature-sensitive

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

phenotype of H plasmid transfer results from temperature-dependent expression of R27 transfer genes.

The proposed link between the Mpf and Dtr protein complexes is the coupling protein. This model was confirmed by bacterial two-hybrid and immunoprecipitation assays which characterized a direct interaction between the R27-encoded coupling protein TraG and the Mpf protein TrhB. Evolution of the H-type transfer system required the adaptation of unrelated Mpf and Dtr determinants by the coupling protein to form an apparatus functional for conjugative plasmid transfer.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

'Dealing with ancient events of necessity involves conjecture' – R.F. Doolittle

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

For my parents, Warren and Janice Gilmour,
and grandparents, George and Mildred Boyce

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Acknowledgements

Thank you to my friends Neil and Angela Matheson, James and Karen Gunton, Dobryan Tracz, and Paola Marcato. The difficulties associated with grad school were shadowed by your presence. I would like to especially thank Sean Connell, Dave Rasko and Trevor Lawley, firstly for their friendship, but also for the guidance and leadership they provided. I am indebted to each of you for the successes that I've had. Also thank you to Chris Mackenzie, Michelle Rooker, Qin Jiang, Gerald Audette, Sharee Kuny, Cathy Trieber, Richard Rothery, Richard Sherburne, Peter Newnham, Ge Wang, George Mulvey, Anne Giles, Nancy Elsinga, Anita-Marie Reff, Christian Rabel, Peter Christie, and Fernando de la Cruz for sharing your time and knowledge with me. Thank you to Jamil Janmohamed and Andrew Ting, two great summer students. Thank you Donald Court, Benjamin Gibbard, Isaac Brock, Jarvis Cocker, Ted Leo, and Daniel Ladant who provided the materials essential for the completion of this thesis. Thank you members of the Sanger Institute pathogen sequencing unit, including Nick Thomson, for teaching me the power of computers in biology and for their contributions to chapter 2.

To my supervisor and mentor, Diane Taylor, it is from you that I have learned the most. My writing, presentation, and traveling skills have been well honed through your support. To the members of my examining committee (past and present): Laura Frost, Glen Armstrong, Markus Stein, Bart Hazes, Jon Dennis, and Anath Das; thank you for your enthusiasm, support and feedback. Any errors remaining in this thesis are my own.

To the Alberta Heritage Foundation for Medical Research, Canadian Institutes of Health Research, and the Andrew Stewart Memorial Graduate Student Prize and Izaak Walton Killam Memorial Scholarship selection committees; thank you for the recognition and financial support.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

To the city of Edmonton: the body of the thesis was written on a park bench in the river valley with a dull pencil, and I thank you for the sun, fresh air, and solitude. I was never rained upon once.

Thank you to my parents, Warren and Janice; my Nan and Grandpa, Mrs. Mildred Boyce and Mr. George Boyce; and my aunts and uncles, Kevin Boyce, Lynn and David Tomlin, Lois and Don Ellis and Irene Jackson. If there is any quality in this work, it was inspired by your ideals. To my brother, Ryan, who also found an unexpected profession: 'strength, courage, integrity, and respect', indeed. And to my bride, Sue: without your love and encouragement, the weight of completing this thesis would have been overwhelming. You are beautiful, and I love you.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

TABLE OF CONTENTS

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

LIST OF TABLES

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

LIST OF FIGURES

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

List of Abbreviations

Ap	ampicillin
BLAST	basic local alignment search tool
bp	base pairs
BTH	bacterial two-hybrid
CDS	coding sequence
Cm	chloramphenicol
DNA	deoxyribonucleic acid
dr	derepressed
Dtr	DNA transfer replication
DTT	dithiothreitol
FPLC	fast protein liquid chromatography
GFP	green fluorescent protein
HRP	horseradish peroxidase
IPTG	isopropyl-β-D-thiogalactoside
Inc	incompatibility group
Km	kanamycin
LB	Luria-Bertani
MDR	multiple drug resistant
Mpf	Mating pair formation
Nal	nalidixic acid
NTP	nucleotide triphosphate
NCBI	National Center for Biotechnology Information
OD	optical density
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PSI-BLAST	position specific iterative – basic local alignment search tool
Rif	rifampicin
RNA	ribonucleic acid
SDS-PAGE	sodium dodecyl sulphate – polyacrylamide gel electrophoresis
T3S	Type III secretion
T4S	Type IV secretion
Tc	tetracycline

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

TMH	transmembrane helix
Tra1	transfer region 1
Tra2	transfer region 2
Tris	Tris (hydroxymethyl) aminomethane
UV	ultraviolet
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Chapter 1

General Introduction

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

1. General Introduction

1.1 The tree of life

The three domains of life include the Eucarya, the Bacteria, and the Archaea, with members of the latter two domains collectively called prokaryotes (Woese *et al.*, 1990). A visual representation of this classification scheme presents each domain as being derived from a universal common ancestor, and an absence of genetic exchange since the divergence of these lineages. This phylogenetic structure is called into question by the growing evidence from genome sequencing projects that transfer of genetic material occurs between members of the same domain, and between members of different domains (Doolittle *et al.*, 2003). General examples include the observation of Archaeal coding sequences in Bacterial genomes, or vice versa. It is possible that orthologous genes present in multiple domains may have been inherited from the universal common ancestor, however if true for all such circumstances, this theoretical ancestral genome would exceed a reasonable size (Koonin, 2003). Instead, it is more plausible that some genes have independently evolved and subsequently transferred between genomes. Such events represent lateral or horizontal gene transfer, and these blur the lines that comprise the universal tree of life.

Gene transfer events have predominantly been reported between prokaryotic genomes (Doolittle *et al.*, 2003). This bias is likely a result of the uni-cellular nature and reproductive strategy of prokaryotes (Lawrence, 2002). Eukaryotes propagate in most instances through sexual reproduction, wherein two distinct meiotic genomes are combined to create a progeny unique from (but representative of) the parental genomes. Conversely, prokaryotic cells generate progeny through binary fission after faithful replication of their own chromosomal and extra-chromosomal genetic elements, without the requirement of genetic input from a second parent cell. In addition, cell fusion, as required during eukaryotic reproduction is prohibited by the peptidoglycan exoskeleton

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

possessed by bacterial cells (Cavalier-Smith, 2002). In the absence of gene transfer, the only diversity introduced into a prokaryotic genome would be errors (i.e. point mutations) or small re-arrangements introduced during DNA replication. The evolutionary force of point mutations to create new species would eventually be seen over the course of millennia, but the immediate effect would be genomic stasis. Alternatively, the integration of 'foreign' genetic determinants (possibly encoding novel functions) into the genome of a prokaryote would result in the inheritance of those determinants after a single generation. For a multi-cellular organism, germ line cells would be the only recipients which would perpetuate transferred genetic information. Furthermore, the only confirmed eukaryotic recipients of trans-domain transfer events have been protists (Doolittle *et al.*, 2003).

The degree of horizontal gene transfer amongst prokaryotes lies somewhere between two extremes: the complete absence of transfer, or a singular 'bacterial superbiosystem' (Sonea and Mathieu, 2001) in which all genes might be freely accessible and transferable. Numerous limitations to gene transfer prevent the occurrence of the latter possibility, including geographical barriers, the inefficiency of recombination between unrelated sequences, the bacterial mis-match repair system, and incompatibility of genetic features such as promoter and termination sequences and codon usage biases between different organisms (Lawrence and Hendrickson, 2003; Majewski *et al.*, 2000). The most obvious physical barrier which must be overcome during horizontal transfer is the bacterial envelope, and there are multiple strategies to transfer DNA molecules into bacterial cells.

1.2 Mechanisms of horizontal gene transfer

The three known mechanisms of horizontal genetic exchange are transduction (by bacteriophage), transformation (by naked DNA), and conjugation (by plasmids) – each

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

capable of transmitting DNA across the bacterial cell envelope. A now-classic example of the evolutionary consequences of horizontal gene transfer is the conversion of the benign bacterium *Escherichia coli* to a pathogenic strain through the acquisition of pathogenicity determinants (Perna *et al.*, 2001). It was estimated that 26% (1.34 Mbp) of the genome of the reference enterohemorrhagic O157:H7 strain *E. coli* EDL933 is not present in the lab strain K-12, and may represent horizontally transferred sequences (Perna *et al.*, 2001). This includes 338 genes that are predicted to be of bacteriophage origin, and in particular, the Shiga-like toxin (Stx2) sequence represents such a prophage determinant (Datz *et al.*, 1996; Perna *et al.*, 2001). Other pathogenicity determinants, such as the locus for enterocyte effacement present in numerous pathogenic *E. coli* stains, and additional adhesion, toxin and secretion systems have also been acquired by mechanisms of horizontal gene transfer (Perna *et al.*, 2001).

The evolution of pandemic *Vibrio cholerae* strains also resulted after transduction of pathogenicity determinants (Faruque and Mekalanos, 2003). *V. cholerae* strain O139 encodes numerous phage pathogenicity islands, including the CXTϕ element which carries the cholera toxin genes (Waldor and Mekalanos, 1996). As bacteriophage cycle through bacterial hosts, it is possible that DNA from the current host bacterium is packaged and transmitted to subsequent hosts, as well as the phage DNA (Birge, 1994). Such gene transfer events represent transduction, and the range of gene transfer is limited by the amount of DNA packaged by the phage and the different bacteria which can be infected (Ochman *et al.*, 2000). During transduction, bacteriophage may recognize specific features on, and gain entry into the cell via the extracellular pili or the associated pilus assembly apparatus (Grahn *et al.*, 1997; Paranchych, 1975). The genomic sequence data from Proteobacteria (including *E. coli* and *V. cholerae*) reveal that in most instances multiple (7-20) prophages are present (Casjens, 2003).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

The first example of genetic exchange, and one of the seminal experiments to demonstrate that DNA stores genetic information, was the natural transformation and associated competence of *Streptococcus pneumonia* (Griffith, 1928). Natural transformation is the process by which bacterial cells are able to take up and maintain extracellular DNA without aid of chemical treatments (i.e. $CaCl_2$-induced competence) (Dubnau, 1999). Because DNA is sampled from the environment, there is the possibility of acquiring distantly related genetic determinants (Ochman *et al.*, 2000). Limitations to this process include the recognition of specific uptake sequences (although this is not observed in all competence systems) and the capacity to incorporate heterologous sequences (Dubnau, 1999). The transformation proteins are similar to the pilus assembly proteins or bacterial conjugation proteins, each of which form multi-component transenvelope structures (Dubnau, 1999; Lawley *et al.*, 2003b). In particular, the competence system of the pathogens *Helicobacter pylori* and *Campylobacter jejuni* are related to proteins of the conjugative transfer apparatus (Bacon *et al.*, 2002; Hofreuter *et al.*, 2001).

The evolutionary significance of horizontal transfer was first highlighted by observations that antibiotic resistance determinants were spread through the bacterial gene pool by plasmid DNA molecules, rather than *de novo* synthesis of these features in each resistant isolate (Davies, 1994; Watanabe, 1963). These extrachromosomal genetic elements, capable of intercellular transfer, may confer upon the host cell (and recipients of a transfer event) several resistance phenotypes, including resistance to antibiotics, toxic heavy metals, colicins and bacteriophage. Notably, most pre-antibiotic era bacteria were not found to contain transferable resistance plasmids (Hughes and Datta, 1983), therefore the spread of resistance determinants to pathogens by conjugative plasmids seems to have occurred in response to human activity.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Among the essential features of conjugative plasmids are gene products and cis-acting elements for plasmid replication, partitioning and conjugal transfer, which are collectively termed the backbone or core functions (Thomas, 2000). Through these components, bacterial plasmids are both stable and transmissible within bacterial populations, and contribute to bacterial genetic exchange. Conjugative transfer is achieved by a macromolecular assembly of proteins which span the bacterial envelope and terminate in the extracellular space with a filamentous pilus.

1.3 Type IV secretion systems

Bacterial conjugation, in addition to being a mechanism of genome evolution, can be considered as a mechanism of macromolecular secretion. In particular, plasmid conjugative transfer is classified as a type IV secretion (T4S) system and represents the only known bacterial system for secretion of DNA (Christie, 2001). Bacteria not only transfer genetic information, such as determinants encoding antibiotic and heavy metal resistance using T4S systems; but also transfer virulence determinants into eukaryotic cells using these machines. Despite being encoded by a diverse range of bacterial genomes, each transferring a different protein or nucleoprotein substrate, these multi-component modules are each classified as T4S systems because there is conservation of both transfer gene sequences and organization (Christie, 2001; Nagai and Roy, 2003). In particular, components of the membrane-associated conjugative apparatus required for plasmid DNA transfer are similar to proteins responsible for the intercellular translocation of virulence factors from pathogenic Gram negative bacteria (Baron *et al.*, 2002).

Those T4S systems involved in pathogenesis have been subcategorized as 'effector translocators' by Cascales and Christie (2003). An increasing number of such effector translocation systems have been reported recently, including the prototypical secretion

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

of CagA by *H. pylori* which results in a hummingbird phenotype (through cytoskeletal rearrangements) in epithelial cells (Segal *et al.*, 1999), pertussis toxin secretion by *Bordetella pertussis* resulting in pleotropic effects after delivery to mammalian cells (Farizo *et al.*, 2000), and plant oncogenesis mediated by the T-DNA nucleoprotein complex comprising proteins VirD2 and E2 secreted by *Agrobacterium tumefaciens* (Vergunst *et al.*, 2000). Other systems include the secretion of RalF by *Legionella pneumophila*, and unknown effectors released by *Bartonella* and *Brucella* spp., each necessary to maintain the lifestyle of these intracellular pathogens (Nagai *et al.*, 2002; O'Callaghan *et al.*, 1999; Schmiederer *et al.*, 2001; Seubert *et al.*, 2003b).

The T4S effector translocation systems have been compared to the type 3 secretion (T3S) systems, also involved in delivering protein effectors during pathogenesis (Cascales and Christie, 2003). Bacteria encoding T3S systems have been visualized with a needle-like extracellular superstructure that is anchored in the cell envelope by a structure resembling the flagellar basal body (Blocker *et al.*, 2001; Kubori *et al.*, 1998). Until recently, the only visible structure associated with T4S was the pilus. However, microscopic studies of *H. pylori* have revealed a filamentous structure having a central channel and covered in a sheath comprised of proteins related to *Agrobacterium* T4S proteins (Rohde *et al.*, 2003; Tanaka *et al.*, 2003). The apparent polar localization of this appendage on *H. pylori* cells also corresponds with the polar localization of the *Agrobacterium* pilus and T4S proteins (Kumar and Das, 2002; Lai *et al.*, 2000; Rohde *et al.*, 2003). The *H. pylori* filamentous appendage may be functionally analogous to the pilus, and have roles in cellular adhesion and/or substrate transfer.

T4S systems can also contribute to pathogenesis by releasing or transferring DNA which encodes virulence determinants. The human pathogen *Neisseria gonorrhoaea* is naturally competent for DNA uptake, and DNA is released from donor gonococcus by a putative T4S system homologous to the F plasmid transfer system. These determinants

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

are encoded on a genomic island which also encodes a peptidoglycan hydrolase that likely contributes to inflammation by producing a peptidoglycan cytotoxin (Dillard and Seifert, 2001; Ding *et al.*, 2003; Hamilton *et al.*, 2001). This putative T4S system may be responsible for the transmission of this genomic island and other determinants amongst *Neisseria* spp. since the T4S and hydrolase determinants were each necessary for DNA secretion (Dillard and Seifert, 2001; Hamilton *et al.*, 2001). Conversion of avirulent *L. pneumophila* to a virulent strain was also reported to occur as a result of a T4S process (Miyamoto *et al.*, 2003). The *icm-dot* determinants (encoding T4S effector translocation systems) were transferred by a conjugative mechanism, and these data further highlight the evolutionary contributions made by horizontal gene transfer mechanisms.

1.4 Bacterial conjugation

Conjugative resistance plasmids are the paradigm for horizontal transfer between bacterial cells. In order to process and transport plasmid DNA, a multitude of cytosolic and membrane-associated proteins are required. Conjugal transfer proteins are believed to form at least three macromolecular complexes: the transenvelope mating pair formation (Mpf) complex, the cytosolic proteins responsible for initiating DNA transfer and replication (Dtr), and the coupling protein which is proposed to localize the Dtr components to the Mpf complex at the cell membrane through protein interactions with each complex (Schroder *et al.*, 2002; Willetts, 1981). In a widely accepted model for plasmid transfer the Mpf complex is responsible for production of the conjugative pili, which, after binding to an appropriate recipient cell, would retract via subunit depolymerization, thereby bringing both donor and recipient cells into close contact (Curtiss, 1969; Ou and Anderson, 1970). The exact nature of the conjugative pore through which DNA can traverse the cell wall of both donor and recipient bacteria is unknown.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

1.4.1 Mating pair formation proteins

The Mpf complex represents a protein assembly analogous to the effector translocation T4S systems. This superstructure is hypothesized to contain both inner and outer membrane proteins, as well as periplasmic proteins, and could therefore span the cell envelope. Known functions of the Mpf complex include the production of conjugative pili and the entry of bacteriophage. The Mpf complex may also represent the route through which DNA traverses the cell envelope of the donor cell during conjugation. The associated pilus may be involved solely in attachment/recognition of recipient cells, or may additionally act in a needle-like fashion and inject the transferring DNA through the recipient cell wall.

The Mpf proteins encoded by plasmids F and RP4 are some of the best characterized of the known conjugative systems (Frost *et al.*, 1994; Lessl *et al.*, 1992a). Lanka and colleagues have identified 11 essential Mpf proteins by determining the role of each gene for RP4 transfer, mobilization of the non-self-transmissible plasmid RSF1010, pilus production, and phage propagation (Grahn *et al.*, 1997; Haase *et al.*, 1995; Lessl *et al.*, 1993). These same Mpf proteins, with the exception of TrbB, have been localized to a membrane fraction of intermediate density between inner and outer membrane fractions in cells containing RP4 or a Mpf-producing derivative, suggesting that the Mpf proteins assemble into a complex capable of joining these two bacterial membranes (Grahn *et al.*, 2000). Additionally, production of the core Mpf proteins in donor cells results in an electron-dense interface (termed conjugative junction) between cells, as observed during transmission electron and light microscopy (Durrenberger *et al.*, 1991; Samuels *et al.*, 2000). This likely represents the route through which DNA passes to the recipient cell, as there is cell-cell contact mediated by the Mpf complex. Furthermore, the point of fusion between mating pairs of donor and recipient bacteria is

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

located along the lateral wall or at the extreme ends of either cell (Lawley *et al.*, 2002b). These data suggest that sites for plasmid transfer are present throughout the cell periphery.

1.4.2 DNA transfer replication components

Processing of plasmid DNA prior to and during transfer resembles plasmid rolling circle replication (Lawley, 2003), as two double-stranded copies of a plasmid are ultimately produced after cleavage of the parental plasmid to generate two single-stranded templates. One of these single DNA strands is transferred to the recipient cell during bacterial conjugation, and the other strand remains in the donor cell. Processing of the plasmid to create the single-stranded transfer intermediate is the role of the relaxase – a member of the cytoplasmic relaxosome. This complex is also comprised of other DNA-binding proteins (such as integration host factor: IHF) which act at the origin of transfer (*oriT*), and cumulatively represent the Dtr aspects of bacterial conjugation (Nunez and De La Cruz, 2001). The *oriT* is generally an AT rich segment represented as a tandem of repeats containing a *nic* site where the relaxase cleaves a single strand of the DNA duplex. The nucleophilic tyrosine of the relaxase remains covalently bound to the 5' end of the cleaved strand (which will serve as the transferred strand), however an equilibrium between cleaved and uncleaved states exists, and the plasmid retains superhelicity (Guasch *et al.*, 2003; Pansegrau and Lanka, 1996). If a transfer event is initiated, the remaining free 3'-OH of the cleaved strand serves to initiate replacement strand synthesis using the host replication machinery and the uncleaved strand as a template. After transfer of the single DNA strand to the recipient is complete, both the replacement strand and transferred strand are recircularized by the relaxase (Zechner *et al.*, 2000). In the recipient cell, the transferred strand is converted into a double-stranded molecule after synthesis of nucleotide primers complementary to the

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

transferred strand, and containing a free 3'-OH from which complementary strand synthesis is initiated (Becker and Meyer, 2000; Pansegrau and Lanka, 1996).

1.4.3 Type IV coupling proteins

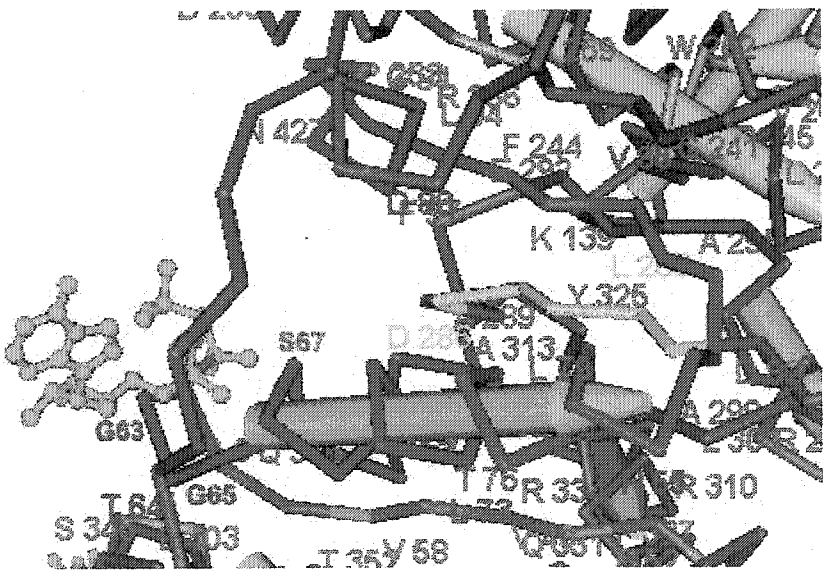
The DNA/relaxosome nucleoprotein complex is targeted to the inner membrane through a specific interaction with an integrally-associated membrane protein, termed the coupling protein (Schroder *et al.*, 2002; Szpirer *et al.*, 2000). This protein likely represents the inner membrane gate for transfer of substrate from the cytosol to the uncharacterized transenvelope structure, as suggested by the structure of the plasmid R388 coupling protein TrwB (Gomis-Ruth *et al.*, 2001). Using a TrwB truncation lacking the first 70 amino acids, a hexameric structure with a 20 Å central channel was revealed, including a nucleotide-binding domain in the central region and an all-alpha helical segment at the C-terminus. This quaternary structure resembles that of DNA ring helicases (such as FtsK), supporting the notion that DNA traverses through the central channel (Gomis-Ruth *et al.*, 2001). TrwB and other coupling proteins TraG (IncP plasmids), TraD (F plasmid), and HP0524 (*H. pylori*) have been shown to have non-specific DNA-binding activity (Schroder *et al.*, 2002). This may enable either docking of the DNA/relaxosome complex, and/or transfer of DNA through the central channel.

The individual proteins that comprise the different conjugative systems have conserved motifs, and in particular, Walker-type nucleotide triphosphate (NTP) binding domains [motif A, GxxGxGKS/T; motif B, hhhhDE; h=hydrophobic (Schneider and Hunke, 1998)] are the most common. The Walker A domain (or P loop) is involved in making contacts with the β- and γ-phosphates of ATP (Figure 1-1), whereas the Walker B domain binds divalent cations associated with the nucleotide (Schneider and Hunke, 1998). Each member of the coupling protein family have these domains, but none have been shown to hydrolyze ATP *in vitro* (Schroder *et al.*, 2002). ATP hydrolysis is

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Figure 1-1. The structure of the Walker A and B domains associated with ATP. The three-dimensional structure of TrwB (structure accession: 1GL7_A) was visualized using the NCBI structure viewing and comparison program Cn3D after primary sequence alignment with R27-encoded TraG (red = identical, blue = similar, grey = unaligned). The Walker A domain (GATGTGKS) is numbered 60-67 and is present as the loop associating with ATP analogue (green) prior to the helical segment (purple). The yellow residues represent the Walker B domain (WLFIDE).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

predicted to be an *in vivo* activity of the coupling proteins, resulting in structural re-arrangements within this inner membrane gate and possibly energizing substrate translocation through the inner membrane to the Mpf complex (Gomis-Ruth *et al.*, 2002a; Gomis-Ruth *et al.*, 2002b).

All known plasmid conjugation systems encode a coupling protein, as well as the CagA (*H. pylori*) and T-DNA/Vir (*A. tumefaciens*) T4S effector translocation systems. The coupling protein may be involved in transporting effectors across the inner membrane, as demonstrated by an interaction between the *A. tumefaciens* VirE2 effector and VirD4 coupling protein (Atmakuri *et al.*, 2003). Other effector translocation T4S systems may circumvent the requirement of a coupling protein for inner membrane transport by using the general secretory system, as in *B. pertussis* Ptl secretion.

The function of the coupling protein may not be limited to the inner membrane targeting and transfer of substrates. In addition to making contacts with cytoplasmic relaxase and effector molecules, an interaction with the Mpf complex or T4S apparatus would deliver the transferring substrate to these envelope-spanning structures, thereby facilitating substrate export to recipient cells. As of yet, there is no evidence supporting a direct interaction between a coupling protein and Mpf/T4S protein.

1.5 Incompatibility group H plasmids

Plasmids of the *Enterobacteriaceae* are categorized into groups based upon their ability (or inability) to co-exist within the same host bacterial cell. These incompatibility groups are each comprised of plasmids which share common features necessary for replication and/or partitioning, and thus one plasmid species will ultimately be favored when multiple related DNA species are present in a single cell. Plasmids of the H incompatibility group (IncH) were first characterized after an outbreak of chloramphenicol -resistant typhoid fever in Mexico in 1972 (Anderson and Smith, 1972). H plasmids are

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

divided into two subgroups: HI and HII. Most of the IncHI plasmids are large genetic elements (>150 kbp), thermosensitive for transfer, frequently associated with human pathogens, encode serologically related pili, and carry antibiotic and heavy metal resistance determinants (Taylor, 1989). The IncHI group was further subdivided into IncHI1, IncHI2 and IncHI3 subgroups following restriction digestion, DNA-DNA hybridization studies, and incompatibility studies which each indicated that members of the HI group shared little DNA similarity (Roussel and Chabbert, 1978; Smith *et al.*, 1973; Whiteley and Taylor, 1983). In contrast to IncHI1 plasmids, IncHI2 plasmids, including R478, are compatible with the F plasmid and this depends on the lack of the RepFIB replicon which is found in both F and IncHI1 plasmids (Smith *et al.*, 1973; Taylor *et al.*, 1985).

IncH plasmids have historically been the sole source of multi-drug resistance (MDR) in *Salmonella enterica* serovar Typhi. The sequences of the IncHI1 resistance plasmids R27 (180 kbp) and pHCM1 (218 kbp) have already been determined (Parkhill *et al.*, 2001; Sherburne *et al.*, 2000). These two plasmids were isolated from different strains of *S. enterica* sv. Typhi over three decades apart and in different regions of the world (R27: England, 1961; pHCM1: Vietnam, 1993), yet over 168 kbp of each sequence are >99% identical (Parkhill *et al.*, 2001). The difference in size is attributed to two clusters of drug resistance determinants present in pHCM1, but not in R27, which may have been acquired and maintained due to selective pressure imposed by exposure to antibiotics. In a recent study, all examined MDR strains of *S. enterica* sv. Typhi contained a plasmid resembling pHCM1, and some demonstrated acquisition of additional resistance determinants (Wain *et al.*, 2003). Several other recent articles have documented the occurrence of Asian and Indian MDR *S. enterica* sv. Typhi isolates containing IncHI1 plasmids (Mirza *et al.*, 2000; Shanahan *et al.*, 1998; Shanahan *et al.*, 2000). This incompatibility group of plasmids has also been isolated from individuals

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

infected with *S. enterica* sv. Typhi living in Africa (Kariuki *et al.*, 2000) and travelers returning from South Asia to North America (Harnett *et al.*, 1998).

One of the characteristic features of IncHI plasmids is a thermosensitive mechanism of conjugation. Since transfer occurs only at lower temperatures (22°C to 30°C) and not at 37°C, it is believed that transfer is optimized to occur in soil and water environments (Maher and Taylor, 1993). IncHI plasmids are one of the only known plasmid groups harboured by human pathogens that are temperature sensitive for transfer.

The IncHI1 plasmid R27 transfer determinants arose from multiple origins, as there is a mosaic of IncF-like and IncP-like elements (Lawley *et al.*, 2002a; Sherburne *et al.*, 2000). The Mpf system of this resistance plasmid is a member of a growing subset of F-like type IV secretion systems, including those from R391, the *Vibrio cholera* SXT element, and the IncHI2 plasmid R478 (Beaber *et al.*, 2002; Boltner *et al.*, 2002; Lawley *et al.*, 2003a). Analogous to the IncP plasmids, the R27 genes encoding Dtr and Mpf functions are located in separate regions of the plasmid, termed Tra1 and Tra2 respectively. The Tra1 region encodes relaxosome components, an origin of transfer, and some Mpf proteins (Lawley *et al.*, 2002a); and the Tra2 region encoding the majority of the R27 Mpf proteins (Lawley *et al.*, 2003a; Rooker *et al.*, 1999). In total, 16 Mpf proteins, 3 relaxosome-associated proteins, and a coupling protein (TraG) are encoded by R27.

1.5.1 TrhC, a putative ATPase

Encoded within the R27 Tra2 region is *trhC*, a gene which is essential for R27 transfer as determined by transposon mutagenesis (Taylor *et al.*, 1999). TrhC was previously reported to contain 630 amino acids (Taylor *et al.*, 1999), but is now known to have 893 amino acids and is designated as a member of the VirB4 family of Type IV secretion proteins, which includes among others, VirB4 from the Ti plasmid, TrbE from

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

plasmid RP4, and TraC from the F plasmid. Of these, TrhC has the closest sequence homology to TraC: 22.5% identity over 888 contiguous amino acids. All of these proteins contain the A and B-type nucleotide-binding motifs, and a considerable degree of sequence identity at the C-terminus (Rabel *et al.*, 2003).

It has been speculated that the NTP-binding motifs of the VirB4 proteins act to power either the assembly of the transfer apparatus or translocation of substrate (Zechner *et al.*, 2000). Similar to the *in vivo* results of purified coupling proteins, preparations of the VirB4-related proteins TrbE and TrwK encoded by plasmids RP4 and R388, respectively, also did not hydrolyze ATP or GTP (Rabel *et al.*, 2003). ATP hydrolysis by TrbE and VirB4 may occur under *in vivo* conditions since the NTP-binding motifs are essential for the function of both proteins (Berger and Christie, 1993; Fullner *et al.*, 1994; Rabel *et al.*, 2003). The precise role of the NTP-binding motifs of VirB4-related proteins in the conjugative donor apparatus is ill-defined.

The subcellular location of the VirB4-like proteins has been examined. Using PhoA fusions, VirB4 was found in the cytoplasmic membrane with two transmembrane domains (Dang and Christie, 1997). The C terminal tail of VirB4 remained in the bacterial cytoplasm and the membrane-association of VirB4 was independent of other T-DNA transfer proteins. Likewise, RP4 TrbE associates with the cytoplasmic membrane in the absence of all other Mpf proteins, however when the entire Mpf complex is present, TrbE will also associate with the outer membrane (Grahn *et al.*, 2000). The TraC protein required for F plasmid transfer is a cytoplasmic protein when expressed in the absence of other Tra proteins, but when accompanied by the remainder of the transfer proteins, TraC then localizes to the cytoplasmic membrane (Schandel *et al.*, 1992). The subcellular localization of TrhC has not been experimentally determined, but it is an essential R27 Mpf protein with sequence identity to the VirB4 family of proteins,

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

and when in the context of the Mpf complex, therefore likely to be present in the bacterial membrane.

1.6 Objectives

Through bioinformatics, biochemistry, and cell biology it was the aim of this thesis to characterize the H plasmid conjugative transfer protein apparatus. From the sequence of the H plasmid transfer determinants it was apparent that the Mpf genes have a different origin than the Dtr and coupling determinants. The function of the conjugative apparatus, however, is dependent on the interaction between members of these different groups, as demonstrated by specific protein-protein interactions and the visualization of a transfer protein subcomplex. For the members of these different lineages to function together, it can be postulated that they were adapted to each other through a single protein – the coupling protein TraG, which also serves as the inner membrane pore. A model for the assembly of a transenvelope structure responsible for secretion of both DNA and proteins is also presented. This work will help define the mechanism and structures associated with DNA transfer.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Chapter 2

The complete nucleotide sequence of R478: defining the backbone components of incompatibility group H conjugative plasmids

Portions of this chapter have been submitted for publication as:

Gilmour, M.W., Thomson, N.R., Sanders, M., Parkhill, J. and Taylor, D.E. The complete nucleotide sequence of R478: defining the backbone components of incompatibility group H conjugative plasmids through comparative genomics. Submitted for publication.

* note: N.R. Thomson, M. Sanders, and J. Parkhill determined the complete nucleotide sequence of R478 and performed the initial sequence annotation at the Sanger Centre, Cambridge, UK.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

2. The complete nucleotide sequence of R478: defining the backbone components of incompatibility group H conjugative plasmids

2.1 Introduction

The large resistance plasmid R478 was isolated in the USA in 1969 from a clinical isolate of *Serratia marcescens* (Medeiros and O'Brien, 1969); and was originally classified as an S plasmid in reference to this genus (Hedges *et al.*, 1975). In 1977, Taylor and Grant re-classified R478 into group HI2, and at this time, resistance to chloramphenicol, kanamycin, tetracycline, and mercuric ions was known to be mediated by this genetic element (Taylor and Grant, 1977a). IncHI2 plasmids were also known to inhibit the development of bacteriophages λ, T1, T5, and T7 (Taylor and Grant, 1977b). The thermosensitive profile of R478 conjugal transfer was determined to be optimal between 22-30°C, and inhibited at 37°C (Rodriguez-Lemoine *et al.*, 1975). Some R478-encoded determinants for replication, conjugative transfer and tellurite, arsenic and silver resistance have previously been described (Gupta *et al.*, 2001; Page *et al.*, 1999, 2001; Ryan and Colleran, 2002; Taylor, 1999; Whelan *et al.*, 1994; Whelan *et al.*, 1995).

In this study R478 was sequenced and annotated to identify the function and evolutionary history of IncHI plasmid backbone and accessory genes. Comparisons with the sequence data of R478, R27 and pHCM1, which have common backbone features but were isolated at different geographical and chronological points, allowed us to investigate the evolution of plasmid genomes.

2.2 Experimental Procedures

2.2.1 Sequencing and assembly of R478

The complete sequence of R478 was obtained from 6,644 paired end sequences (giving 11-fold coverage) derived from two pUC18 shotgun libraries (with insert sizes of 1.4 to 2 kb and 2 to 4 kb) using dye terminator chemistry on ABI3700 automated sequencers. All identified repeats were bridged by read-pairs or end-sequenced PCR

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

products. The sequence was assembled, finished and annotated as described previously (Parkhill *et al.*, 2000), using the program Artemis (http://www.sanger .ac.uk/Software/Artemis) to collate data and facilitate annotation (Rutherford *et al.*, 2000).

2.2.2 Annotation of R478 sequence data

Initial coding sequence predictions were performed using Orpheus (Frishman *et al.*, 1998) and Glimmer2 software (Delcher *et al.*, 1999). The two predictions were amalgamated, and codon usage (Nakamura *et al.*, 2000), positional base preference methods and comparisons to the non-redundant protein databases using BLAST (Altschul *et al.*, 1990) and FASTA (Pearson and Lipman, 1988) software were used to curate the predictions. Protein motifs were identified using Pfam (Bateman *et al.*, 2002), the CDD (Marchler-Bauer *et al.*, 2003), and Prosite (Falquet *et al.*, 2002), transmembrane domains were identified with TMHMM (Krogh *et al.*, 2001), and signal sequences were identified with SignalP version 2.0 (Nielsen *et al.*, 1997). The full-length sequence and annotation of plasmid R478 has been deposited in the EMBL database under accession number: BX664015. Comparison of the genome sequences was facilitated by using the Artemis Comparison Tool (ACT) (K. Rutherford, unpublished) which enabled the visualization of TBLASTX comparisons (Altschul *et al.*, 1990) between the genomes. Orthologous proteins were identified as reciprocal best matches using FASTA with subsequent manual curation (Table A-1). Additional local and global alignments were generated with ALIGN (at the EMBL website) using the BLOSUM62 matrix. Multiple sequence alignments were generated using ClustalW in MegAlign (DNAStar) and GENEDOC for visual representation.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

2.3 Results and Discussion

2.3.1 Physical features

The complete DNA sequence of plasmid R478 was obtained by a whole-plasmid shotgun method (Parkhill *et al.*, 2000), and it was determined to be a circular element of 274,762 bp. From the R478 sequence data, 295 coding sequences (numbered SMR0001-0304) representing a coding percentage of 83.5% were predicted (Figure 2-1). Homologous coding sequences were identified and are reported in Table A-1. The base composition of R478 is 45.5% G+C, whereas the relative buoyant density of R478 previously indicated a G+C content of 50% (Whiteley and Taylor, 1983). The G+C profile is illustrated in Figure 2-1, and areas in which G+C content highly diverged from this average directly relate to possible insertions (discussed below).

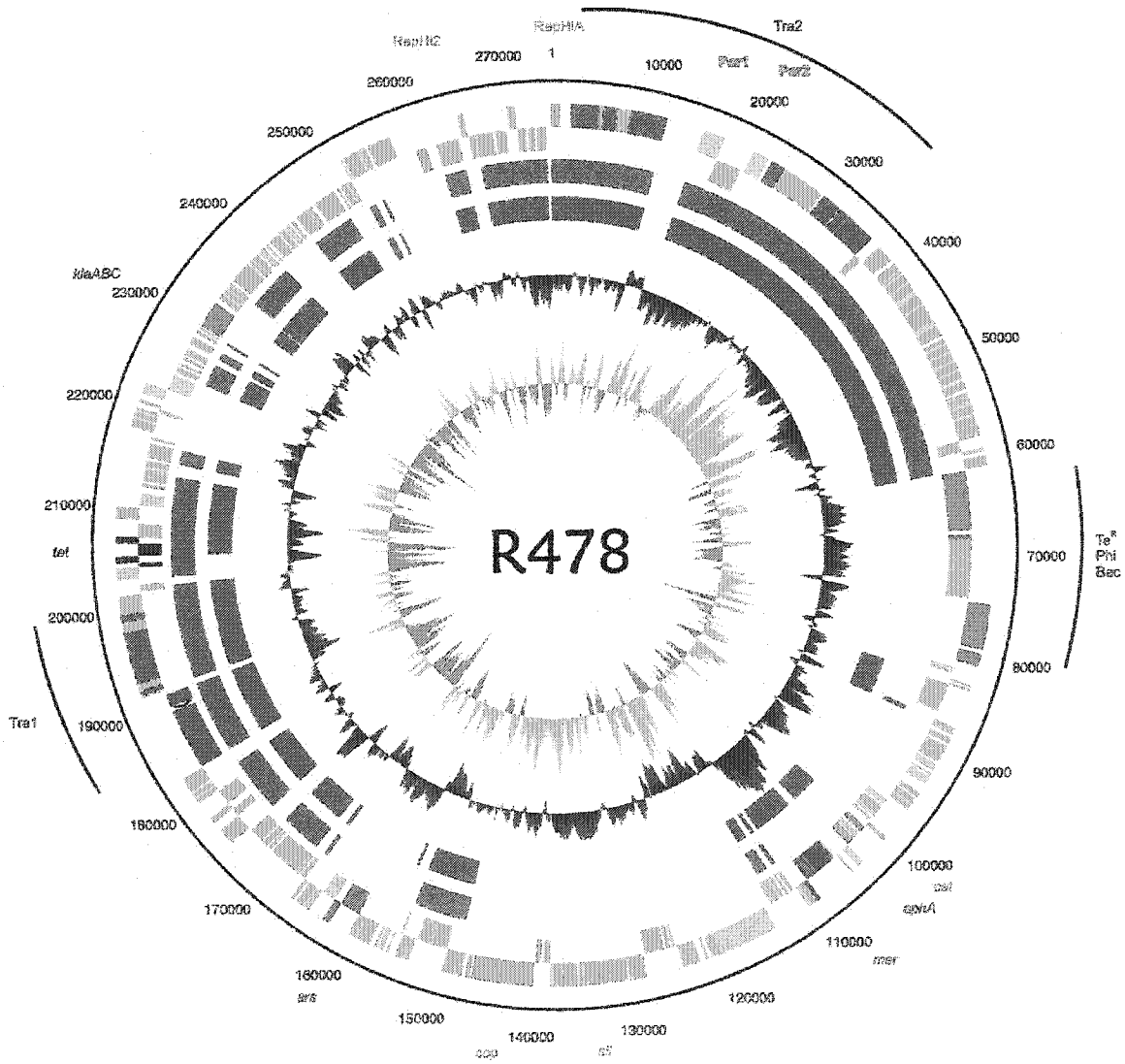
2.3.2 Replication and partitioning

The copy number of IncH plasmids has been estimated to be 1-2 copies per chromosome (Taylor, 1989). Plasmid replication is initiated at repeat iteron sequences after binding of a plasmid-encoded Rep protein and subsequent recruitment of host replication proteins (i.e. DnaA, helicases, primases) as the DNA helix is unwound revealing single-stranded template (del Solar *et al.*, 1998). Iteron-based replicons are indicative of theta-type replication (del Solar *et al.*, 1998), however it is unknown if uni- or bi-directional replication is initiated by any of the IncH-type replicons. R478 contains two functional iteron-based replication determinants termed RepHIA and RepHI2 (Page *et al.*, 2001). The RepHIA replicon includes the *repHIA* structural gene (SMR0001) surrounded by five upstream and nine downstream 19 bp repeated sequences (Figure 2-2). The equivalent replicons to RepHIA in R27 (RepHIA) and pHCM1 (RepA2) encode Rep initiations proteins that are 86% to the R478 RepHIA protein (Table A-1). Surrounding the R27-encoded RepHIA replication initiation protein are four upstream

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Figure 2-1. Circular representation of plasmid R478 coding sequences, base composition, and comparison to other H plasmids. The outer solid ring shows the nucleotide co-ordinates, where base 1 is in the first codon of the *repHIA* gene. The second and third rings represent predicted coding sequences, in the forward and reverse reading frames, respectively. Genes are color-coded as follows: red, conjugative transfer; orange, replication; yellow, partitioning; purple, tellurite resistance; light pink, insertion elements; dark pink, chloramphenicol resistance (*cat*); brown, kanamycin resistance (*aphA*); dark green, mercuric ion resistance; grey, silver resistance; light green, copper resistance; blue, arsenic resistance; black, tetracycline resistance; light blue, miscellaneous or unknown function. The black arrowhead at base position 189001 represents the putative *ter* site. The fourth ring (blue) represents sequences that are shared between R478 and R27, whereas the fifth ring (red) represents sequences shared between R478 and pHCM1. The sixth ring (black) illustrates the G+C content variation from the overall average of 45.5%. The seventh, innermost ring (yellow and purple), represents the GC skew (see text).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

and three downstream 19 bp repeats with a consensus sequence of AAAAGCATTTG-ATGAATG (Newnham and Taylor, 1994). The occurrence of this replicon in all sequenced IncHI plasmids partly explains the observed incompatibility.

In contrast, the RepHI2 replicon is not present in other sequenced IncHI plasmids; however the replication initiator protein is similar to numerous other iteron-based Rep proteins (Table 2-1). A previous annotation of the $RepHI2_{R478}$ replicon suggested that it comprised eight 18 bp iterons (four upstream and four downstream of the *repHI2* structural gene, SMR0292) and nine downstream 76 bp iterons, which each include an equivalent to the 18 bp sequence (Page *et al.*, 2001). We have redefined the iteron repeat sequences downstream of the *repHI2* gene as twelve 80-81 bp sequences that are positioned end-to-end (Figure 2-2). The previously defined four upstream repeats remain as previously described. An additional disagreement relates to the RepHI2 protein which was previously reported to be 319 aa (accession: AAB67987), whereas we predict a protein of 362 aa. The difference between these predictions occurs at the N-terminal coding region, and our prediction represents a coding sequence of similar length to the related *repA* coding sequences from pMT1 (accession: NP_395385) and pHCM2 (NP_569603). $RepA_{pMT1}$ is predicted to be 354 aa, and displays 51% identity to the RepHI2 protein over a range of 346 aa. $RepA_{pHCM2}$ is predicted to be 351 aa (Parkhill *et al.*, 2001), and is 51% identical to the RepHI2 protein over a range of 333 aa.

Replication termination systems in iteron-controlled plasmids are rarely observed, with the *ter*-Tus system of plasmids R6K, R1, and R100 being the only reported systems (Horiuchi and Hidaka, 1988). During uni-directional replication of R6K (or bi-directional replication of the *E. coli* chromosome), binding of the chromosomally-encoded Tus protein at the inverted *ter* repeats serves to halt the replication machinery (Lee *et al.*, 1989). In the sequence of R478 a single *ter* sequence was detected between *trhY* (SMR0206) and *trhF* (SMR0205) at base positions 188981-189001 that is nearly

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 2-1. Predicted function of R478 coding sequences. Related sequences were determined by BLAST analysis (see Experimental Procedures).

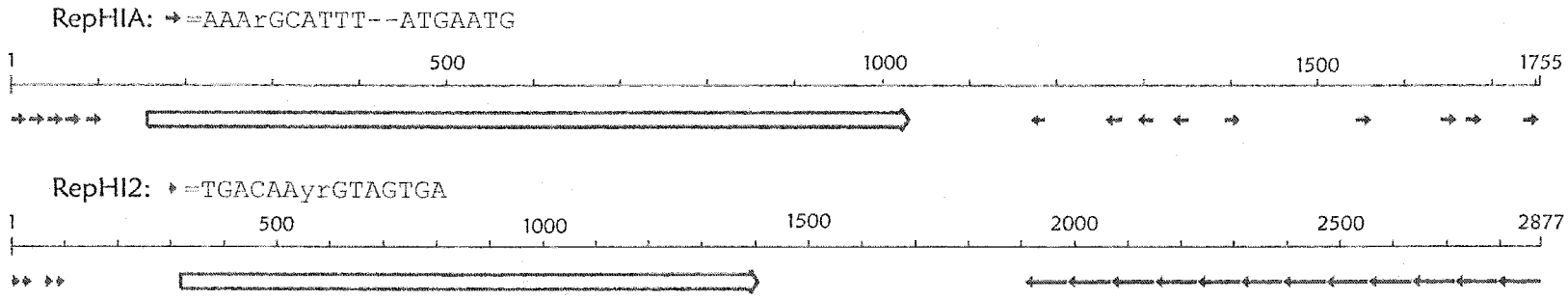
Function	R478 Sequence	Homologues encoded in: [a]
Replication	RepHIA	pMT1 (*Yersinia pestis*), pHCM2 (*S. typhi*), pSLT (*S. typhimurium*), F plasmid, pO157 (*Escherichia coli*), Rts1 (*Proteus vulgaris*), pS268a (*Serratia* sp.)
	RepHI2	
Transfer	TrhA-L-E-K-B-V-C-P-TraW-U-N; TrhG-H-F-Y-R-TraH-I-G-J	SXT (*Shigella flexneri*), Rts1, R391 (*Providencia rettgeri*), pRS241d (*Rhodobacter sphaeroides*), *Vibrio vulnificus*, pCAR1 (*Pseudomonas resinovorans*), pSLT, R100 (*S. flexneri*), pNL1 (*Sphingomonas aromaticivorans*), Salmonella genomic island 1 (SGI1)
Partitioning	ParA-B	pMT1, P7, P1, pSLT, Rts1, pCP301 (*S. flexneri*)
	ParM-R	pB171 (*E. coli*), pCP301, R1 (*E. coli*), pCTX-M3 (*Citrobacter freundii*)
Tellurite, phage, and colicin resistance	SMR0062-63-TerY3-Y2-X-Y1-W-69-70-71-72-73-74-Z-A-B-C-D-E-F	*Y. pestis, E. coli* O157:H7, *Streptomyces coelicolor, Deinococcus radiodurans, Haemophilus somnus, Pseudomonas syringae, Proteus mirabilus*
Organic mercury resistance	MerE-D-A-C-P-T-R	R100, pSB102 (from rhizosphere), pDU1358 (*S. marcescens*), R391, pKLH256 (*Enterobacter cloacae*), *Ralstonia metallidurans*
Silver resistance	SilE-S-R-C-SMR0145-B-A-148-P	pMG101 (*S. typhimurium*), *E. coli*
Copper resistance	CopE2-A-B-C-D-R-S-E1	pRJ1004 (*E. coli*), *R. metallidurans*, *P. syringae*
Arsenic resistance	ArsC-B-R-H	pSX267 and pI258 (*Staphylococcus* sp.), pYV (*Yersinia* sp.), *P. syringae*, R773 (*E. coli*), R46 (*S. typhimurium*)
Tetracycline	TetD-C-A-R	*Tn10* (R100)
Chloramphenicol resistance	Cat	mini-*Tn10*
Kanamycin resistance	AphA	Numerous bacterial genera
DNA damage repair	MucA-B	pKM101, R391, Rts1, pCTX-M3, R46, R394 (*S. typhimurium*), R446b (*Morganella morganii*), R471a (*S. marcescens*)
Sulphate permease	SfpA-B	Rts1, R391, *Tn1404*, *R. sphaeroides*, *R. metallidurans*, Yersinia plasmid fragment
Putative DNA binding proteins	SMR0042, primase; 55, Tus; 58, Cre recombinase; 60, Hha; 88, DNA helicase; 126, H-NS; 138, HigB regulator 139, OLD family endonuclease; 161, TlpA; 165, cytosine methylase; 166, nuclease; 191, RetA maturase; 272, methylase; 300, adenine-methylase	Numerous bacterial genera
Enzymes	SMR0037-39-40, nitric oxidase; 194, disulphide isomerase; 198, DsbC; 215, protease; 266, phosphoadenosine phosphosulphate reductase; 104/AdhC, alcohol dehydrogenase	Numerous bacterial genera

a. for references see text and Table A-1

Figure 2-2. Linear illustration of the iteron-controlled replicons of R478. Hollow arrows indicate the *rep* structural gene, while the solid arrows indicate the surrounding repeat sequences (iterons). Consensus sequences for the iterons are provided for each replicons (y = pyrimidine; r = purine; solid dash = unconserved sequence). The hatched solid line represents the nucleotide scale. The downstream RepHI2 iterons, represented as twelve tandem solid arrows, include the consensus sequence present in the four upstream iterons.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.



identical to the *ter* consensus, as well as a coding sequence, SMR0055, having ~20% identity to characterized Tus proteins (Figures 2-1 and 2-3). To my knowledge, this is the first observation of a plasmid encoding both *ter* and Tus. In the sequence of plasmids R27 and pHCM1 *ter*-like sequences were detected that were more divergent from the *ter* consensus than the R478-encoded *ter* sequence, but identical to each other (Figure 2-3). If the R478-encoded termination system is functional, a replication fork moving clockwise (in relation to the orientation shown in Figure 2-1) may be attenuated or arrested at the *ter* site, since the orientation of *ter* dictates which face of the *ter*-Tus complex is capable of blocking the approaching replication helicase (Lee *et al.*, 1989). The *repHIA* and *repHI2* structural genes are encoded on opposite strands of R478, and if uni-directional replication is initiated by both RepHIA and RepHI2 replicons, the presence of two oppositely-oriented replicons may ensure complete replication of R478 if one fork is blocked at the *ter* site. Plasmid R6K also has multiple replicons to ensure that replication of the entire plasmid is completed (Crosa, 1980).

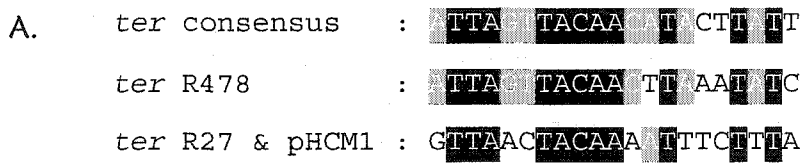
Additionally, an asymmetric distribution of bases was observed in the R478 sequence, visualized as the GC skew, or (G-C)/(G+C) (Figure 2-1). Mutational selection causes an enrichment for G and T in the leading strand (Lobry and Sueoka, 2002), and this was observed in the sequence of R478, wherein a large contiguous stretch from base positions 1600 to 166 200 of one strand has largely a positive skew, whereas a majority of the remainder of that strand has a negative skew (Figure 2-1). If the apparent inverted GC skew profile of R478 resulted from mutations favouring the replicative process, this suggests that either bi-directional replication occurs from a single replicon, as observed with circular bacterial chromosomes (Parkhill *et al.*, 2003), or that multiple uni-directional replicons are functional, each possibly creating localized GC skews.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

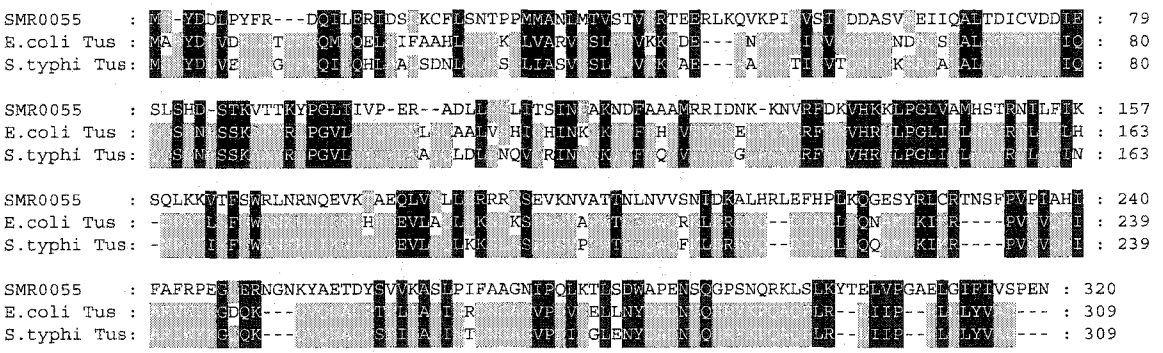
Figure 2-3. Multiple sequence alignment of *ter* and Tus sequences. The consensus *ter* sequence (23) is aligned with putative *ter* sequences from H plasmids R478, pHCM1, and R27 (A). Known Tus proteins (*E. coli* and *S. enterica* sv. Typhi) are aligned with the putative Tus protein encoded in R478 (B). Alignments were created using ClustalW and conserved residues shaded using Genedoc. Accession numbers: *E. coli* K-12 (NP_288045); *S. enterica* sv. Typhi (CAD01897).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

A.



B.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Active partitioning of R27 is a result of two distinct modules (Par1: *parSAB* and Par2: *parCMR*) inserted among the transfer genes, each capable of supporting partitioning, and Par1 being the major determinant (Lawley and Taylor, 2003). Each module is comprised of an array of direct repeat sequences (*parS* or *parC*) which serve as the centromere region for binding by ParB or ParR proteins, with which the respective ATPases (ParA or ParM) interact (Lawley and Taylor, 2003). R478 contains a 3080 bp region preceding the predicted *parA* and *parB* genes (SMR0014/15) that includes 11 repeats ranging from 233-257 bp, which are termed *parS* (Figure 2-4). Additionally, each of the 11 repeats contains a highly conserved sub-region (ATTTAACACCCTTAAT CGCCAAACTGG) that aligns with the 26 direct repeats (34 bp with consensus: cccCctTAaTcGcCAg--ccATGG-gg-a--c-g) present in the R27 *parS* region. Analysis of R27 *parS* outside of the 34 bp repeats did not reveal any additional conservation with the remainder of the larger R478 *parS* repeats.

A region similar to the second R27 partitioning module, Par2, was also detected in the R478 sequence (SMR0020/21; Figure 2-4). The ParM ATPase of this module is predicted to be an actin-type ATPase, rather than a Walker-type ATPase such as ParA encoded in the Par1 module. The nucleating proteins (ParR) of actin-type partitioning modules are less conserved, however the genetic organization of the Par2 structural genes of both R27 and R478 is conserved. Repeat sequences which could act as a centromere region in the Par2 region of R478 were not detected. The dual partitioning modules encoded on R478 resemble numerous plasmid-encoded sequences (Table 2-1). The presence of multiple replicons and partitioning modules may be a result of multiple evolutionary events leading to the duplication of these core plasmid systems, with selection and maintenance of each module ensuring the continuance of the plasmid in a broad range of hosts.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Figure 2-4. Linear illustration of sequences encoding conjugative transfer functions. Individual coding sequences are represented as coloured arrows, where solid colored arrows represent sequences homologous to the same-coloured essential R27 transfer gene; hollow coloured arrows represent sequences homologous to the same-coloured non-essential R27 gene; hollow black arrows represent unique sequences, and black solid arrows represent repeat sequences found in the partitioning and origin of transfer (*oriT*) regions. On the left, the source of the sequence is listed and the prefix for the nomenclature of the genes is in parentheses (example: the R478 gene 'A' represents '*trhA*', whereas R478 genes not having a *trh* prefix are labeled in their entirety). Solid lines above the arrows indicate contiguous sequences, and the R478 sequence is labeled with a nucleotide scale. Genes are transcribed in the direction indicated by the respective arrow. Citations: pCAR1, (Maeda *et al.*, 2003); R27, (Lawley *et al.*, 2002a; Lawley *et al.*, 2003a); SXT, (Beaber *et al.*, 2002); R391, (Boltner *et al.*, 2002); Rts1, (Murata *et al.*, 2002); pNL1, (Romine *et al.*, 1999)

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Within the R478 sequence additional sequences which may be involved in plasmid segregation were detected. SMR0243a displays 69% amino acid identity with Hok, a 52 amino acid peptide encoded on plasmid R1 which acts in post-segregational killing through its toxic effects to membranes (Gerdes *et al.*, 1997). Synthesis of Hok is prevented by antisense *sok* RNA, and only daughter cells receiving the plasmid (and thereby the ability to produce *sok*) will survive, ensuring the maintenance of the Hok/*sok*-carrying plasmid. This module in R478 is surrounded on either side by insertion sequences (*insLKJ*), and was likely to have been acquired as a cassette. SMR0278 (92 aa) is also related to a killer peptide: the 92 aa HigB of plasmid Rts1, shown to have post-segregational killing activity (Tian *et al.*, 1996). However, an equivalent to the suppressor protein HigA could not be detected, and it is likely that because of the modest level of conservation between the R478 and Rts1 HigB proteins (30% identity), that R478 HigB has an alternative function. The implicit genetic burden on the host cell by the large size of R478 may have resulted in the presence of multiple partitioning and killing functions, thereby decreasing the chance for elimination of R478 after segregation.

2.3.3 Conjugative transfer

One of the defining phenotypes of IncHI plasmids is the ability to transfer between cells via bacterial conjugation. Conjugation is a paradigm for horizontal gene transfer, and facilitates the spread of various genetic determinants throughout bacterial populations. This process is presumed to be initiated by the H-pilus, an extracellular filament comprised solely of pilin subunits (Lawley *et al.*, 2003b). The H-pilus is responsible for contact with recipient cells, then retracting to allow the transfer pore to form between the two bacterial cell walls. The transmembrane apparatus (the Mpf complex) that is required for pilus production, is also considered to be involved in pore

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

formation and the subsequent translocation of DNA. Plasmids are transferred to recipient cells as single-stranded DNA, a product generated by the cytoplasmic relaxosome complex, which includes DNA-binding proteins and a relaxase responsible for DNA cleavage at the origin of transfer (*oriT*). The Mpf complex and relaxosome are linked via an integral inner membrane coupling protein, completing the conjugative apparatus.

The transfer genes of R27 and pHCM1 are present as two separate regions, termed Tra1 and Tra2. A majority of the Mpf proteins are encoded in Tra2, whereas the *oriT*, some Mpf genes, and all relaxosomal components are encoded in Tra1 (Lawley *et al.*, 2002a; Lawley *et al.*, 2003a). The same transfer gene organization occurs in R478 and the individual gene products are 45-97% identical to the respective R27 transfer proteins (Figure 2-4). Essential R27 transfer proteins were previously determined in large-scale mutagenesis studies (Lawley *et al.*, 2002a; Lawley *et al.*, 2003a). In comparison with proteins encoded in the R27 transfer system, but which are not essential for R27 transfer, the respective R478 transfer proteins have an average percent identity of 64% (n=12; range: 45 – 85%), whereas essential transfer proteins are more conserved (average: 80%; n=21; range: 63 – 97%). The high level of both sequence and organizational conservation of the R27 and R478 transfer sequences implies that they share a common ancestor; however, mutations have accumulated in both essential and non-essential genes, with the greater occurrence in the later. The majority of the R478 transfer genes have been renamed with the prefix *trh*, as previously used for both R27 and pHCM1 transfer determinants, in preference to the prefix *htd* initiated by Page and colleagues (Page *et al.*, 1999).

The Mpf genes of all three sequenced H plasmids resemble those encoded on the genetic elements SXT, R391, Rts1, pNL1, and pCAR1 (Figure 2-4 and Table A-1). Each of these transfer systems resemble the prototypical transfer determinants of the F

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

plasmid in regards to both sequence identity and genetic organization. During the search of the genome databases for transfer sequences related to R478, two novel sequences having the same organization as the H plasmid transfer system were found. The genome data available for *Rhodobacter sphaeroides* 2.4.1 (accession: NZ_AAAE01000159) and *Vibrio vulnificus* CMCP6 (NC_004460) each contained putative T4S systems which encode an equivalent to each of the essential H plasmid Mpf transfer genes, but lack the inserted elements such as the partitioning modules and many of the non-essential transfer genes of H plasmids (Figure 2-4). Also, both the *R. sphaeroides* and *V. vulnificus* Mpf-like sequences are encoded in one contiguous segment, rather than in separate Tra regions, as observed in H plasmids. These novel sequences may represent a prototype of the H plasmid Mpf transfer systems since they each lack insertions (i.e. par modules) and non-essential Mpf genes. *R. sphaeroides* 2.4.1 contains a 100 kbp conjugative plasmid identified as pRS241d, and the R478-like transfer sequences were localized to this element (S. Kaplan and R.C. Mackenzie, personal communication). The *V. vulnificus* transfer genes were localized to chromosome II, as reported in the available submission. It is possible that the *Vibrio* genes are not involved in conjugative transfer, but rather T4S of an unknown substrate.

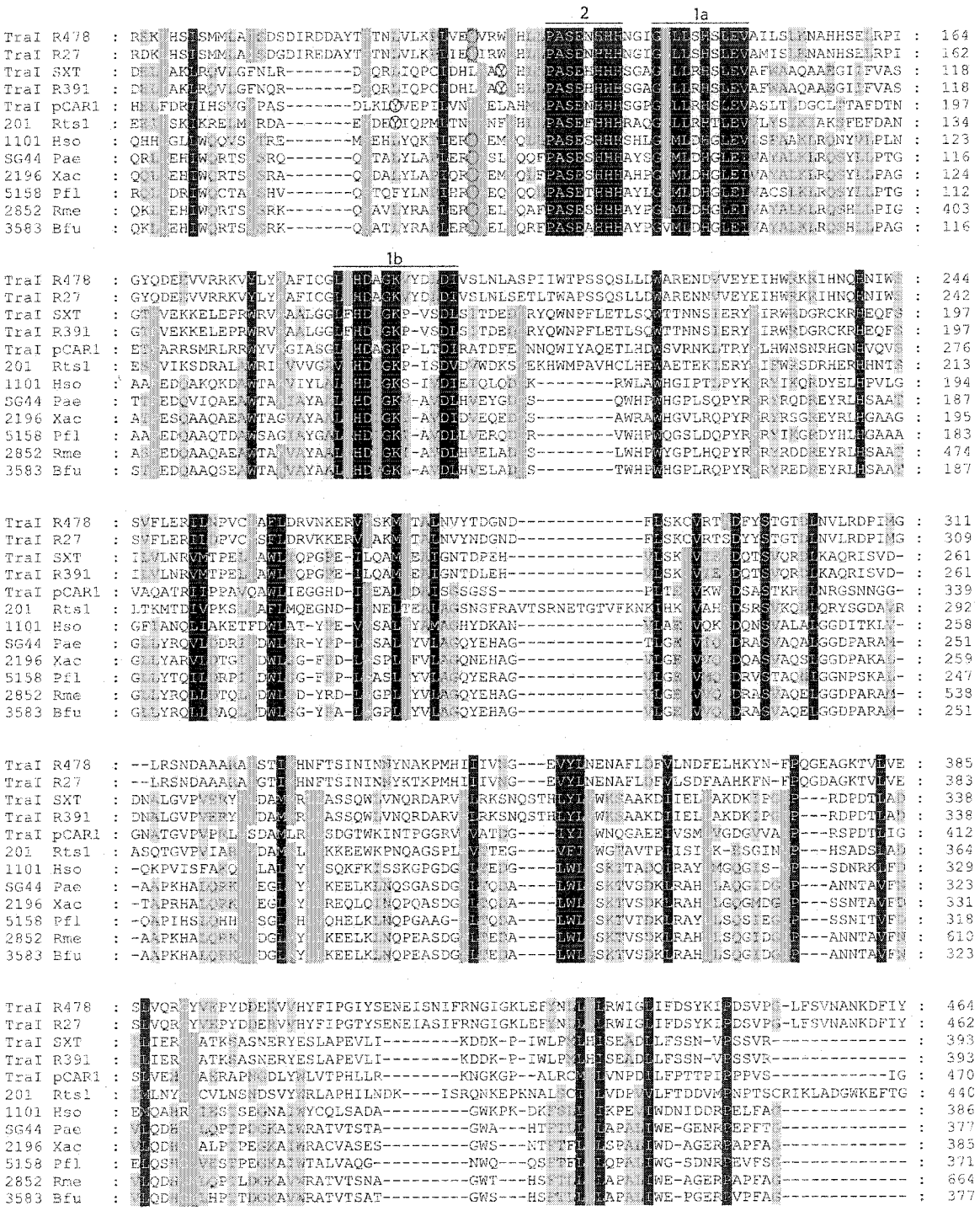
The relaxosomal (TraI and TraJ) and coupling proteins (TraG) of the H plasmids, SXT, R391, Rts1, and pCAR1 are not strict members of the F-type family, and constitute a novel phylogenetic group (Figures 2-5, 2-6, and 2-7). For each of TraI, TraG, TraJ, and the respective related proteins there is conservation of both primary amino acid sequence (including highly conserved subdomains), and predicted secondary elements, such as transmembrane domains (Figures 2-5, 2-6, and 2-7). The genomic context of *traI*, *traG*, and *traJ* is also conserved in multiple genetic elements, including the above sequences, plasmid pHG1, and the genomic sequences of *Haemophilus somnus*, *Xanthomonas axonopodis pv. citri*, *Pseudomonas fluoroscens*,

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Figure 2-5. Multiple sequence alignment of a conserved region from IncH TraI-related sequences (A). The alignment was generated using ClustalW in MegAlign (DNAStar) and conserved residues were shaded using GENEDOC. Previously characterized relaxase motifs containing conserved His and Ser residues are indicated with a '2' (see text), whereas novel conserved regions of the IncH-type TraI proteins are indicated with a '1'. Possible nucleophilic tyrosine residues are circled. Accession numbers: R478 (CAE51736); R27 (NP_058333); SXT (AAL59675); R391 (AAM08003); pCAR1 (NP_758664); Rts1 (NP_640161); *Haemophilus somnus* (Hso, ZP_00123099); *Pseudomonas aeruginosa* (Pae, AAN62266); *Xanthomonas axonopodis* pv. *citri* (Xac, NP_642513); *Pseudomonas fluoroscens* (Pfl , ZP_00087872); *Ralstonia metallidurans* (Rme, ZP_00023889); *Burkholderia fungorum* (Bfu, ZP_00030759). Rooted phylogenetic tree generated by MegAlign using the multiple alignment of the complete sequences for each of the above TraI homologues (B).

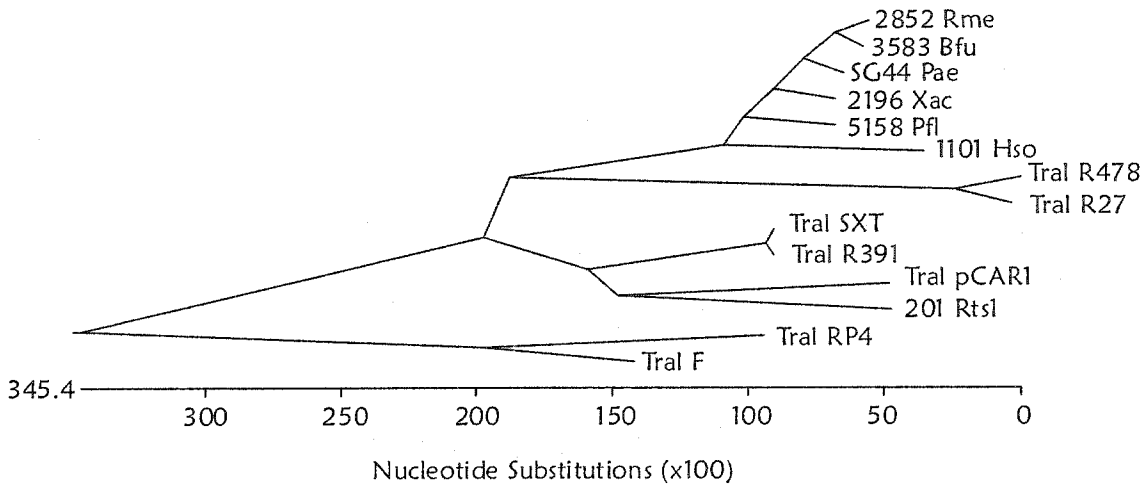
Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

A.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

B.

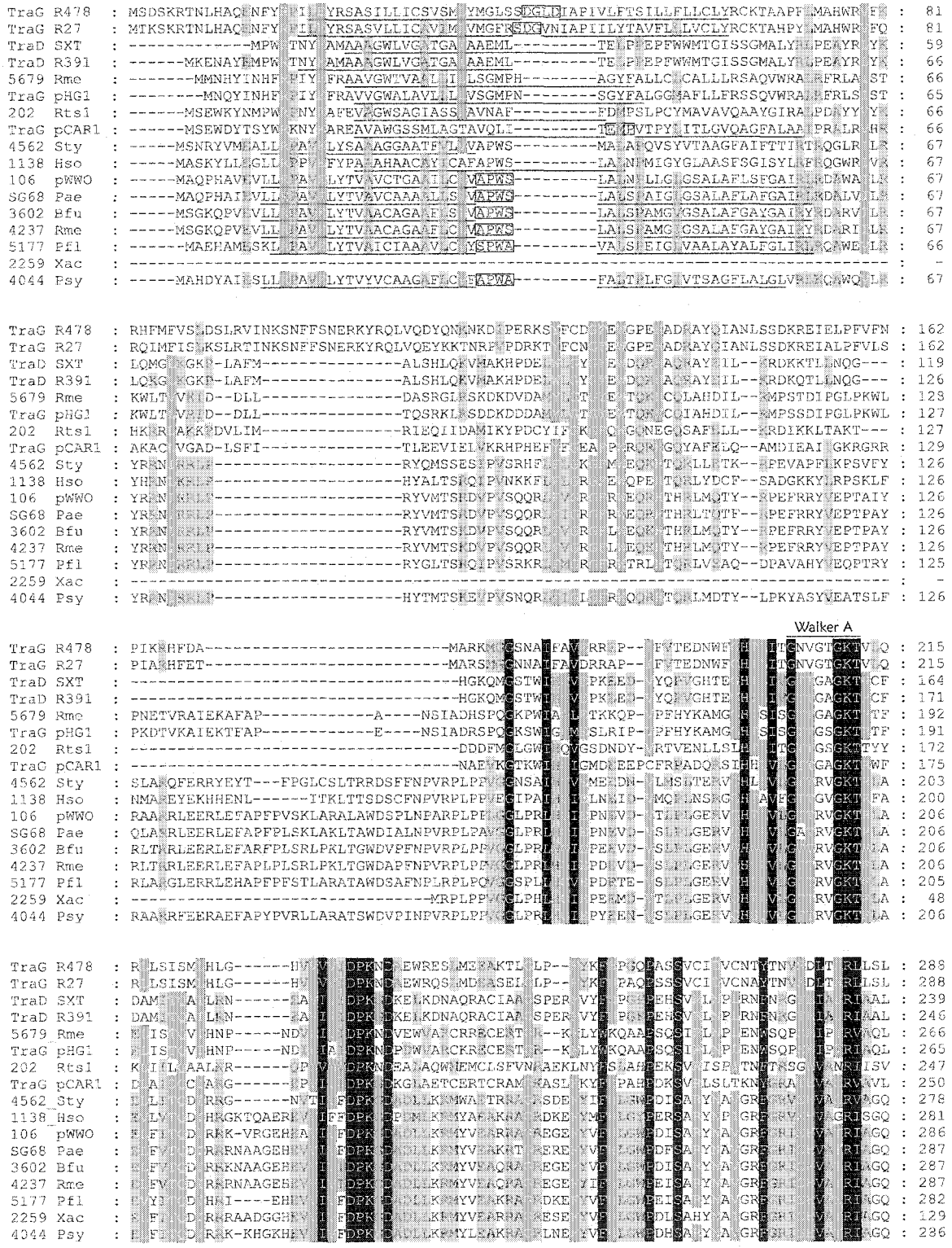


Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

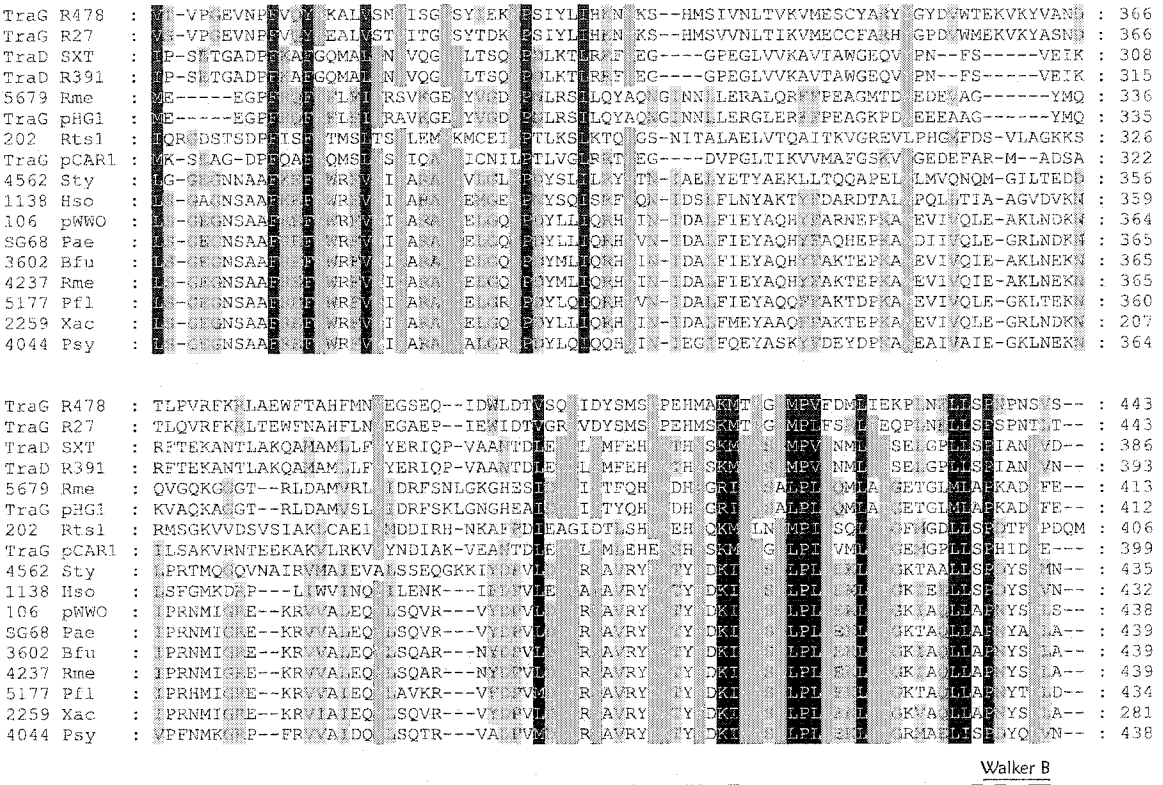
Figure 2-6. Multiple sequence alignment of TraG-related sequences (A). The alignment was generated using ClustalW in MegAlign (DNAStar) and conserved residues were shaded using GENEDOC. Underlined residues indicate transmembrane helices predicted using TMHMM, boxed regions are predicted periplasmic domains, and Walker NTP-binding motifs are indicated. Accession numbers: R478 (CAE51737); R27 (NP_058332); SXT (AAL59680); R391 (AAM08004); *Ralsotonia metallidurans* (Rme 5679, ZP_00026653); pHG1 (AAP86115); Rts1 (NP_640162); pCAR1 (NP_758665); *Salmonella enterica* sv. Typhimurium (Sty, NP_458648); *Haemophilus somnus* (Hso, ZP_00123135); pWWO (NP_542873); *Pseudomonas aeruginosa* (Pae, AAN62290); *Burkholderia fungorum* (Bfu, ZP_00030778); *R. metallidurans* (Rme 4237, ZP_00025244); *Pseudomonas fluorescens* (Pfl, ZP-00087890); *Xanthomonas axonopodis* pv. citri (Xac, NP_642576); *Pseudomonas syringae* (Psy, ZP_00127721). Rooted phylogenetic tree generated by MegAlign using the multiple alignment of the complete sequences for each of the above TraG homologues (B).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

A.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

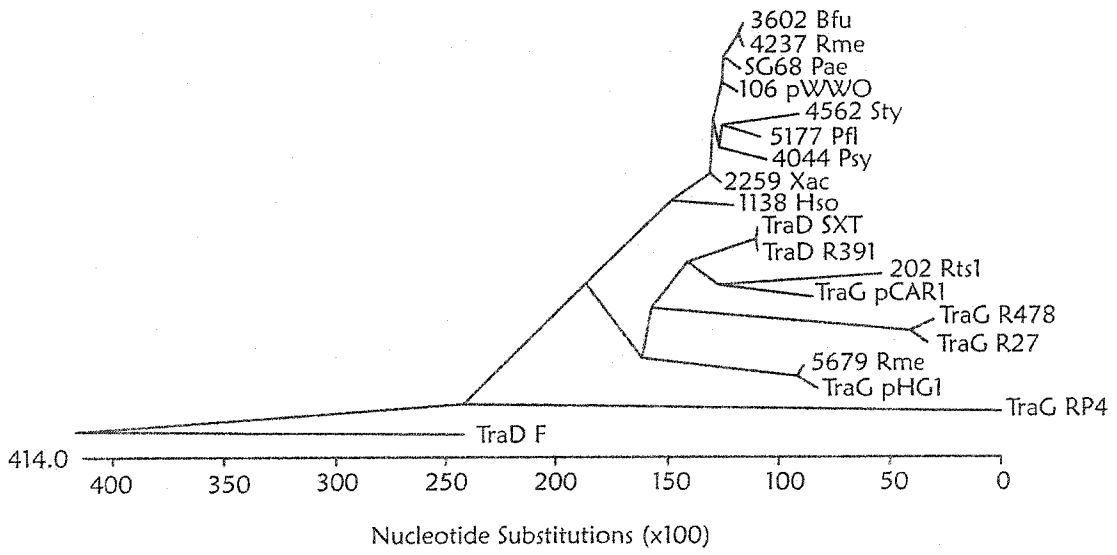


Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

```
TraG R478  : TGSETS-LPHNN GSISE KQTTLEESIPKDLLGQV MFHIVARLQD RKVVGQIPIAVAEKQMKPNTTLSEMLFKKAGK : 674
TraG R27   : TGSETT-LPHNN GSISE KQTTLEESIPKELLGQV FHIVARLQD RKVVGQIPIAVSEKAMKPNTTLEMFLKPAGK : 674
TraD SXT   : MSSN--DSPAL GSHGE LMEEEGD FPSQLLGQL NLYIAKLS RIKGIPILTSSTQAV--------------- : 599
TraD R391  : MSSN--DSPAL GSHGE LMEEEGD FPSQLLGQL NLYIAKLS RIKGIPILTSSTQAA--------------- : 606
5679 Rme   : TTG-TEAHVGE GSVSRSAQLEKVP IPHDLVSQL NLYFLRLS AYQGIPIIQD------------------- : 622
TraG pHG1  : TTG-TEAHIGE GSVSRSAQLEKVP IPHDLVSQL NLYFLRLS AYQGIPIIKD------------------- : 621
202  Rts1  : SSG---TDPEQI GSIGE LMEEESE FPAPILGWL NLFFANQAKV IKGVPILTKPKDENEIKRIKAKAAKFQAV : 638
TraG pCAR1 : NAG---EGGVVQGGVGE LMQEEGERFPTALLGEL DLYLAVFA DRKGLPILTGPTAN--------------- : 609
4562 Sty   : VDINPDQGSDI SVQDI SSTRVPI SIAEVNL KG AFILIN RIKIMPLPASVDDEMMFVNIQSIVTTMRKH : 674
1138 Hso   : FDSINPDDDKA SKTGE ISQKEVP LDVANVTNL KG AFILMN STYKVMPLPAIDKEDVIPASMQEILDKMHKH : 674
106  pWWO  : ADSIDPHGNTA SVTQD ISSNSVP IEFAHVVAETMG CFALTE NMFVMPLPAPDPDEAMPKDIQELAGYMRQH : 678
SG68 Pae   : ADSIDVRGPTD SVAQD ISTASVP IEFAHVVGL KG AFALIQ QWKIMPLPAPDPDEALPEDIQQLAGYMRQH : 679
3602 Bfu   : ADSIDIRGATD SVTQD ISMSSVP IEFSHIVGL KG CFALLQ QWKVMPLPAPDPDEVMPQDIQQLAGYMRQS : 679
4237 Rme   : ADSIDIRGATD SVTQD ISMASVP IEISHVVGL KG CFALLQ QWKVMPLPAPDPDEVMPTDIQQLAGYMRQS : 679
5177 Pfl   : ADTIDPEAHTD SSSQD VSSTSVP IEFAHIVSL KG MFSFQA QWKVMPLPKPSNDDAMPKDIQELTQRMRAT : 675
2259 Xac   : ADSIDPTGNTA SVTQD IASTSVL IEFAHVVAL KG AFALNE NWKVMPLPAPDPDEAMPRDIQELAAGMRLH : 521
4044 Psy   : ANDS--ATGNTA SSSHDQVQQVSVP IEFAHVVGL KG CFALNE NWKIMPLPGRDPDEVMPADIQALAESMRQT : 676

TraG R478  : VT-------------LRQNLDIKNINKFIRKLH---------------------------------------- : 694
TraG R27   : VT-------------LRQNVGLSYINKYIRKLH---------------------------------------- : 694
TraD SXT   : ------------------------------------------------------------------------- : -
TraD R391  : ------------------------------------------------------------------------- : -
5679 Rme   : ------------------------------------------------------------------------- : -
TraG pHG1  : ------------------------------------------------------------------------- : -
202  Rts1  : V--------------EKKVRGEKVRSKLANA------------------------------------------ : 655
TraG pCAR1 : ------------------------------------------------------------------------- : -
4562 Sty   : V----RTGDTWWSGSVLPG-------DNLVDAL------------------------------------------ : 696
1138 Hso   : V----HVATNWWEPAFKDYTPSKDIADSFENMVTEEKMAKADLNWGADIDLPEDDEAESISQADTDEQDMEEE : 743
106  pWWO  : V----VEAGDWWENQGIPGLQDEAIPDDILDDFR----QMVAA----------EEAEA-------------- : 718
SG68 Pae   : V----AEAGDWWENQSLPGLQDQPIPDDILDGFR----HLAAP------DTAVDGEARP------------- : 724
3602 Bfu   : V----SEATQWWEFTSSPALQDGAIPDDILDDAAAVEPGLVAT------DDSAGNEASP------------- : 728
4237 Rme   : V----SEATQWWEFTSSAALPHAAIPDDILDDAAPAEPDAVATGA----DDSTG-EAAP------------- : 729
5177 Pfl   : VN---TQAGQWWS--ASGGGPTTNFDLDQVG---------------------------------------- : 701
2259 Xac   : V----VEAGAWWDGQIAPGLQDATIPDDILAGFAP--PEDAGG---------NPGAAAP------------- : 565
4044 Psy   : VQVGKVDDTGWWSAPGYQTLN-ESIPADILDDFKK---IAVAP----------EDAA--------------- : 719
```

B.

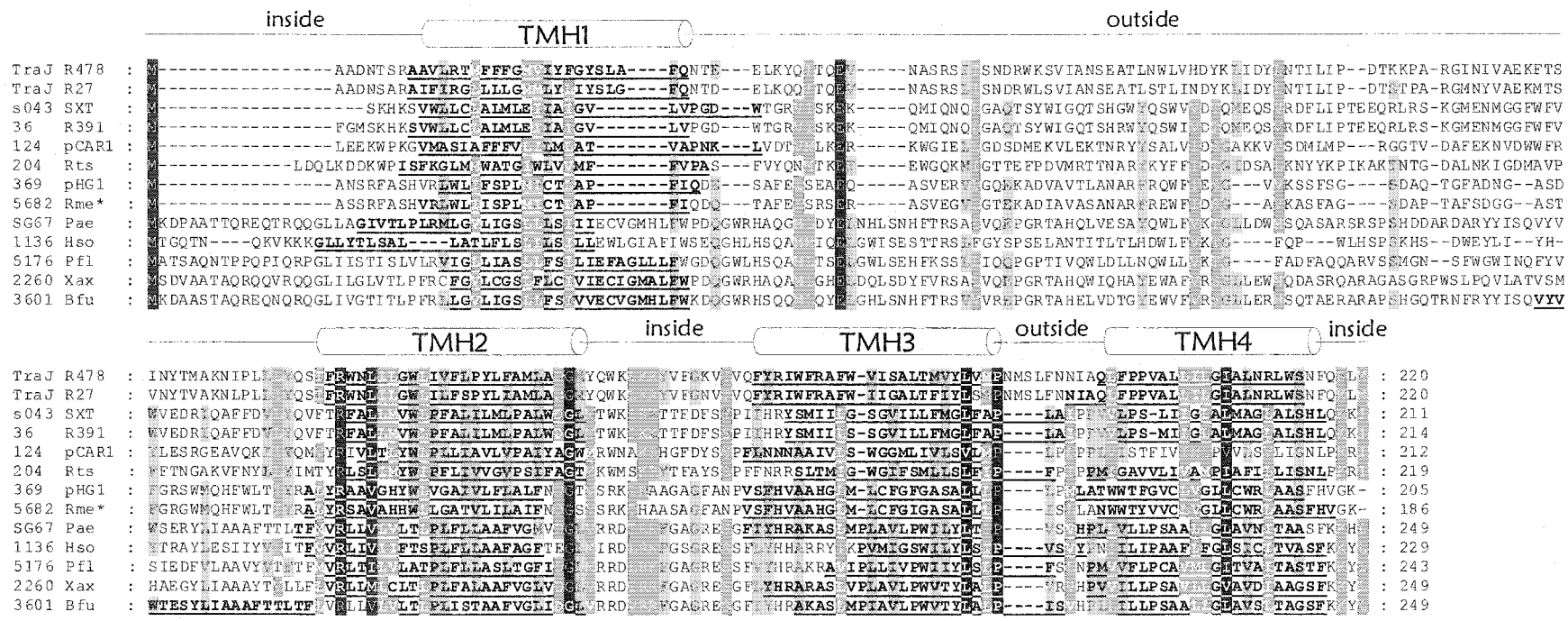


Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Figure 2-7. Multiple sequence alignment and second structure analysis of TraJ and homologues. Bold/underlined residues indicate transmembrane helices (TMH) predicted by TMHMM. The schematic of TraJ represents predicted TMH (cylinders), cytoplasmic domains (inside), and periplasmic domains (outside). Accession numbers: R478 (CAE51739); R27 (NP_058330); SXT (AAL59721); R391 (AAM08039); pCAR1 (NP_758667); Rts1 (NP_640164); pHG1 (AAP86118); *Haemophilus somnus* (ZP_00123133); *Xanthomonas axonopodis pv. citri* (NP_642577); *Pseudomonas fluoroscens* (ZP_00087889); *Burkholderia fungorum* (ZP_00030777); *Pseudomonas aeruginosa* (AAN62289); *Ralstonia metallidurans* (ZP_00026655, *note: the *Ralstonia metallidurans* coding sequence 5682 has been re-annotated from the entry at NCBI to include 19 more codons and an alternate start codon).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.



Burkholderia fungorum, *Pseudomonas aeruginosa*, and *Ralstonia metallidurans* (Table 2-2). This indicates that *traI-G-J* sequences have been co-inherited by numerous genetic elements from a common ancestor. When the list of genomes having H-type Dtr/coupling determinants is compared to the genomes having F-like Mpf genes, there is only overlap with the highly related H plasmid, SXT, R391, Rts1, and pCAR1 elements. Otherwise, plasmids pNL1, pRS241d, and the F-like plasmids have F/H-type Mpf genes, but not the H-type *traI-G-J* genes. These observations indicate that the IncH plasmid Mpf and Dtr/coupling determinants have evolved from different lineages.

It was previously reported that relaxosomal and coupling determinants of the IncH plasmids were derived from an IncP lineage (Lawley *et al.*, 2002a). This classification was made partly because the R27-encoded relaxase protein TraI lacks the C-terminal helicase domain encoded on the F plasmid TraI and has the conserved histidine triad motif present in IncP TraI sequences (Figure 2-5) which is involved in Zn^{2+} co-ordination in the active site (Guasch *et al.*, 2003). There is however, little overall sequence identity between $TraI_H$ and $TraI_P$ (14% identity in a global alignment comprised of 46% gaps). In contrast, each member of the H-type TraI family have multiple conserved domains outside of the previously described relaxase motifs, and in a phylogenetic tree, cluster together outside the IncP (RP4) and IncF TraI sequences (Figure 2-5). The coupling proteins of the H-type Dtr/coupling group are also more similar to each other, ranging from 23-31% identity (Figure 2-6, Table A-1), than to TraG of IncP plasmid RP4 (15% identity) or TraD of F (16% identity). There is also conservation of predicted transmembrane and periplasmic domains (Figure 2-6). TraJ is unique to the H-type Dtr/coupling group, and all of the TraJ homologues are of similar length, all have four predicted transmembrane helices, and these helices are in relatively the same site as they are predicted to occur in R478-encoded TraJ (Figure 2-7). In addition, between predicted transmembrane helices number 3 and 4 (Figure 2-7), there is a conserved

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 2-2. Co-inheritance of *traI*, *traG*, and *traJ* sequences. Homologues to R478-encoded relaxosomal proteins and the coupling protein were detected in the protein database at NCBI. For those genomes described in *genus species* format, these are preliminary annotations deposited to NCBI, and may represent chromosomal or plasmid sequences. Accession numbers: R478 (CAE51736, CAE51737, CAE51739); R27 (NP_058333, NP_058332, NP_058330); SXT (AAL59675, AAL59680, AAL59721); R391 (AAM08003, AAM08004, AAM08039); pCAR1 (NP_758664, NP_758665, NP_758667); Rts1 (NP_640161, NP_640162, NP_640164); pHG1 (AAP86114, AAP86115, AAP86118); *Haemophilus somnus* (ZP_00123099, ZP_00123135, ZP_00123133); *Xanthomonas axonopodis pv. citri* (NP_642513, NP_642576, NP_642577); *Pseudomonas fluoroscens* (ZP_00087872, ZP_00087890, ZP_00087889); *Burkholderia fungorum* (ZP_00030759, ZP_00030778, ZP_00030777); *Pseudomonas aeruginosa* (AAN62266, AAN62290, AAN62289); *Ralstonia metallidurans* (ZP_00026652, ZP_00026653, ZP_00026655).

Genome	*traI*[a]	bp between *traI-traG*[b]	*traG*[a]	bp between *traG-traJ*[b]	*traJ*[a]
R478	***traI***	0	***traG***	1095	***traJ***
R27	***traI***	10	***traG***	1089	***traJ***
SXT	***traI***	69	***traD***	557	**043**
R391	***traI***	48	***traD***	547	**36**
pCAR1	***traI***	-4	***traG***	372	**124**
Rts1	**201**	-8	**202**	382	**204**
pHG1	**365**	-4	***traG***	785	**369**
Haemophilus somnus	**1107**	27166	**1138**	334	**1136**
Xanthom. axonopodis pv. citri	**2196**	68646	**2259**	-2	**2260**
Pseudomonas fluoroscens	**5158**	18591	**5177**	90	**5176**
Burkholderia fungorum	**3583**	16902	**3602**	-4	**3601**
Pseudomonas aeruginosa	**SG44**	18376	**SG68**	-4	**SG67**
Ralstonia metallidurans	**5678**	-2	**5679**	805	**5682**[c]

a. gene designation of homologous sequences are as reported in the accession file listed in the legend

b. overlapping coding sequences are marked with a negative value

c. the annotation of *Ralstonia metallidurans* coding sequence 5682 has been re-annotated from the entry at NCBI to include 19 more codons and an alternate start codon

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

tandem of positively charged residues (Lys-Arg), which may play a role in orienting the protein in the membrane (i.e. positive-inside rule), or substrate interactions. Helix-turn-helix and leucine zipper motifs were previously predicted for TraJ (Lawley *et al.*, 2002a), but these domains were not detected in the current analysis. The apparent membrane-localization of TraJ is a feature not normally associated with Dtr components, and this essential transfer protein may play a role in the membrane-targeting of the relaxosome during conjugation (J.E. Gunton and D.E. Taylor, unpublished data).

2.3.4 Resistance determinants

Resistance mechanisms for both antibiotics and toxic metals in environmental bacteria pre-date human surveillance, but the combination of selective pressures and horizontal gene transfer have provided the motive and means, respectively, to cause rapid dissemination of these traits throughout the bacterial population, including medically relevant pathogens (Davies, 1994). This is demonstrated by the appearance of multiple resistance phenotypes in *Salmonella* and *Serratia* species due to the carriage of H plasmids. The resistance determinants on R478 are encoded primarily in one contiguous region from co-ordinates 61238-162175. The exceptions are a Tn*10* insertion encoding tetracycline resistance and a secondary tellurite-resistance operon that lie outside of this region (Figure 2-1).

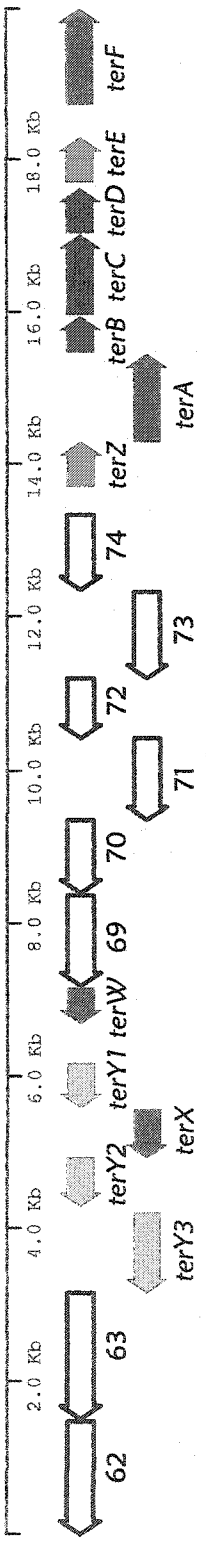
2.3.4.1 Tellurite, bacteriophage, and colicin resistance

Tellurium (Te) compounds have a long standing history as therapeutic agents and as selective agents for the isolation of pathogens (Taylor, 1999). Five distinct groups of plasmid and chromosomally-encoded tellurite resistance determinants have been identified (Taylor, 1999). The IncHI2-type determinants are characterized by a conserved set of sequences (*terZABCDEF*) found on numerous IncHI2 plasmids isolated from different bacterial hosts (Hou and Taylor, 1994) (Figure 2-8). These determinants

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Figure 2-8. Genetic organization of the operons encoding tellurite resistance, Phi and Pac. Individual coding sequences are represented as arrows, and coloured arrows have been previously characterized, except for TerY2 and TerY3, which are similar to TerY1. Genes are transcribed in the direction indicated by the respective arrow.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

provide a high level of resistance to tellurite oxyanions (MIC ~1024 μg/ml) by lessening the oxidation of thiol groups (Turner *et al.*, 1999), however the mechanism of this protection is unknown.

Besides providing resistance to tellurite, the IncHI2-type determinants inhibit host cell infection by bacteriophage Lambda, T1, T5, T7 and ϕ80 (Phi phenotype), and resistance to pore-forming colicins A, B and K (PacB phenotype) (Taylor and Grant, 1976; Whelan *et al.*, 1995). TerC, TerD and TerZ appear to be the principle determinants, since transposon mutants of the respective genes were reduced in or lacked each of the above phenotypes (Whelan *et al.*, 1995). It was also noted that cloning of *terZABCDEF* in a multicopy vector resulted in a lethal phenotype and filamentous morphology of the host cell; and the inclusion of an additional DNA segment containing sequences for *ter*WYX 'protected' cells from these effects (Whelan *et al.*, 1997). TerW contains a predicted helix-turn-helix and is a member of COG2378 (transcriptional regulator), corresponding to the previous prediction that TerW mediates protection from TerZABCDEF by modulating gene expression (Whelan *et al.*, 1997).

Tellurite resistance determinants related to the IncHI2 determinants are present in a large number of chromosomal sequences, including many strains of *E. coli* O157:H7, *Yersinia pestis* and the genus *Proteus* (Taylor *et al.*, 2002; Toptchieva *et al.*, 2003) (Table 2-1). *E. coli* and *Y. pestis* sequences include both the *terW* and *terZ* divergently opposed operon sequences, and the six genes of unknown function present in the interceding region between *terW* and *terZ* (Figure 2-8), implying that that this entire region has transferred between numerous bacteria and genetic elements. Using the complete sequence of R478, there may be additional genes involved in mediating the phenotypes of tellurite resistance, Phi, and PacB. Two additional copies of *terY* were detected (and the *terY* genes have been renamed *terY1*, *terY2*, and *terY3*), and at the

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

end of this putative operon (Figure 2-8) there are two coding sequences (SMR0062/63) with strong sequence identity to known serine/threonine kinases and phosphatases. Amino acids 140-273 of SMR0062 align with the protein kinase domain pfam00069 (e value: 2e-8), and 298-545 of SMR0063 align with the phosphatases domain cd00143 (e value: 6e-4). This may represent a signaling module responsive to a yet unknown signal.

A surprise on the R478 sequence was the detection of IncP-type tellurite determinants *klaABC* (otherwise called *kilA*, *telA* and *telB*). The R478 coding sequences SMR0256-258 are 23-57% identical to full-length coding sequences encoded on the pseudomonad plasmids pCAR1 (Maeda *et al.*, 2003) and RK2, and it was previously found that all three RK2-encoded determinants are essential for resistance to tellurite (Goncharoff *et al.*, 1991; Maeda *et al.*, 2003; Turner *et al.*, 1994). The two essential cysteine residues present in TelB (Turner *et al.*, 1994) are lacking in both the pCAR1 and R478 KlaC sequences, and *klaC* is the least conserved gene of this tripartite operon, indicating that if functional, this module may have a different mode of action. No obvious insertion elements surround this cassette of resistance genes, nor do any other genes that might point towards a Pseudomonad origin (Table A-1). Additionally, the *klaABC* genes are not present in either R27 or pHCM1.

2.3.4.2 Mercury resistance

Toxicity to inorganic mercuric ions results from Hg(II) binding with high affinity to protein thiol groups (Silver and Phung, 1996). The *merTPCAD* and *merR* genes encoded by R478 (SMR0118-223) are highly similar in both organization and sequence to respective sequences on R100 (Tn*21*), R391, pDU1358, pSB102, Tn*501*, pKLH201, pKLH256, and other genetic elements (Table 2-1). Mercury resistance is mediated by efflux, and the structural genes for Hg transport being *merTPC*, with the latter gene

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

being absent from some *mer* resistance operons. After transport from the cytosol, MerA, the mercuric reductase volatizes Hg^{2+} to the less reactive Hg^0 gas. Both *merD* and *merR* are regulatory genes (Summers, 1992) that are not part of the *mer* operon. The most striking change in G+C content on R478 occurs within the *mer* region, the left border of which has been disrupted by transposable elements which are now themselves present as pseudogenes (Figure 2-1).

2.3.4.3 Silver resistance

Silver compounds are used in hospitals, industry, and agriculture as antimicrobial agents due to their bactericidal effects (Gupta and Silver, 1998). The initial identification of a silver resistance operon (*sil*) encoded on an IncHI2 plasmid was after isolation of pMG101 from a nosocomial outbreak strain of *S. enterica* sv. Typhimurium (Gupta *et al.*, 2001). In this study, *sil* sequences were also detected on R478, as well as resistance to silver cations by cells harbouring R478. Silver resistance was found to be inducible by exposure to silver cations, possibly due to the presence of a two-component regulatory system comprised of an inner membrane-associated sensor histidine kinase (SilS) which phosphorylates and activates the respective cytoplasmic response regulator (SilR), which in turn activates transcription of the structural genes (*silABC*). The SilABC proteins form a tripartite, membrane-spanning chemiosmotic cation/proton antiporter (SilA: inner membrane antiporter; SilB: periplasmic-spanning protein; and SilC: outer membrane protein). Also present in the pMG101 sequence is a P-type ATPase (SilP) related to other cation transport ATPases, therefore multiple mechanisms for silver cation export seem to be present. The sequence data for R478 confirms that all of the *sil* sequences present on pMG101 are highly conserved on R478 (SMR0141-149 range from 92-100% amino acid identity compared to pMG101 Sil sequences). These two

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

IncHI2 plasmids are likely to be very closely related, as pMG101 similarly encodes resistance to mercury, tellurite, chloramphenicol and tetracycline (McHugh *et al.*, 1975).

Encoded upstream of the R478 *sil* sequences are the Tn*7* transposition genes *tnsABCD* (SMR0132-135). A complement of TnsABC+E are generally observed for Tn*7* insertions into conjugative plasmids, whereas the TnsABC+D array is involved in recognition of *attTn7* for insertion into bacterial chromosomes (Peters and Craig, 2001). The R478-encoded sequences are also atypical in that there is a low level of sequence conservation (ranging from 21-39% overall length-wise identity), and neither the tandem 22-bp repeats at the Tn*7* left and right border, or an *attTn7* site (Peters and Craig, 2001) were observed. However, TnsC has retained the characteristic Walker A NTP-binding domain, and TnsD has a helix-turn-helix appropriate for a DNA-binding protein (data not shown).

2.3.4.4 Copper resistance

Although transition metals such as copper and zinc are essential for function of many oxidizing enzymes, in excess they are toxic to cells due to non-covalent interaction with cysteine and histidine residues (Gatti *et al.*, 2000). The mechanism of copper tolerance was determined to be copper sequestration in the periplasm by CopA and CopC (Cha and Cooksey, 1991). A two-component regulatory system, CopS and CopR, controls expression of the structural genes (Munson *et al.*, 2000). The putative copper tolerance determinants encoded on R478 (CopABCDRSE, SMR0153-159) are 92-100% identical to the PcoABCDRSE proteins encoded on pRJ1004 (Brown *et al.*, 1995), a 125 kbp conjugative plasmid belonging to incompatibility groups I1 or K (Tetaz and Luke, 1983). Furthermore, the two-component regulatory system of the *cop* operon is similar to that regulating silver resistance operons (see above), the *E. coli* chromosomally-encoded YlcA and YbcZ (Munson *et al.*, 2000), and the CzcRS proteins which promote

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

expression of *czcABC* encoding a cation-proton antiporter in *Ralstonia metallidurans* (Nies, 1995). These latter functions are encoded on pMOL30, a 240 kbp self-transmissible plasmid that also encodes resistance to mercury, zinc, cadmium, and cobalt (Mergeay *et al.*, 1985). Plasmid-encoded copper resistance mechanisms are apparently highly conserved, and are controlled by a two-component system utilized by numerous resistance operons.

2.3.4.5 Arsenic resistance

Ryan and Colleran (2002) characterized the R478 resistance determinants for arsenate and arsenite by using a 5.5 kbp subclone containing six R478 coding sequences (*sfpAB* and *arsCBRH,* SMR0174-179), for which they could detect similar sequences on numerous other IncHI2 plasmids. Interestingly, arsenic resistance is mediated by ArsC which reduces arsenate, As(V), into the more toxic arsenite, As(III), which is then transported out of the cytosol by ArsB. R478, like the Staphylococcal plasmids pSX267 and pI258 (Silver and Phung, 1996), does not encode the P-type ATPase ArsA which interacts with ArsB for anion transport. Consequently, arsenite must be transported chemiosmotically. ArsH has an unknown function, but is required for arsenic resistance in Yersiniae carrying plasmid pYV (Neyt *et al.*, 1997). ArsR acts as a transcriptional activator for the *ars* operon, and possibly for the divergently transcribed *arsH* (Ryan and Colleran, 2002). Furthermore, ArsR is thought to co-regulate *sfp* and *ars* operons (both encoding anion transport systems), as binding sequences of this regulator are also found upstream of the *sfpA* gene (Ryan and Colleran, 2002). Immediately after the R478 *arsH* sequence is a transposase sequence (similar to the IS*1202* transposase), and this region also has an elevated G+C content. Each of these characteristics suggest that this resistance module was acquired as an

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

insertion, possibly from a sequence related to Tn*2502* on plasmid pYV (Neyt *et al.*, 1997).

2.3.4.6 Tetracycline resistance

The common resistance determinant that is present in all three sequenced IncHI plasmids is Tn*10*, encoding *tetRACD* (R478: SMR0220-223). These genes encode a tetracycline efflux system, in which TetA is the integral membrane transporter, TetR represses expression of the operon, and TetC and TetD have unknown function (Lawley *et al.*, 2000). The *tet* genes are bounded on each side by insertion sequence elements (IS*10*L and IS*10*R), which cumulatively represent a composite transposon. Tn*10* is also present on plasmid R100 (accession: NC_002134). Despite being common to all H plasmids, Tn*10* is present in a different location in each plasmid, implying that Tn*10* was not present in an ancestral H plasmid, and has been independently acquired by each of these plasmids. Additionally, a subsequent insertion in *tetC* of pHCM1 by an IS*1*-related element disrupted *tetC* and lead to the deletion of *tetD*. The IS*10*R sequence is still present in pHCM1, and the intervening sequences between IS*10*R and the *tetC* pseudogene include the numerous resistance determinants encoding β-lactamase, chloramphenicol acetyltransferase, streptomycin phosphotransferase and a mercury resistance operon, which account for the majority of the differences between pHCM1 and R27.

2.3.4.7 Chloramphenicol and kanamycin resistance

Similar to the insertion sequence-rich region of pHCM1, R478 encodes resistance determinants *cat* (chloramphenicol acetyltransferase, SMR0108) and *aphA* (an aminoglycoside phosphotransferase that provides resistance to kanamycin, SMR0110). In R478 these sequences are surrounded by five identical IS*26* elements (Figure 2-9). The IS*26* sequences appear to have disrupted the existing IntI1 class 1 integrase

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Figure 2-9. Genetic organization of R478 sequences encoding *cat*, *aphA*, and IS*26*. Individual coding sequences are represented as arrows, and pseudogenes are coloured brown. Genes are transcribed in the direction indicated by the respective arrow.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

(SMR0096) and recombinase/resolvase (SMR0096b) elements, revealing a succession of insertion events in this region as part of the expansion and gene loss that has occurred during the evolution of R478.

The determinants for metal and antibiotic resistance encoded on R478 may have been acquired from various other plasmids, as indicated by the presence of orthologous sequences on both related and unrelated plasmids (Table 2-1). It is also apparent that various transposons and insertion elements (Tn*7*, Tn*10*, IS*26*, and IS*1202*) have contributed to the acquisition of these determinants on R478. Consistent with this notion is the observation that each of the resistance determinants have markedly higher G+C content than the R478 average.

2.3.5 Other functions

Numerous proteins with putative DNA-binding motifs are encoded by R478 (Table 2-1). It is possible that these genes have been steadily acquired and maintained because of potential roles in regulation (i.e. H-NS, TlpA, and Hha), stability and modification (i.e. several putative nucleases and DNA methylases) of this large extrachromosomal genetic element, or that they may be involved in plasmid replication (i.e. presence of a putative helicase and a primase, see Table 2-1). There are also several genes predicted to encode enzymes which may be required for the function of other plasmid-encoded proteins (i.e. putative disulphide isomerases, Table 2-1).

Upstream of the Tra1 region are a conserved set of coding sequences (SMR0191-193) that are similar to a putative retroelement (coding *retA* and *mucAB*) encoded in the IncL/M plasmid R471a, isolated from *S. marcescens*. The *retA* gene encodes a putative reverse transcriptase (RetA) and *mucAB* encode homologues of the chromosomally-encoded UmuCD mutagenesis proteins (Kulaeva *et al.*, 1998). In plasmid R471a, *mucB* is followed by a gene encoding the transposase of an IS1396 element, which led

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Kulaeva and colleagues (Kulaeva *et al.*, 1998) to suggest that this putative retroelement has been mobilized from a different genome, as *mucAB* are commonly found on plasmids (Table 2-1). Such an insertion sequence is absent in R478, indicating that if these determinants were previously acquired by transposition from R471a, or a similar element through this proposed mechanism, subsequent genomic insertions or deletions in R478 have immobilized the retroelement. Notably, the *mucAB* genes are present in the same location of R27 and pHCM1, although the upstream *retA* gene is lacking. Again, it is possible that subsequent genomic rearrangements have caused the deletion of this feature.

2.3.6 Comparison of three H plasmid sequences

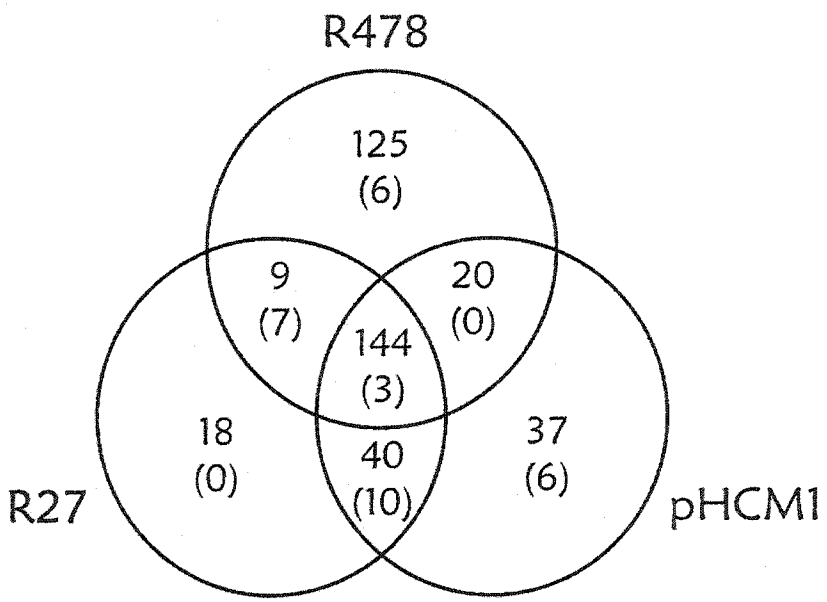
Comparative genomics of microbial chromosomal sequences has permitted the identification of niche-specific, virulence and strain-specific determinants (Deng *et al.*, 2003; Parkhill *et al.*, 2003; Perna *et al.*, 2001). The comparison of related plasmid sequences allows, on a smaller scale, an analysis of evolutionary events which have created transmissible genetic elements encoding multiple resistance phenotypes.

Global sequence comparisons demonstrate that the IncHI2 plasmid R478 is extremely similar to the previously sequenced IncHI1 plasmids R27 (Sherburne *et al.*, 2000) and pHCM1 (Parkhill *et al.*, 2001) (Figures 2-10 and 2-11). The original classification of three subgroups for IncHI plasmids was partly based upon the results of a DNA-DNA hybridization study that suggested IncHI1 and IncHI2 plasmids are unrelated at the DNA sequence level (Roussel and Chabbert, 1978). Another study attempting to determine the relatedness of plasmids in the same incompatibility group found that of the Inc F, I, N, B, and H groups, only members of the IncH group do not share sequence similarity (Grindley *et al.*, 1973). A more recent study concerning a different set of IncHI plasmids also found a low level of similarity between IncHI1 and

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Figure 2-10. Venn diagram detailing the unique and shared gene complements of plasmids R478, R27 and pHCM1. Numbers in parentheses indicate the number of insertion sequence elements that are included in the grouping.

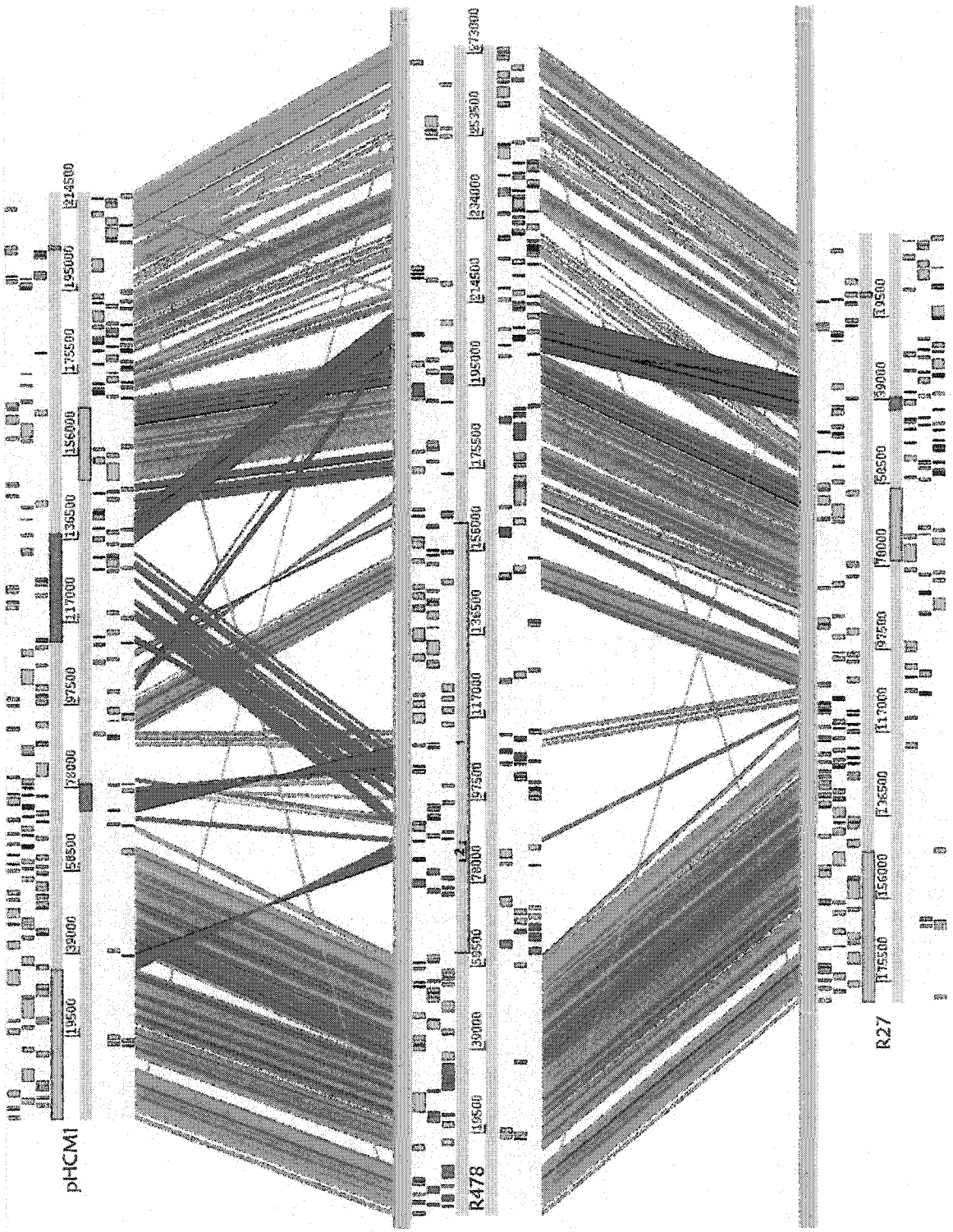
Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Figure 2-11. Linear genomic comparison of R478 to R27 and pHCM1. The six-frame translation of the R478 sequence was compared to the six-frame translation of R27 or pHCM1 using tBlastX and displayed using the Artemis Comparison Tool (ACT). Regions having similarity above a threshold of approximately 27% identity are connected via a red line. R478 coding sequences are color-coded as follows: red, conjugative transfer; orange, replication; yellow partitioning; purple, tellurite resistance; light pink, insertion elements; dark pink, chloramphenicol resistance; brown, kanamycin resistance; dark green, mercuric ion resistance; grey, silver resistance; light green, copper resistance; blue, arsenic resistance; black, tetracycline resistance; light blue, miscellaneous or unknown function. R27 and pHCM1 core genes (replication, transfer and partitioning) are marked with orange boxes, while resistance genes marked with purple boxes. R478 sub-regions labeled 1 or 2 are discussed in the text.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

IncHI2 groups, where R478 restriction digests were probed with radiolabelled IncHI1 plasmid DNA (Whiteley and Taylor, 1983). IncHI1 probes that bound to IncHI2 restriction fragments were attributed to common resistance and transfer determinants (Whiteley and Taylor, 1983). It is likely that DNA hybridization would only be positive with highly related sequences. With the complete sequence data now available, it is clear that there is in fact considerable 'relatedness' between members of different H incompatibility subgroups. This was likely not detected by the less precise methods used in earlier studies.

Approximately half of the R478-encoded genes are also present in both pHCM1 and R27 (Figure 2-10). In particular, there is conservation of both gene order and individual coding sequences in a large contiguous section comprising RepHIA (SMR0001), the *tus* gene, partitioning modules Par1 and Par2, the Tra2 conjugative transfer region, and the downstream genes (of mostly unknown function) up to and including SMR0061 (Figure 2-11). Another segment (SMR0185-218), comprising the Tra1 transfer region, *mucAB*, and surrounding genes of unknown function, is also highly conserved. It is likely that the common ancestor of R478, pHCM1 and R27 had a genetic organization similar to these conserved regions.

Divergence from this ancestral sequence may have occurred after transfer events, resulting in separate lineages. The principal difference between R478 and the other sequenced H plasmids is a large segment downstream of Tra2 and upstream of Tra1 in which there are few shared sequences. This region is termed the principal plasticity zone, and is labeled as section 1 in Figure 2-11. In this region are encoded the majority of the resistance determinants (*ter*, *cat*, *aphA*, *mer*, *sil*, *cop*, *ars*), and also Tn*7* and IS*26* sequences. In pHCM1, this same region has been subject to insertion of *bla*, *cat*, *str*, and *mer* resistance determinants (Parkhill *et al.*, 2001), and it also accounts for the primary difference between pHCM1 and R27. Additionally, the ancestor common to

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

R478, R27, and pHCM1 encoded no resistance determinants, and insertion of each determinant constituted an independent acquisition. Tn*10* is the only resistance determinant common to all three sequenced H plasmids, and this transposon is present at unique locations in each plasmid. Notably, R478 coding sequences SMR0088 (encoding a putative ATP-dependent DNA helicase) and SMR0089 (unknown function, but containing a predicted signal peptide cleavage site), which are 99% identical to pHCM1 coding sequences HCM1.102 and 103, also appear to be independent insertions into the principal plasticity zone of each respective plasmid, since they are present in a different genetic context and orientation (Figure 2-11, section 2).

In terms of incompatibility classifications, the absence of both RepHIB and RepFIB replicons in R478, and the presence of the unique RepHI2 replicon, verifies the validity of the HI2 subgrouping, rather than being subsumed into HI1, despite the observed high level of sequence conservation. The RepHI2 replicon (SMR0292) is in a region containing many repeated coding sequences of unknown function (SMR0236, 299, and 304 are similar to each other, as are 280, 290, 293; data not shown) that each have related sequences in R27 and pHCM1 (Table A-1). There are no obvious insertion sequences surrounding RepHI2, but the large number of duplication events in the surrounding region indicate a localized history of genetic rearrangements. Another distinctive feature of this region suggesting prior rearrangements is the absence of insertion sequences adjacent to each side of the putative Hok/*sok* systems encoded in R27 and pHCM1, whereas SMR0243a/*hok* is bordered by such features.

The presence of multiple functional replicons may permit the IncHI plasmids to extend both host range and limit incompatibility with other plasmids (Gabant *et al.*, 1993). The evolutionary events resulting in the presence of multiple replicons, however, are unclear. It is plausible that a singular ancestral IncHI plasmid encoded each of the RepHIA, RepHIB and RepHI2 replicons, with subsequent recombination events resulting

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

in the loss of individual replicons and the generation of the currently observed IncHI plasmids. Alternatively, individual IncHI plasmid species encoding a single replicon may have independently recombined with co-resident plasmids, thereby expanding the number of replicons. Each of the HI replicons are similar to numerous other plasmid replicons (Table 2), and these plasmids (or their ancestors) may have been the source of the present IncHI replicons. This latter possibility is also more likely because no H plasmid carrying all of these replicons has been observed, and the genetic organization around the RepHI2 replicon indicated prior re-arrangements.

Using the sequence data from R478, R27, and pHCM1, and experimental data concerning individual coding sequences, it is possible to define the backbone components for IncHI replication, partitioning and maintenance, and conjugative transfer. The minimal IncHI determinants, as defined by these three H plasmid sequences, include the RepHIA replicon, the essential *trh*/*tra*/*oriT* sequences encoded in Tra1 and Tra2, and the Par1 partitioning module. The RepHIA replicons from both R27 and R478 were found to independently support plasmid replication (Newnham and Taylor, 1994; Page *et al.*, 2001), and this was the only common replicon amongst the three sequenced IncHI plasmids. Essential R27 and R478 transfer genes have been determined (Lawley *et al.*, 2002a; Lawley *et al.*, 2003a; Page *et al.*, 1999), and a difference in the level of conservation between essential and non-essential transfer sequences was observed. The higher sequence identity between essential R478 and R27 transfer orthologues implies that these genes are subject to additional selective forces to conserve functionality, impeding the divergence from the ancestral sequence. Similarly, the Par1 regions of R478 and R27 are more conserved than the putative Par2 regions, and the Par1 region of R27 was found to be the principal partitioning determinant (Lawley and Taylor, 2003). The backbone includes no resistance determinants, but rather these are accessory functions which in part distinguish the different plasmid species.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

A successful combination of backbone determinants that ensures continuing plasmid survival (through vertical and horizontal transmission) would be expected to be highly conserved. The presence of the Par module in the Tra2 region may ensure the co-inheritance of partitioning and conjugative transfer genes. Similarly, the RepHIA determinant is present immediately upstream of the Tra2 region in all sequenced IncHI plasmids, which may also serve as evidence of co-inheritance. Notably, the putative replication initiator protein RepC of *R. sphaeroides* plasmid pRS241d (accession: ZP_00008306) is also encoded immediately upstream of the H/F-type transfer determinants, and the RepC protein of *A. tumefaciens* plasmid pTiC58 is encoded adjacent to the *trb* transfer determinants (Li and Farrand, 2000). Thomas (2000) has previously suggested that clustering and intricate control networks between the core plasmid determinants promote co-inheritance.

2.3.7 Conclusions

The evolution of R478, R27 and pHCM1 was seemingly predominated by a series of acquisition events, rather than deletion events, as evidenced by the conserved co-linear nature of shared sequences. These shared co-linear sequences include each of the designated backbone determinants. Once a productive backbone of core plasmid determinants has evolved, as demonstrated with the H plasmid backbone, it is likely to be highly conserved because each backbone determinant encodes a function essential for vertical or horizontal transmission. Insertions of accessory determinants into vertical transfer determinants may result in the loss of the individual recipient plasmid from the gene pool, whereas insertions into horizontal transfer determinants may restrict the plasmid to the present host. In contrast, insertion or deletion of accessory determinants in non-core regions may allow conversion to a genetic element conferring novel phenotypic traits. This evolutionary pattern may have resulted in the apparent clusters

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

of backbone and accessory determinants observed in the H plasmid sequences. R478 may have originated from an ancestral IncHI plasmid, encoding the backbone sequences described above, but independently acquired accessory determinants (including each of the resistance determinants). Likewise, the IncHI1 plasmids R27 and pHCM1 are also derivatives of this ancestral sequence and have also independently acquired accessory determinants. Incompatibility groups are therefore an abstract classification scheme defined by a conserved set of plasmid backbone features, whereas individual plasmid species are, in addition to the backbone, defined by accessory features. The predominance of resistance determinants in IncHI plasmids highlights the influence of the host on the processes of selection and maintenance of plasmid accessory features.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Chapter 3

Visualization and Characterization of Conjugative Transfer Protein Complex Assembly via a Mpf Protein-GFP Fusion

Portions of this chapter have previously been published as:

Gilmour, M.W., Lawley, T.D., Rooker, M.M., Newnham, P.J., and Taylor, D.E. (2001) Cellular location and temperature-dependent assembly of IncHI1 plasmid R27-encoded TrhC-associated conjugative transfer protein complexes. *Mol Microbiol* **42**: 705-715.

Gilmour, M.W and Taylor, D.E. (2004) A subassembly of R27-encoded transfer proteins is dependent on TrhC nucleotide triphosphate-binding motifs for function but not formation. *J Bacteriol* **186**: In Press.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

3. Visualization and characterization of conjugative transfer protein complex assembly via an Mpf protein-GFP fusion

3.1 Introduction

The 238 amino acid (27kDa) green fluorescent protein (GFP) of *Aequorea victoria* has proven useful as a reporter for protein localization in bacteria. Not only does GFP permit *in situ* visualization of fusion proteins in real time, but visualization occurs in a non-invasive fashion in live bacteria without substrate addition (Margolin, 2000). GFP mutants created by Cormack and colleagues (Cormack *et al.*, 1996) provide considerable improvements over wild-type GFP. GFPmut2 (Ser65Ala and Val68Leu) is a red-shifted variant with an excitation maxima at 488nm. At this wavelength GFPmut2 fluoresces 19-fold more intensely than wild-type GFP. This variant protein is also significantly more soluble than the wild-type and has a decreased maturation time, as wild-type GFP is not detectable until 1-2 hours after synthesis, while GFPmut2 fluoresces within 8 minutes (Cormack *et al.*, 1996). These properties permit more accurate time course and localization studies to be conducted than with wild-type GFP.

The proteins which facilitate conjugative transfer of plasmid DNA represent a dynamic module that is functional in the cytosol, cell envelope, and extracellular environment. Besides the pilus, no distinguishable superstructures have been visualized for the conjugative apparatus, although several macromolecular complexes have been proposed to exist. These include the cytoplasmic relaxosome responsible for DNA processing prior to and during transfer (Lanka and Wilkins, 1995), the membrane-associated mating pair formation complex (Mpf) responsible for construction of the pilus and DNA transfer (Grahn *et al.*, 2000), and a multimeric coupling protein which links the relaxosome to the Mpf complex (Gilmour *et al.*, 2003; Gomis-Ruth *et al.*, 2001; Schroder *et al.*, 2002). Despite the absence of a readily visible transmembrane apparatus during electron microscopy, such as that visualized in type III secretion systems (Blocker *et al.*,

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

2001), several studies have identified specific interactions between type IV secretion proteins or identified functional subgroups indicative of protein assemblies (Gilmour *et al.*, 2003; Grahn *et al.*, 2000; Harris *et al.*, 2001; Liu and Binns, 2003; Ward *et al.*, 2002).

This study focused on the VirB4-related protein TrhC, an inner membrane-associated Mpf protein required for the conjugative transfer of the large IncHI1 resistance plasmid R27 (Lawley *et al.*, 2003a; Taylor *et al.*, 1999). TrhC is encoded in both R478 and R27 (84% identity), and despite being a member of the VirB4 family, is more similar to F-type TraC proteins (Figure 3-1). A TrhC-GFP fusion construct was created which permitted visualization of the cellular location of TrhC-associated protein complexes. TrhC was present in membrane-associated protein complexes, visualized as distinct fluorescent foci at the cell periphery. In addition, the R27-encoded proteins that are required for the formation of TrhC-associated protein complexes were identified, and the role of the Walker NTP-binding motifs within TrhC for protein complex assembly, DNA transfer, and pilus formation were addressed. A key feature of this study is the insertion of *gfp* into R27 plasmid DNA to create a *trhC-gfp* fusion under control of natural expression determinants. This revealed the temperature-dependent expression profile of *trhC* and mechanisms for the thermosensitive transfer of R27.

3.2 Experimental Procedures

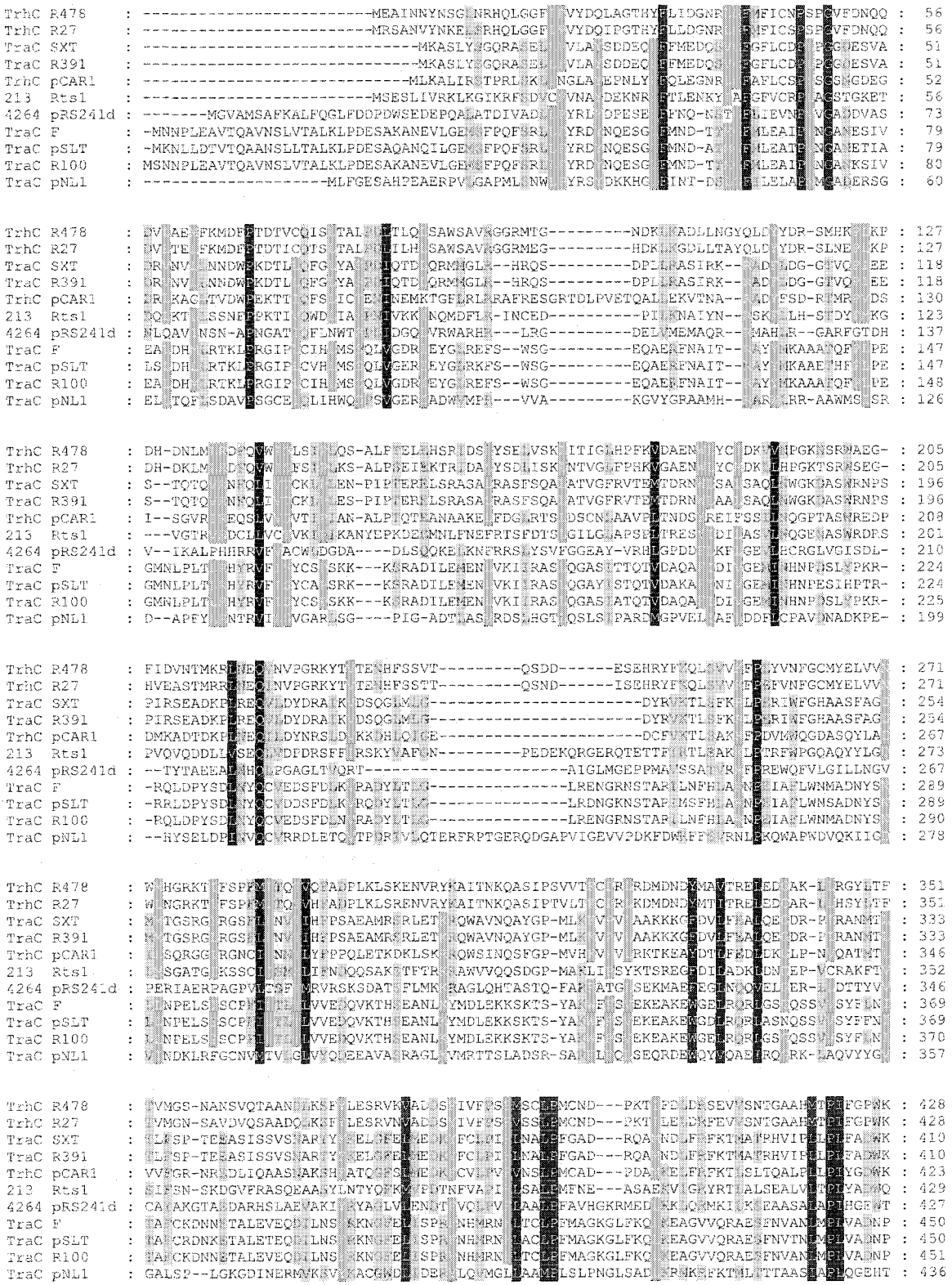
3.2.1 Bacterial strains and plasmids

All strains and plasmids used in this chapter are listed in Table 3-1. Antibiotics were used at the following concentrations, when appropriate: ampicillin (Ap) 100 μg/ml, chloramphenicol (Cm) 16 μg/ml, rifampicin (Rif) 20 μg/ml, nalidixic acid (Nal) 30 μg/ml, tetracycline (Tc) 10 μg/ml, and kanamycin (Km) 10 μg/ml. Conjugal matings were performed for the indicated time and carried out as previously described (Taylor and Levine, 1980).

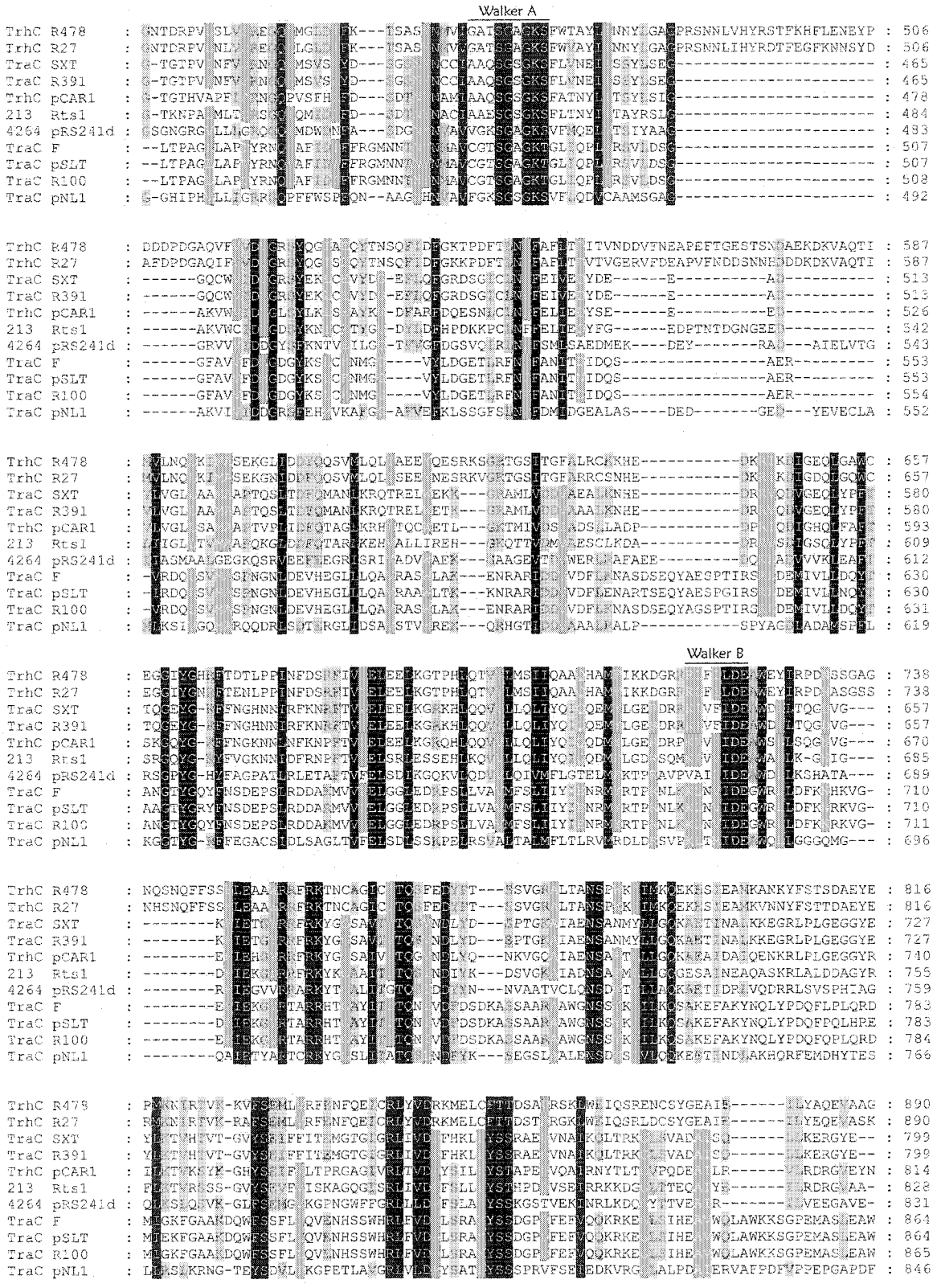
Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Figure 3-1. Multiple sequence alignment of TrhC-related sequences. The alignment was generated using ClustalW in MegAlign (DNAStar) and conserved residues were shaded using GENEDOC. Walker NTP-binding motifs are indicated. Accession numbers: R478 (CAE51543), R27 (NP_058238), SXT (AAL59681), R391 (AAM08001), pCAR1 (NP_758682), Rts1 (NP_640173), *Rhodobacter sphaeroides* (Rsp: ZP_00008315), F (AAC44180), pSLT (NP_490573), R100 (NP_052960), pNL1 (NP_049162).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

```
TrhC R478     : LAA--------------- : 893
TrhC R27      : SAA--------------- : 893
TraC SXT      : ------------------ :   -
TraC R391     : ------------------ :   -
TrhC pCAR1    : PDPAEVA----------- : 821
213  Rts1     : ------------------ :   -
4264 pRS241d  : ------------------ :   -
TraC F        : LEEHEKYRSVA------- : 875
TraC pSLT     : LQEHETFRGTYEYKAETD : 882
TraC R100     : LEEHEKYRSVA------- : 876
TraC pNL1     : AEGELAEAAE-------- : 856
```


Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 3-1. Plasmids and *E. coli* strains used in this chapter.

Strain or Plasmid	Description	Selective marker(s)[b]	Source
E. coli strain			
DH5α	λ⁻ ϕ80d*lacZ*ΔM15 Δ(*lacZYA-argF*)*U169 recA endA1 hsdR17*($r_k^-m_k^+$)		Invitrogen
DH10B	F-*mrcA* Δ(*mrr-hsd*RMA-*mrcBC*) ϕ80d*lacZ*ΔM15 Δ*lacX74 deoR recA1 endA1 araD139* Δ(ara,leu) 7697 *galU galK* λ-*rpsL hupG*		Invitrogen
DY330	W3110 Δ*lacU169 gal490* λ*c1857* Δ(*cro-bioA*); spontaneous Nal^r or Rif^r	Nal or Rif	Yu *et al.*, 2000
J53	*pro met* F^-	Nal	Bachman, 1972
JE2571	*leu thr str fla pil*	Nal or Rif	Bradley, 1980
RG192	*ara leu lac*	Rif	Taylor, 1983
XL1-Blue	*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1* lac [F' *proAB* $lacI^qZ\Delta m15$ Tn*10*]	Tc	Stratagene
Plasmids			
pDT1942	Derepressed transfer mutant of R27	Tc, Km	Maher *et al.*, 1991
pDT2953	pDT1942 with mini::Tn*10* inserted into *trhU*	Tc, Km, Cm	Rooker *et al.*, 1999
pDT2957	pDT1942 with mini::Tn*10* inserted into *trhC*	Tc, Km, Cm	Taylor *et al.*, 1999
pDT2959	pDT1942 with mini::Tn*10* inserted into *trhW*	Tc, Km, Cm	Rooker *et al.*, 1999
pDT2969	pDT1942 with mini::Tn*10* inserted into *trhC*	Tc, Km, Cm	Taylor *et al.*, 1999
pDT2970	pDT1942 with mini::Tn*10* inserted into *trhE*	Tc, Km, Cm	Rooker *et al.*, 1999
pDT2971	pDT1942 with mini::Tn*10* inserted into *trhB*	Tc, Km, Cm	Rooker *et al.*, 1999
pDT2978	pDT1942 with mini::Tn*10* inserted into *trhC*	Tc, Km, Cm	Taylor *et al.*, 1999
pDT2981	pDT1942 with mini::Tn*10* inserted into *trhK*	Tc, Km, Cm	Rooker *et al.*, 1999
pDT2984	pDT1942 with mini::Tn*10* inserted into *trhH*	Tc, Km, Cm	Lawley *et al.*, 2002a
pDT2987	pDT1942 with mini::Tn*10* inserted into *trhF*	Tc, Km, Cm	Lawley *et al.*, 2002a
pDT2988	pDT1942 with mini::Tn*10* inserted into *trhN*	Tc, Km, Cm	Rooker *et al.*, 1999
pDT2989	pDT1942 with mini::Tn*10* inserted into *traG*	Tc, Km, Cm	Lawley *et al.*, 2002a
pDT2991	pDT1942 with mini::Tn*10* inserted into *trhG*	Tc, Km, Cm	Lawley *et al.*, 2002a
pDT2995	pDT1942 with mini::Tn*10* inserted into *traI*	Tc, Km, Cm	Lawley *et al.*, 2002a
pDT3048	R27 with *gfp/cat* inserted in-frame with *trhC*	Tc, Cm	Gilmour *et al.*, 2001
pGEM-T	Cloning vector for PCR products	Ap	Promega
pJEG51	pDT1942 with *cat* inserted into *traJ*	Tc, Km, Cm	Lawley *et al.*, 2002a
pJEG94	pDT1942 with *cat* inserted into *trhR*	Tc, Km, Cm	Lawley *et al.*, 2002a
pJEG104	pDT1942 with *cat* inserted into *traH*	Tc, Km, Cm	Lawley *et al.*, 2002a
pJEG115	pDT1942 with *cat* inserted into *trhY*	Tc, Km, Cm	Lawley *et al.*, 2002a
pGUN110	pDT1942 with *cat* inserted into *trhA*	Tc, Km, Cm	Lawley *et al.*, 2003a
pMS119EH	Cloning vector, P_{tac}-$lacI^q$, pMB1 origin of replication	Ap	Strack *et al.*, 1992
pMWG36	$TrhC_{His6}$ expression vector (pMS119EH)	Ap	Gilmour *et al.*, 2001
pMWG133	*gfp/cat* cloned into pGEM	Ap, Cm	Gilmour *et al.*, 2001
pMWG191	TrhC-GFP expression vector (pMS119EH)	Ap	Gilmour *et al.*, 2001
pMWG253	pDT1942 with *cat* inserted into *trhL*	Tc, Km, Cm	Lawley *et al.*, 2003a
pMWG265	pDT1942 with *cat* inserted into *trhV*	Tc, Km, Cm	Lawley *et al.*, 2003a
R27	IncHI1	Tc	Taylor, 1983

a. Antibiotic resistance markers: Ap, ampicillin; Tc, tetracycline; Km, kanamycin; Cm, chloramphenicol; Nal, nalidixic acid; Rif, rifampicin

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

HgaI RNA bacteriophage was prepared as previously described (Maher *et al.*, 1991). To determine HgaI infectivity of TrhC mutants, phage spot tests were performed. A 50 μL sample of mid-logarithmic phase culture was suspended in 8 ml of brain-heart infusion (BHI) broth (Difco) + 0.6% agar at 48°C. The suspension was immediately overlayed on a fresh BHI plate. When cultures contained a pMS119EH expression vector, ampicillin and IPTG were included in the overlay. After solidification of the overlay 10 μL of HgaI stock (9.5×10^{10} PFU/ml) was spotted at discrete points on the overlay. Plates were incubated at 27°C and analyzed for zones of clearing in the phage-treated regions.

3.2.2 DNA and RNA techniques

The *gfp/cat* linear DNA cassette was constructed by recombinant PCR. First round PCR amplification of the green fluorescent protein gene *gfpmut2* (Cormack *et al.*, 1996) was achieved using primers GIL78 (5'-GATTCTGTACGAACAGGAAGTTGCGAGCAAATCAGCTG CCATGAGTAAAGGAGAAGAAC) and GIL79 (5'-GAAGTGATCTTCCGTCACAGTTATTTGTATAGTTCA TCCATG) and amplification of the Tn*9* chloramphenicol acetyl transferase gene *cat* (including the native promoter) was with primers GIL80 (5'-CATGGATGAACTATACAAATAAC TGTGACGGAAGATCACTTC) and GIL81 (5'-ACCCGGGCTATTATTTCGGGGGAGGGCAGGCAATGA CAGCTTATTCAGGCGTAGCACCAG). Each first round product was ethanol precipitated, re-dissolved in de-ionized water, and used as templates for a second round PCR to adjoin *gfp* and *cat* via overlap extension and amplification with the outside primers GIL78 and GIL81. The recombinant product was purified from an agarose gel using the Qiaquick gel extraction kit (Qiagen) and ligated into pGEM-T (Promega) to create pMWG133 and transformed into DH5α. Clones containing the *gfp/cat* construct were selected for with Ap and Cm.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

To insert *gfp* in-frame with the end of *trhC*, linear *gfp/cat* DNA was produced through PCR of pMWG133 using primers GIL78 and GIL81, purified using Qiaquick PCR purification columns (Qiagen), and electroporated into *E. coli* strain DY330(R27) following the procedure outlined by Yu *et al.* (Yu *et al.*, 2000). DNA was treated with *Dpn*I (Life Technologies) prior to purification to remove remaining circular pMWG133. Recombinants were selected on Tc and Cm. To verify insertion of the *gfp/cat* cassette into R27 after the penultimate codon of *trhC*, clones were sequenced using primers GIL54 (5'TGGTGCTACGCCTGAATAAG) and GIL88 (5'-AAGAATTGGGACAACTCCAGTG). Strain DT3048 contains R27::*trhC-gfp*. The *trhC-gfp* fusion product was amplified and subsequently cloned into pMS119EH (to create pMWG191) using the forward primer GIL20 (5'-TATGGTACCATGAGATCAGCTAACGTCTAT) and reverse primer GIL89 (5'-TATCTGCAGTTATTTGTATAGTTCATCCATG) and pDT3048 as the template DNA.

To clone *trhC* with a C terminal His_6-coding tag, the forward primer GIL20 the reverse primer GIL06 (5'-TATCTGCAGTCAGTGATGGTGATGGTGATGTGCGGCAGCTGATTTGCTCGCAAC) were used. The PCR product was initially cloned into pGEM-T (Promega) and then subcloned into pMS119EH (see Table 3-1) using the restriction enzymes *Kpn*I and *Pst*I and T4 Ligase (Invitrogen).

Plasmids pMWG191 (encoding *trhC-gfp*) and pMWG36 (encoding $trhC_{His6}$) served as templates for site-directed mutagenesis of Walker A and B-encoding regions of TrhC. To introduce substitutions and deletions into *trhC*, a complementary pair of oligonucleotide primers containing mismatches or deletions in comparison to the template sequence were used in a thermocycling reaction with *PfuTurbo* (Stratagene) to synthesize full-length DNA molecules from the plasmid templates. These primers were based upon the following sequence from the *trhC* sense strand: 5'GCAACGTCTGGTGCTGGTAAATCATTCTGGGTCGCATAC (Walker A) and 5'-CGGGTTGTTCATACTTGATGAGGCATGGGAGTATATCCGTCCG (Walker B), where the underlined segments represent regions containing

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

mutation-specific mismatches or deletions (detailed below). The parental wild-type plasmid DNA was selectively destroyed using *Dpn*I (Invitrogen), which recognizes and cleaves methylated DNA substrates. The newly synthesized, un-methylated DNA containing mutations was transformed into XL1-Blue *E. coli* ultra-competent cells (Stratagene) for circularization and propagation. The plasmid constructs were sequenced in the target area (using primers: Walker B: 5'-ACACCTTCAGACTGTCGTTCTG; Walker A: 5'-CTCCATCAGGGTCAAATGCATC) and analyzed by restriction digestion to ensure that mutated constructs of the proper size had been generated. All pMWG191 wild-type and mutation constructs were then transformed into RG192 cells containing pDT2969 for fluorescence assays, whereas pMWG36 wild-type and mutation constructs were transformed into RG192 cells containing pDT2957, pDT2969, or pDT2978 for conjugative transfer complementation and bacteriophage propagation assays.

For RT-PCR experiments, RNA was purified from 20 ml cultures of DY330(pDT3048) grown overnight at 27°C or 37°C. The cells were pelleted and then used in the GenElute Mammalian Total RNA Isolation Kit (Sigma), following manufacturer's directions. The resulting RNA was then quantitated with an Ultrospec 3000 (Pharmacia). The RNA was treated with DNase I (Invitrogen) to remove any residual DNA that could contaminate the sample. cDNA was made from the RNA using Superscript II reverse transcriptase (Invitrogen), following manufacturer's protocols. The primer used for cDNA synthesis, specific for *gfp* sequence, was GFP-rev (5'-TTTCGAAAGGGCAGATTG). The cDNA was then used in a PCR reaction with GFP-rev and the forward primer TrhC-RT-PCR (5'-ATTCCGGGTACACATTATTTTCTA) specific to *trhC* sequence to amplify a product of 3230 bp. The PCR samples were analyzed by agarose gel electrophoresis and stained with ethidium bromide.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

3.2.3 Microscopy

For visualization of H-pili with the transmission electron microscope the method of Maher *et al.* (Maher and Taylor, 1993) was followed, with the exception that JE2571 cells were grown in LB broth, Lennox formulation (Difco).

For the fluorescence assay, *E. coli* cells were grown overnight in 20 ml LB broth containing antibiotics at either 27°C or 37°C. A 5% inoculum was transferred to 20 ml fresh LB-Lennox containing antibiotics (pre-equilibrated at either 27°C or 37°C) and incubation continued at that temperature with shaking until cells reached OD_{600}~0.4. A 1 ml aliquot was taken and the cells were pelleted in a microfuge. The supernatant was aspirated and cells were washed in 1 ml MOPS minimal media supplemented with 0.5% glucose and 20-amino acid mixture (Gordon *et al.*, 1997), pelleted and aspirated again. The cells were resuspended in a final volume of 50 μl of the above MOPS media. A 1 μl aliquot was spotted on a MOPS+glucose-1.5% agarose slab on a glass microscope slide. A cover slip was placed on and sealed, and the cells visualized using a UV light source through a Leica DMRE microscope equipped with a HQF filter and a Cooke SensiCam. Images were collected and processed using SensiControl 4.0 and PhotoPaint 8.0 (Corel) software.

3.2.4 Purification of TrhC and antibody production

Bacterial cultures of DH10B containing pMWG36 (4 litres) were grown at 30°C to OD_{600}~ 0.6 and induced with 0.5 mM IPTG for 1.5 hours. Cells were harvested by centrifugation (Beckman JLA 8.1 rotor, 15 minutes, and 7000xg) and resuspended in buffer A (20 mM Tris-HCl pH 7.6, 10% glycerol, and 0.5 M NaCl) with 50mM imidazole at a concentration of 3 ml/g of cells and 1 ml/20g of cells protease inhibitor cocktail (Sigma). Cells were lysed by three passages through a French press (16 000 p.s.i.) and the cytoplasm collected by centrifugation (Ti70.1 rotor, 100 000xg, 2 hours). The

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

supernatant was applied to 2 ml of Ni-NTA superflow resin (Qiagen) equilibrated in buffer A with 50 mM imidazole in a 1.5 x 10 cm Glass Econo-Colomn (Biorad). The column was washed three times with 5 ml of buffer A with 50 mM imidazole and TrhC was eluted with buffer A plus 200 mM imidazole. Fractions containing TrhC were concentrated with Centriprep-50 centrifugal concentrators (Amicon). Glycerol content was increased to 40% and the concentration of TrhC was estimated to be 2mg/ml. To further purify TrhC during antigen preparation, total protein (with addition of 100 mM Tris-HCl pH 7.6, 2% SDS, and 100 mM DTT) was loaded into a 10% SDS acrylamide gel. After electrophoresis the gel was washed three times in Milli-Q water then stained for ten minutes in 0.05% Coomassie stain. The gel was destained in Milli-Q water until the band corresponding to TrhC was discernable. This band was excised from the gel, air-dried, and ground to a fine powder with a mortar and pestle. The band isolated was verified to be TrhC by N-terminal sequencing through Edman degradation. The protein-acrylamide powder was resuspended in PBS until it reached the consistency to pass through a 21-gauge syringe, and then frozen until required for injection. Antibodies to TrhC were raised in New Zealand White rabbits. A primary inoculation of 1 mg of TrhC with complete Freund's adjuvant (Sigma) per rabbit was proceeded by a secondary inoculation on day 50 (0.5 mg) and a third injection on day 95 with another 0.5 mg, each with incomplete Freund's adjuvant. Production of TrhC-specific antisera was followed by immunoblot analysis. Proteins were resolved on 8% acrylamide gels, transferred to nitrocellulose (Micron Separations Inc.), and blocked overnight in 15% skim milk, PBS, and 0.1% Tween-20. Primary antibody was added at 1:1000 and secondary antibody (horseradish peroxidase conjugated goat anti-rabbit IgG, Sigma) at 1:2000. Detection was by either the ECL kit (Amersham Pharmacia) or with the SuperSignal West Dura extended duration substrate (Pierce).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

3.2.5 Purification of TrhC for enzymatic analysis

Bacterial cultures were grown and induced as described in section 3.2.4, and prior to lysis via the French press, cells were resuspended in 25 ml of buffer C (10mM Tris-HCl, pH 7.6, 5 mM $MgCl_2$, 2.5 M NaCl) with 50 mM imidazole. TrhC was purified using Ni-NTA resin as described in section 3.2.4, except buffer C was used for washes (with 50 mM imidazole) and elution (7 ml with 350 mM imidazole). Protein content of the 1ml fractions was estimated using a standard Bradford solution assay, and three fractions containing the highest concentrations of protein were desalted using a PD-10 column (Amersham Biosciences) equilibrate with buffer D (50 mM Tris-HCl, pH 7.6, 10 mM $MgCl_2$). Seven 1 ml fractions were collected and again analyzed for protein content using the Bradford assay. Protein-containing fractions were loaded into a MonoQ anion-exchange fast protein liquid chromatography (FPLC) column using buffer D at flow rate of 1 ml/min for 10 min. To elute bound proteins, a linear gradient of buffer E (1 M KCl) was used, starting at 0% and rising to 100% over the course of 20 min at a flow rate of 1ml/min. Protein content of was analyzed at 215 nm and 280 nm during elution and 1ml fractions were collected. An aliquot from each fraction was analyzed by SDS-PAGE.

3.2.6 ATPase assay

Reaction buffer (50 mM Tris-HCl, pH 7.6, 5 mM $MgCl_2$, 1 mM DTT) and ATP mix (1 mM cold ATP and 0.04 μCi ^{32}P-γ-ATP) were combined in 1.5 ml tube on ice. At time = 0, preparations of TrhC were added and the assay mixture (total volume, 100 μl) was incubated at 28°C. At specific intervals, 30 μl of the assay mixture was removed, added to 500 ml of stop solution (67% vol/vol 2-butanol, 133 mM H_2SO_4, 0.4 mM NaH_2PO_4,13 mM $MoNa_2O_4 - 2H_2O$), and vortexed for 1 min. The organic phase containing cleaved γ-phosphates was recovered after centrifugation (5 min, 14000 *g*) and added to 5 ml of scintillation fluid. The radioactivity of each sample was quantified in a scintillation

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

counter calibrated for ^{32}P isotopes. To determine the background level of ATP hydrolysis in the liquid phase, assay mixtures containing only reaction buffer and ATP mix were incubated at 28°C for 30 min and processed as described above.

3.2.7 Localization of TrhC

To determine the approximate cellular location of TrhC, membrane isolation and fractionation as described by Grahn *et al.* (Grahn *et al.*, 2000) was performed. Using this procedure, cytoplasmic and whole membrane fractions were isolated. The membrane fraction was further separated into cytoplasmic and outer membrane fractions by isopycnic sucrose gradient centrifugation. 1 ml fractions collected from the bottom of the ultracentrifugation tube were analyzed for TrhC content by immunoblotting (see above).

3.3 Results

3.3.1 Insertion of *gfp/cat* into R27 to create a *trhC-gfp* fusion

In order to create a *trhC-gfp* fusion that would be expressed at the same rate and under the same conditions as native *trhC*, the fusion of *gfp* to *trhC* was performed using whole R27 plasmid DNA (Figure 3-2). The *gfpmut2* gene (Cormack *et al.*, 1996) was included as part of a linear DNA construct comprising the chloramphenicol acetyl transferase (*cat*) gene and terminal arms homologous to R27. Utilizing the λ prophage recombination system in the host strain *E. coli* DY330 (Yu *et al.*, 2000), the homologous arms directed recombination of the *gfp/cat* cassette into R27 at positions 28485-28639 to create pDT3048 (R27::*gfp/cat*, see Table 3-1). This resulted in a fusion construct encoding a C terminal fusion of GFP to TrhC.

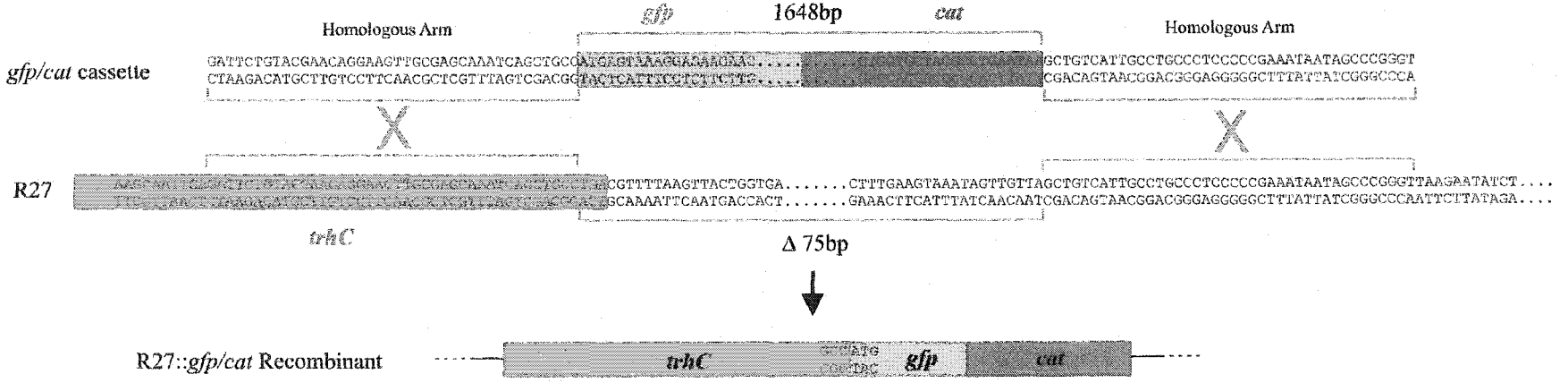
The *trhC* gene was a candidate for *gfp* fusion due to the genetic organization of *trhC* in R27 and the predicted membrane orientation of TrhC. Knowing the membrane

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Figure 3-2. Schematic diagram of *gfp/cat* cassette insertion into R27 to create a *trhC-gfp* translational fusion. The *cat* segment includes 197 bp upstream of the start codon and 54 bp downstream of the stop codon, therefore the native cat promoter is present.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.



topology of the VirB4 family of proteins, TrhC is predicted to be an inner membrane protein with the C terminal end of the protein remaining in the cytoplasm. Since GFP is not stable in the periplasm (Feilmeier *et al.*, 2000), a C-terminal fusion to TrhC is an appropriate choice to detect GFP fluorescence. In addition, the next open reading frame downstream of *trhC*, R0023, is 668 bp away, thereby allowing insertion of the *gfp/cat* construct without disrupting essential transfer genes.

The R27::*gfp/cat* construct pDT3048 was competent for plasmid transfer, albeit with a two-log decrease in transfer frequency compared to wild-type R27 (Table 3-2). *E. coli* DY330R carrying pDT3048 were resistant to lysis by the H-pilus-specific RNA phage Hgal, whereas *E. coli* carrying wild-type R27 were sensitive to Hgal, as measured in a phage spot assay (Table 3-2). In addition, no pili were observed on JE2571 cells carrying pDT3048 under transmission electron microscopy (data not shown). The lack of pili would result in the concurrent lack of RNA phage propagation, as there would be no structure to which the phage could adhere. The *gfp/cat* insertion was confirmed to have no polar effects as production of TrhC with a C terminal His_6 tag (pMWG36) *in trans* restored pDT3048 plasmid transfer frequency back to the wild-type level (Table 3-2). Using R27 with a transposon knockout mutant of *trhC* (pDT2969), the ability of TrhC-GFP to act in plasmid transfer was confirmed. Insertion of mini-Tn*10* into *trhC* abolished transfer of R27, H-pilus production, and infectivity of Hgal, whereas expression of TrhC from pMWG36 complemented pDT2969 to wild-type transfer frequency and restored H-pilus production and Hgal propagation (Table 3-2). The *trhC-gfp* fusion was cloned into the pMS119EH expression vector (pMWG191) and expression of TrhC-GFP enabled pDT2969 to transfer, however at a lower frequency than achieved by $TrhC_{His6}$. Furthermore, complementation with TrhC-GFP was unable to restore Hgal propagation and no pili were detected (Table 3-2). Therefore, the fusion of GFP to TrhC still supports R27 plasmid transfer with both an *in situ* R27 fusion and during *trans* complementation

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 3-2. Conjugative transfer, RNA phage propagation, and H-pilus production by R27 and derivatives

Plasmid(s) in the donor strain	Transfer frequency[a]	Propagation of Hgal[b]	Production of H-pilus[c]
R27	3.4×10^{-1}	+	+
pDT3048	1.3×10^{-3}	-	-
pDT3048 + pMWG36	2.8×10^{-1}	+	+
pDT2969	0	-	-
pDT2969 + pMWG36	8.3×10^{-1}	+	+
pDT2969 + pMWG191	1.5×10^{-5}	-	-

a. transconjugants/donor, 18 hour matings

b. as determined in a agar plate spot test; "-" indicates no zones of lysis were observed

c. as determined in the transmission electron microscope; "-" indicates no pili were observed on the examined cells

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

of a *trhC* knockout, but the proficiency of TrhC-GFP to fulfill roles in pilus production, phage entry, and conjugation is reduced. It seems likely that TrhC-GFP is maintaining association with the conjugative apparatus because plasmid transfer is still occurring.

3.3.2 Localization of TrhC to the cytoplasmic membrane

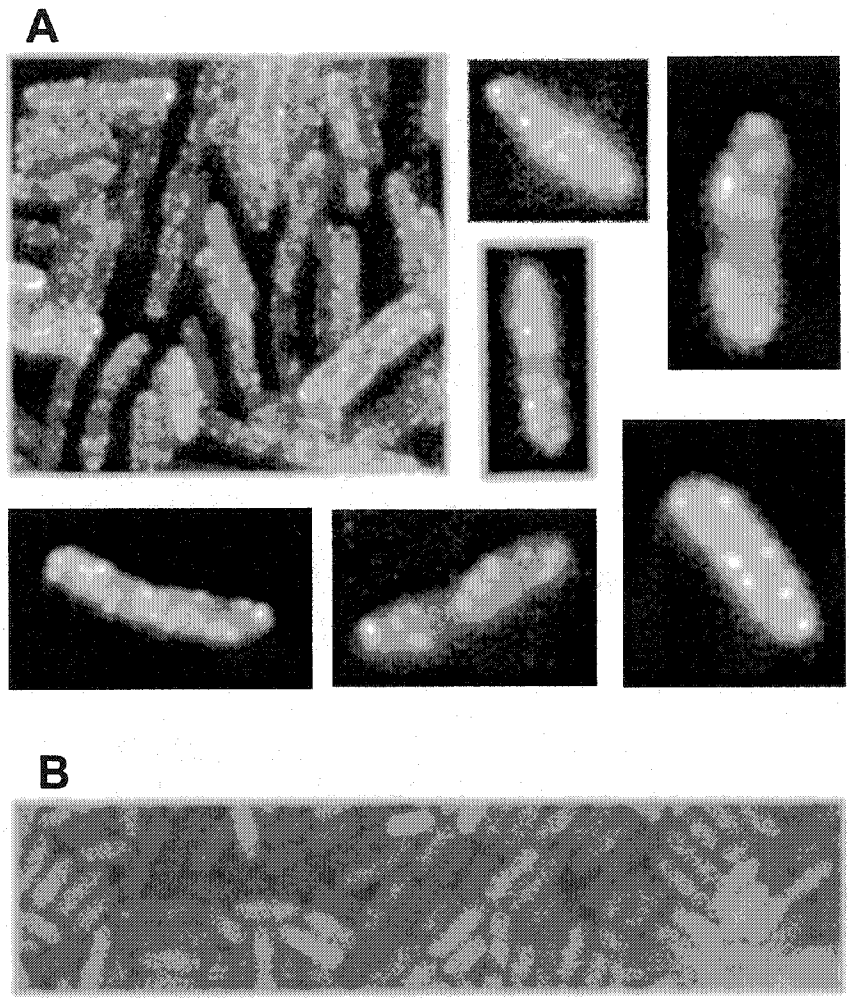
GFP labeling of both DNA and protein have been used as a tool for subcellular localization. The LacI-GFP system has been used to track the segregation of bacterial chromosomes (Gordon *et al.*, 1997) and conjugative plasmids (Pogliano *et al.*, 2001). Fusion of GFP to the DNA replication protein PolC in *Bacillus subtilus* allowed determination of distinct sites for the DNA replication machinery (Lemon and Grossman, 2000). Under fluorescence microscopy, *E. coli* cells carrying pDT3048 contained discrete fluorescent foci primarily at the cell periphery, implying a membrane association of TrhC (Figure 3-3A).

R27 has recently been shown to segregate to and maintain a distinct subcellular localization at the quarter cell (1/4 and 3/4) and mid cell positions (Lawley, 2003). To determine if there could be a correlation between the cellular location of R27 and the cellular location of the R27 transfer protein TrhC, the relative cellular position of TrhC-GFP was examined through spatial analysis of the GFP foci in 118 cells grown at 27°C. There was no evidence that TrhC-GFP co-localizes to these plasmid domains. Each cell was divided into six equal portions across the lateral surface and individual foci cataloged into one of the six segments. The first and last segments represented the polar regions and a combined one-third of the entire cell length. Likewise, the second and fifth segments were combined to represent the quarter cell position, and the third and fourth segments added together and designated as the mid cell position, where each of these positions also represents one-third of the entire cell length. There was an average of 6 discrete fluorescent foci per cell with 39% of foci found in the polar

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Figure 3-3. Fluorescent visualization of *Escherichia coli* DY330R(pDT3048) at 27°C and 37°C. Bacteria were grown exclusively at either 27°C (panel A) or 37°C (B) and visualized under UV illumination.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

positions, 32% in the quarter cell positions, and 29% in the mid cell position. Since it is unlikely a single GFP molecule would be detected in our system, TrhC-GFP therefore seems to be present in protein complexes which are randomly distributed throughout the bacterial cell wall, with a moderate bias for the polar regions. Additionally, the location of pili on the outside of cells carrying R27 or TrhC-complemented pDT3048 appeared to be random, as observed during electron microscopy (data not shown).

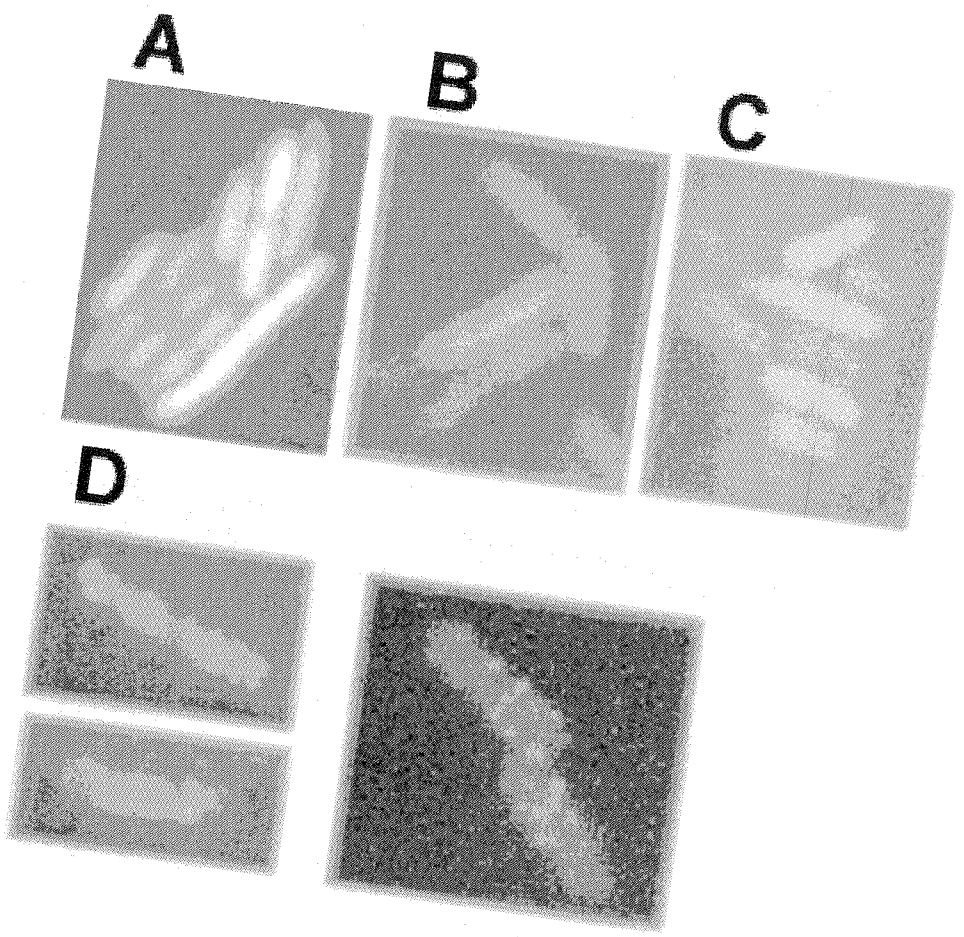
When *trhC-gfp* expression was placed under control of the P_{tac} promoter (pMWG191) there was a uniform pattern of GFP fluorescence at both 37°C and 27°C (Figures 3-4A and 3-4B, respectively). Cells fluoresced with an entirely homogeneous and ubiquitous pattern: there were neither discrete foci nor a distinguishable subcellular localization pattern for the fusion protein. When pMWG191 was transformed into *E. coli* RG192 containing the R27 transfer mutant pDT2969 (see Table 3-1), cells still fluoresced with a homogeneous pattern when grown at 37°C (Figure 3-4C). However, when grown at 27°C, fluorescent foci were visualized at the cell periphery (Figure 3-4D). These data imply that R27-encoded proteins are required for the formation of membrane-associated fluorescent foci of TrhC-GFP, and this process is temperature-dependent.

E. coli cells expressing TrhC (pMWG36) and TrhC-GFP (pDT3048) were separated into whole cytoplasm and membrane fractions. The total membrane fraction was then further resolved into outer and cytoplasmic membrane fractions by isopycnic sucrose gradient centrifugation and the TrhC content of each fraction was assessed by immunoblot analysis with anti-TrhC antibodies (Figure 3-5). Membrane separation was confirmed by SDS-PAGE (data not shown) and density measurement of individual sucrose fractions. TrhC expressed from pMWG36 localized to both the cytoplasm and membrane, and within the latter preparation, entirely to the cytoplasmic (inner) membrane in sucrose fractions of 1.14g/cm^3. The cytoplasmic TrhC may represent unsedimented membrane vesicles containing TrhC or an artifact of over-expression.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Figure 3-4. The formation of fluorescent foci is dependent upon the presence of R27-encoded proteins. DH5α(pMWG191) grown at 37°C (panel A) or 27°C (B) and RG192(pMWG191+pDT2969) grown at either 37°C (C) or 27°C (D) were visualized under UV illumination. Cultures were not induced with IPTG.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

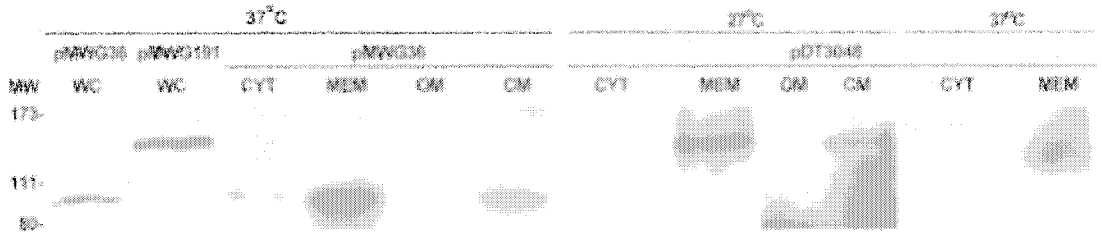


Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

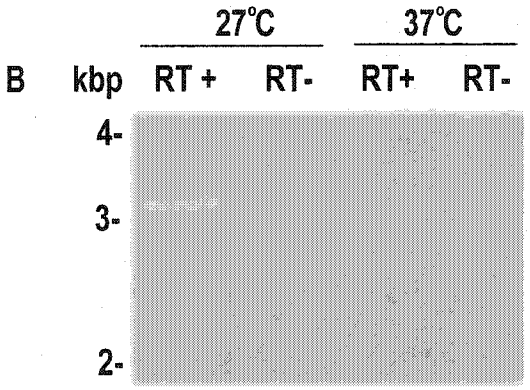
Figure 3-5. TrhC and TrhC-GFP localization and temperature-dependent expression. Whole cell lysates (WC), total cytoplasmic fractions (CYT), total membrane fractions (MEM), outer membrane fractions (OM), and cytoplasmic membrane fractions (CM) prepared from cells expressing TrhC (pMWG36) and TrhC-GFP (pMWG191 and pDT3048) were probed with anti-TrhC polyclonal antibody after SDS-PAGE and transfer to nitrocellulose. All cells were grown at 37°C unless indicated (panel A). DY330R(pDT3048) was grown at either 27°C or 37°C and mRNA was isolated for RT-PCR experiments specific for *trhC-gfp* mRNA (B). PCR reactions were performed on cDNA templates prepared in the presence (RT+) or absence (RT-) of reverse transcriptase.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

A



B



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Since these microscopic studies involved examination of TrhC-GFP, it was also appropriate to determine into which membrane the GFP fusion protein inserted. TrhC-GFP, expressed from pDT3048, was found in the cytoplasmic membrane in sucrose fractions of 1.15g/cm^3 (Figure 3-5). These results confirm that both TrhC and TrhC-GFP are components of the cytoplasmic membrane.

3.3.3 Specific R27 transfer proteins are required for assembly of the TrhC-GFP complex

To verify that the fluorescent foci visualized in cells with pDT3048 (or pDT2969 + pMWG191) result from an interaction between TrhC-GFP and other transfer proteins, TrhC-GFP was produced in the presence of R27 transfer mutants. The key observation relating to this experiment is the dependency of foci formation on the presence of R27-encoded proteins (see Figure 3-4). This implied that R27-encoded proteins were responsible for foci formation, and were likely part of a protein complex containing enough TrhC-GFP molecules to be visible.

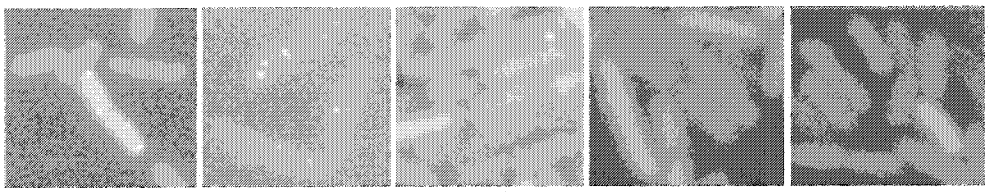
Our group have previously generated transfer mutants in two separate regions of R27 encoding transfer functions (Tra1 and Tra2) by transposon and site-directed mutagenesis (Table 3-1), and in this study each of these transfer-deficient mutants were used in conjunction with pMWG191 to evaluate the requirements for TrhC-GFP to enter into membrane-associated protein complexes. Notably, each of the R27 transfer mutants have previously been complemented for transfer by *in trans* production of the respective wild-type transfer protein (Lawley *et al.*, 2002a; Lawley *et al.*, 2003a), therefore extreme polar effects caused by the gene disruptions were not observed.

After separately transforming the *trhC-gfp* expression plasmid pMWG191 into *E. coli* cells containing one of the nineteen different R27 transfer mutants, preparations of live cells were visualized under UV illumination (Figure 3-6). Twelve of the nineteen R27 transfer mutant were unable to support formation of fluorescent foci, as these cells

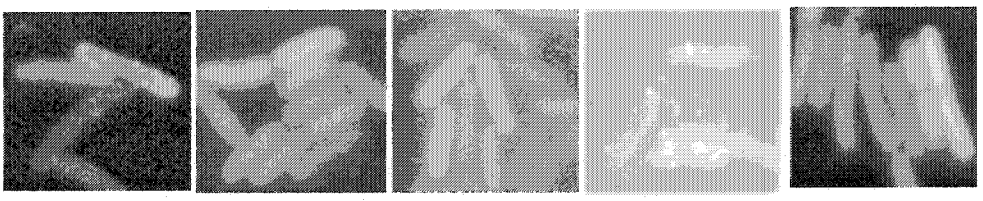
Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Figure 3-6. Fluorescence microscopy of live *E. coli* cells harbouring pMWG191 and a R27 transfer mutant. "n" cells were evaluated for the presence of fluorescent foci under the same criteria as previously described in the Experimental procedures. Solid boxes indicate cells having fluorescence patterns similar to R27 + pMWG191, while broken-line boxes indicate fluorescence patterns having foci, but not to the full extent as the positive control.

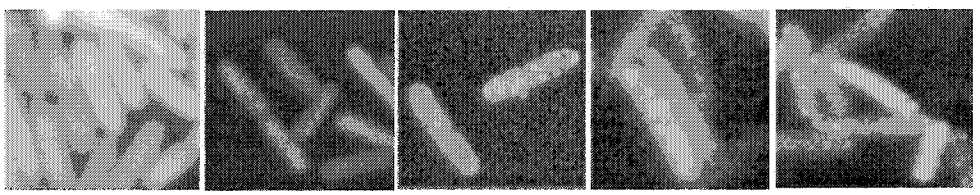
Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.



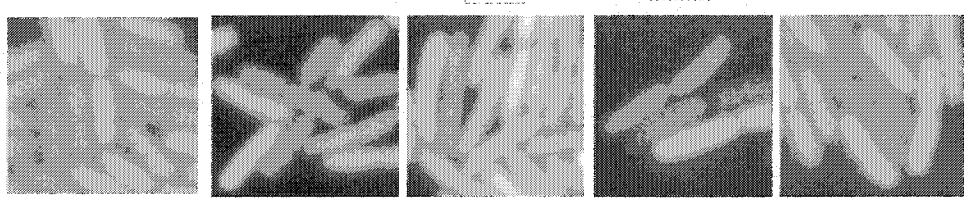
R27 mutant	wild-type	*trhC*	*trhA*	*trhL*	*trhE*
n	259	603	99	101	149
Ave. # foci/cell	3.4	3.8	2.7	0.1	0.1
% cells w/foci	89.2	92.9	70.1	1.0	5.4



R27 mutant	*trhK*	*trhV*	*trhB*	*trhW*	*trhU*
n	82	405	142	48	114
Ave. # foci/cell	2.2	0.2	0.0	3.5	0.2
% cells w/foci	51.2	6.2	1.4	91.7	13.2



R27 mutant	*trhN*	*traJ*	*traG*	*traI*	*traH*
n	103	155	44	59	94
Ave. # foci/cell	0.2	0.4	1.3	1.4	4.4
% cells w/foci	9.7	12.9	43.2	50.9	79.8



R27 mutant	*trhR*	*trhY*	*trhF*	*trhH*	*trhG*
n	140	120	189	212	248
Ave. # foci/cell	0.0	0.0	0.2	0.3	0.2
% cells w/foci	3.6	0.8	8.5	15.1	10.1

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

contained on average between 0.0 and 0.4 foci per cell, and the percentage of cells containing ≥1 focus ranged from 0.8%-15% among the different mutants (Figure 3-6). In these instances where a particular mutation prohibits foci formation, the missing transfer protein is probably a component of the TrhC-GFP complex or is directly required for its assembly. The loss of these requisite components likely results in incomplete (i.e. transfer-deficient) subassemblies that are below the threshold level of visualization.

In cells containing pMWG191 and R27 transfer mutants of *trhC*, *trhA*, *trhK*, *trhW*, *traG*, *traI*, or *traH* (Figure 3-6), fluorescent foci were observed. The average number of foci per cell ranged from 1.3 to 4.4, whereas the percentage of cells containing foci ranged from 43% - 93% among these mutants (Figure 3-6). These genes, with the exception of *trhC*, are not strictly required for the assembly of fluorescent foci and, in particular, cells lacking TrhW (predicted inner membrane Mpf protein) or TraH (predicted relaxosomal component) demonstrated no reduction in foci. In contrast, intermediary effects on the formation of foci were observed in cells lacking TrhA (pilin subunit), TrhK (predicted outer membrane secretin), TraG (inner membrane coupling protein), and TraI (DNA relaxase). These results suggest that several of the essential conjugative transfer proteins are not required for focus formation. The exact composition of the TrhC-GFP protein complex was not obtained in these experiments, and each of the above six transfer proteins may still be components of the TrhC-containing complex, but are not absolutely required for its assembly. The intermediate effects seen with four transfer mutants might have been elicited by unknown consequences of each mutation (including the altered stability of other essential transfer proteins).

In this fluorescence assay, TrhC is produced in two forms: TrhC-GFP (from pMWG191) and TrhC (from all R27 transfer mutants except the *trhC* mutant). To determine the effect of co-production of the two TrhC species for focus formation, cells carrying pMWG191 and wild-type R27 were examined (Figure 3-6). The fluorescence

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

patterns of these cells was similar to cells carrying pMWG191 and a R27 *trhC* mutant, therefore the presence of native TrhC does not affect TrhC-GFP condensation into foci. Furthermore, the over-production of TrhC-GFP from pMWG191 results in approximately a two-log imbalanced stoichiometry between the fusion protein and the other transfer proteins (our unpublished observations). The excess TrhC-GFP in cells containing foci was likely seen as the confluent (non foci-associated) fluorescence, similar to that seen in cells lacking R27 but producing TrhC-GFP.

3.3.4 Temperature-dependent mechanisms of R27 transfer

The transfer of IncHI plasmids is temperature-sensitive, as conjugation proceeds between 22-30°C, but not at 37°C. The temperature of the donor cells prior to and during mating is critical, not the temperature of the recipient (Taylor and Levine, 1980). The mechanism(s) underlying this phenomenon has not been revealed. Peter Newnham has previously examined the temperature-dependent kinetics of IncHI1 plasmid transfer by determining the effect of high temperature on mating frequency (Newnham, 1995). After pulsing mating mixtures at 37°C for various times, a direct negative logarithmic relationship existed between the time of exposure at 37°C and extent of DNA transfer, with a transfer frequency reduction rate of approximately 27% per minute pulsed at 37°C (Newnham, 1995). This suggests that exposure to 37°C causes a rapid and prolonged suppression of the transfer system of R27. This rapid response indicates that a pre-formed component in the transfer apparatus is immediately affected by the temperature shift.

The temperature-dependent synthesis of TrhC was examined using the *trhC-gfp* fusion R27 construct pDT3048. Cultures containing pDT3048 that were grown exclusively at 27°C (Figure 3-3A) showed a marked difference in the level of fluorescence compared with cells grown at 37°C (Figure 3-3B). The majority of the cells

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

grown at 37°C had no detectable GFP fluorescence, as only a small proportion of cells contained GFP foci (<1%). Alternatively, the cells grown exclusively at 27°C almost all contained GFP foci. It is therefore apparent that a 10°C increase in incubation temperature dramatically alters *trhC-gfp* expression and/or the stability of TrhC-GFP.

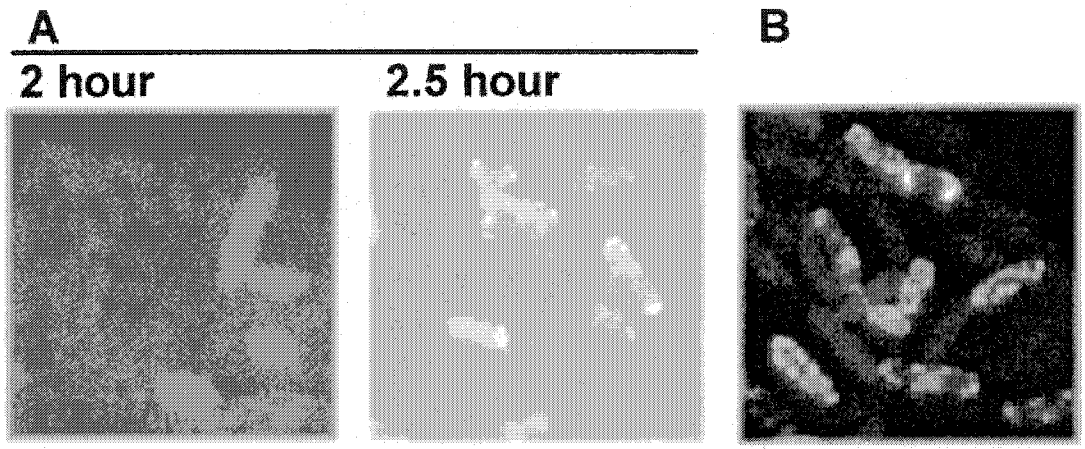
To differentiate between two possibilities underlying temperature-sensitive transfer (transfer protein instability or repressed transfer protein production at 37°C) the fluorescence pattern of TrhC-GFP expressed from pMWG191 (under control of the P_{tac} promoter) at 37°C was examined (Figure 3-4A). The fusion protein fluoresced with same ubiquitous pattern as observed when expressed at 27°C (Figure 3-4B). The fusion protein is therefore stably maintained at the elevated temperature. The temperature-dependent production of TrhC-GFP protein and mRNA was confirmed by immunoblot analysis (Figure 3-5A) and RT-PCR (Figure 3-5B). Neither TrhC-GFP protein or mRNA could be detected in DY330R(pDT3048) grown at 37°C, but was found when cells were cultured at 27°C. These results support a model in which the synthesis of TrhC is temperature-sensitive, likely at the level of transcription.

To further reveal the kinetics of TrhC-GFP expression from pDT3048, temperature shift experiments were conducted. DY330R(pDT3048) was grown at 37°C to mid-logarithmic phase, therefore having little to no TrhC-GFP expression, and then equilibrated to 27°C in a water bath. Incubation (with shaking) was then continued at this lower temperature and samples were taken every half hour after the switch and processed as described (section 3.2.3). On average, 2.5 hours were required for regeneration of GFP foci (Figure 3-7A). Conversely, during shifts from 27°C to 37°C for 60 minutes, the GFP foci remained stable (Figure 3-7B). It appears that TrhC is not the transfer element which is immediately affected by high temperature, but that a considerable lag period is required for TrhC-GFP synthesis after growth at 37°C.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Figure 3-7. Temperature-dependent synthesis of TrhC-GFP foci and stability of TrhC-GFP foci at 37°C. *Escherichia coli* DY330R(pDT3048) was grown at 37°C until mid-logarithmic growth phase, cooled to 27°C (t=0 hour) and incubated at 27°C with shaking. Samples taken at the given time points were visualized using fluorescent microscopy (panel A). DY330R(pDT3048) was grown at 27°C, pulsed at 37°C for one hour, re-equilibrated to 27°C then processed for fluorescent microscopy (B).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

3.3.5 Mutations in the NTP-binding motifs of TrhC do not affect fluorescent foci formation

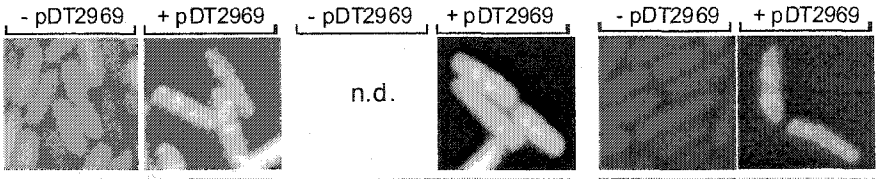
After site-directed mutagenesis using primers to introduce deletions and substitutions in the Walker A and B motif-encoding regions of *trhC* (pMWG36) and *trhC-gfp* (pMWG191, Figure 3-8), the resulting plasmid constructs were transformed into both DY330R cells and DY330R cells containing pDT2969 (a R27 *trhC* transfer mutant). In the Walker A motif (Figure 3-1), mutations were introduced in the region encoding Gly-Lys-Ser (GKS). These include substitution of all three residues to the small, non-polar amino acid alanine (GKS to AAA), a single Ala substitution of the similarly small, non-polar Gly (AKS), Gly substituted to a small, polar Ser (SKS), and the more dramatic substitution of Lys (positively charged, polar) to Glu (acidic side chain, polar; GES). In the Walker B motif the Asp-Glu-Ala-Trp sub-region (DEAW; Figure 3-1) was deleted; and also substituted the highly conserved charged polar residues Asp or Glu (DE sub-region) with the uncharged Asn (NE) and Glu (EE). A tandem substitution in both Walker A and B motifs (AAA/EE) was created using two successive rounds of primer-mediated site-directed mutagenesis.

In the fluorescence assay (Figure 3-8), production of the TrhC-GFP fusion proteins (wild-type or NTP-binding motif mutations) in cells lacking pDT2969 resulted in confluent, uniform fluorescence. For any of these trials in which the complement of R27-encoded proteins was absent, the average number of foci per cell was ≤0.2 and the percentage of cells having one or more foci was ≤5.0% (Figure 3-8). In contrast, production of TrhC-GFP (wild-type or NTP-binding motif mutations) within cells containing pDT2969 resulted in an average of ≥2.9 foci per cell and ≥91% of cells having one or more foci (Figure 3-8). When the extent of foci present in cells producing wild-type TrhC-GFP versus mutant TrhC-GFP were compared, little difference was observed.

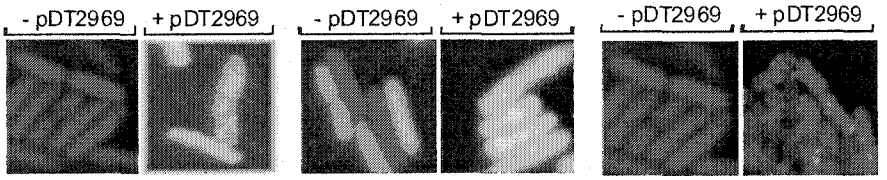
Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Figure 3-8. Fluorescence microscopy of live *E. coli* cells harbouring pMWG191 or derivatives encoding mutations in Walker A or B motifs; with (right panels) or without (left panels) pDT2969. "n" cells were evaluated for the presence of fluorescent foci under the same criteria as described in Experimental procedures. "n.d" is not determined.

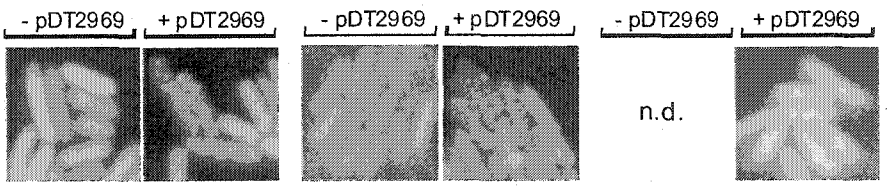
Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.



TrhC-GFP	Wild-type		GKS→AAA		GKS→AKS	
codon change	-		ggtaaatca→gctgcagca		ggt→gct	
n	173	603	-	85	163	153
Average # foci/cell	0.1	3.8	-	2.9	0.1	3.1
% of cells with foci	2.3	92.9	-	94.1	4.3	92.2



TrhC-GFP	GKS→SKS		GKS→GES		Δ DEAW	
codon change	ggt→tct		aaa→gaa		delete gatgaggcatgg	
n	133	107	121	134	111	51
Average # foci/cell	0.1	4.0	0.1	3.3	0.1	3.6
% of cells with foci	3.8	90.7	5.0	93.1	4.3	96.1



TrhC-GFP	DE→NE		DE→EE		AAA/EE tandem	
codon change	gat→aat		gat→gaa		gctgcagca/gaa	
n	107	246	143	47	-	107
Average # foci/cell	0.2	4.1	0.2	3.2	-	3.2
% of cells with foci	2.8	98.3	4.9	91.5	-	93.4

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

These data indicate that the NTP-binding motifs are not required for the assembly of TrhC-associated transfer protein complexes.

3.3.6 Mutations in the NTP-binding motifs of TrhC prohibit plasmid transfer

To determine the effect of NTP-binding motif mutations on conjugative transfer, standard conjugative transfer complementation experiments were conducted. The ability of pMWG36 and derivatives to restore, *in trans*, the transfer of three different R27 *trhC* transfer mutants was tested (Table 3-3). Production of wild-type TrhC from pMWG36 restored transfer of pDT2957, pDT2969, and pDT2978-containing donor cells to wild-type R27 transfer levels (~10^{-1} transconjugant/donor under these conditions, data not shown). All mutated pMWG36 constructs had zero or severely impaired (>10^6-fold reduction) complementation of pDT2957 and pDT2978-containing cells (Table 3-3). The tandem mutation of both Walker A and B motifs also did not complement transfer, and cumulatively, these results suggest that the NTP-binding motifs of TrhC are required for conjugative DNA transfer. Comparable results were also found during mutational studies of VirB4 from the Ti plasmid (Berger and Christie, 1993) and the related TrbE from IncP plasmid RP4 (Rabel *et al.*, 2003).

Complementation of pDT2969 with Walker A mutation constructs of pMWG36, in comparison with the complementation results with wild-type pMWG36, resulted in greater than 10^4-fold reduction in transfer frequency (Table 3-3). The insertion of mini-Tn*10* into the *trhC* gene of pDT2957 and pDT2978 occur upstream of the Walker A-encoding region, whereas in pDT2969 mini-Tn*10* is inserted downstream of the Walker A sequence, but upstream of the Walker B sequence (Taylor *et al.*, 1999). It is possible that a truncated TrhC peptide containing an unaltered Walker A motif is produced from pDT2969 which can weakly support transfer when full-length, albeit mutated, TrhC is supplied *in trans*. TrhC-like peptides might also be produced from pDT2957 and

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 3-3. Complementation of conjugative transfer and susceptibility to bacteriophage of donor cells containing i) R27 *trhC* transfer mutants, and ii) pMWG36 or derivatives encoding TrhC Walker motif mutations. Underlined residues indicated amino acid substitutions in sub-regions of the Walker A (GKS) and/or Walker B (DEAW) motifs encoded in the pMWG36 derivatives. Units of transfer are transconjugants/donor. A transfer frequency of "0" refers to the absence of colonies on plates selecting for transconjugants, whereas "<10^{-7}" refers the presence of a small amount of transconjugant colonies that lay outside the dilution range of the experiment to accurately define a transfer frequency. The susceptibility to the H-pilus specific bacteriophage Hgal was determined in phage spot assays, and lysis was scored as "+", whereas resistant strains were scored as "-".

R27 / pMWG36	pDT2957		pDT2978		pDT2969	
	transfer	phage	transfer	phage	transfer	phage
Wild-type	8.3×10^{-2}	+	2.3×10^{-1}	+	2.4×10^{-1}	+
AAA	0	-	0	-	9.7×10^{-5}	-
AKS	0	-	0	-	6.0×10^{-5}	-
SKS	0	-	0	-	7.7×10^{-5}	-
GES	0	-	0	-	3.6×10^{-5}	-
del. DEAW	0	-	$< 10^{-7}$	-	0	-
NE	0	-	0	-	0	-
EE	$< 10^{-7}$	-	$< 10^{-7}$	-	1.6×10^{-4}	-
AAA/EE	0	-	0	-	7.2×10^{-7}	-

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

pDT2978, but considering the point of mini-Tn*10* insertion in these plasmids (Taylor *et al.*, 1999), no peptides containing Walker A or B motifs would likely be produced.

It is also possible that recombination between pMWG36 mutants and pDT2969 occurred, generating a self-transferable form of pDT2969 which resulted in the observed transfer events. To rule out this possibility, transconjugants from each pMWG36 mutation construct + pDT2969 complementation experiment were used as donors. No transconjugants were detected in this subsequent mating (data not shown), suggesting that transconjugants from the primary mating received intact pDT2969, and not a recombinant form containing a functional copy of *trhC*.

Complementation of pDT2969 with Walker B mutation constructs of pMWG36 gave variable results (Table 3-3). The deletion of amino acids DEAW or substitution of Asp to Arg did not support transfer, whereas a substitution of Asp to Glu decreased transfer only ~1000-fold. The capability of the Asp-Glu substitution to partially complement transfer may result from the subtle biochemical difference between these two amino acids (the chain length of Asp and Glu R-groups differ by one CH_2 group).

3.3.7 Mutations in the NTP-binding motifs of TrhC prohibit phage uptake/propagation

Another method used to determine the functionality of *trhC* clones encoding NTP-binding motif mutations was the susceptibility of donor cells to a pilus-specific bacteriophage that lyses cells with functional Mpf determinants. Using Hgal, a bacteriophage specific for IncH plasmid-containing cells (Maher *et al.*, 1991), the donor cells harboring pDT2957, pDT2969, or pDT2978 + pMWG36 or pMWG36-derivatives vulnerable to this RNA phage were determined (Table 3-3). For those cells containing wild-type pMWG36 and any of the three R27 transfer mutants, there was significant lysis, whereas cells containing any of the pMWG36 constructs encoding mutations in the

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

trhC Walker motifs and any of the three R27 *trhC* transfer mutants were resistant to Hgal (Table 3-3). Phage susceptibility depends on the Mpf proteins, and the inability of TrhC NTP-binding motif mutants to support this function (or conjugative transfer) indicates that these are critical domains for the overall functions supplied by TrhC.

3.3.8 Production and stability of TrhC and TrhC-GFP

Mutations made in the Walker A and B motifs of TrhC and TrhC-GFP, and the expression of TrhC-GFP in cells containing different complements of transfer proteins could result in altered stability of TrhC and TrhC-GFP. If true, insufficient amounts of TrhC protein may be present to support complementation of conjugative transfer or the formation of fluorescent foci. To determine the level of intact TrhC or TrhC-GFP produced in each strain used during fluorescence, complementation, and bacteriophage susceptibility assays, whole cell lysates were probed using anti-TrhC polyclonal sera (Figure 3-9). In all circumstances, either TrhC or TrhC-GFP was detected in cells carrying R27 transfer mutants and pMWG36 or pMWG191. The predominant species detected in each lysate correlated to whole TrhC or TrhC-GFP, and few degradation products were observed. These data suggest that the reduced transfer frequency of R27 when supported by TrhC Walker A or B mutants resulted from alteration of the NTP-binding domain, rather than decreased stability of the mutants. Additionally, the stability of TrhC-GFP was not drastically influenced by the absence of individual transfer proteins; therefore the deficiency in foci production with twelve R27 transfer mutants is likely due to the necessity of those transfer proteins for protein complex assembly, rather than TrhC-GFP degradation in their absence.

3.3.9 Enzymatic analysis of TrhC

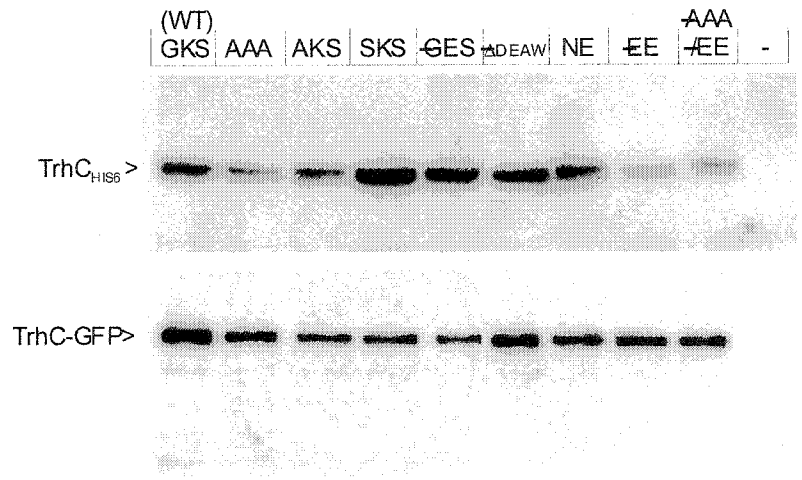
TrhC and TrhC Walker mutants were purified from *E. coli* lysates using a combination of immobilized metal affinity chromatography and anion-exchange fast liquid

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

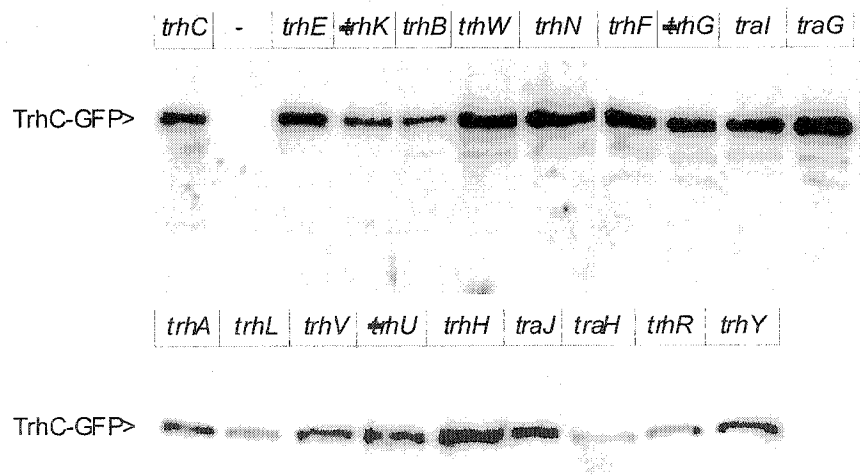
Figure 3-9. Production and stability of TrhC and TrhC-GFP. Polyclonal TrhC antisera was used to detect TrhC or TrhC-GFP in whole lysates of IPTG-induced cultures harbouring (A) wild-type or mutant pMWG36 or pMWG191 + pDT2969, or (B) pMWG191 + R27 transfer mutants. Underlined residues indicated amino acid substitutions in sub-regions of the Walker A (GKS) and/or Walker B (DEAW) motifs. "-" represents the whole cell lysate of RG192 + pDT2969 used as a negative control. Approximately equal amounts of total protein were loaded from each lysate, as the amount of sample used was equalized using optical density measurements of each culture prior to sample preparation, and verified using Coomassie-stained SDS-PAGE (data not shown).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

A. pMWG36 or pMWG191 mutants and pDT2969



B. pMWG191 and drR27 transfer mutants



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

chromatography. No detergents were required to isolate TrhC from the bacterial membrane, as a high concentration of salt was sufficient for isolation. To approximate the amount of ATPase activity of the purified TrhC preparations, aliquots from the protein-containing fractions from the FPLC were incubated with ^{32}P-γ-ATP and excess cold ATP. Inorganic phosphate (the product of an ATPase reaction) was extracted with butanol, and the amount of radioactivity in this organic phase was determined. Purified wild-type TrhC and the subsequent enzymatic analysis is shown in Figure 3-10. The rise in apparent ATPase activity correlates with TrhC-containing fractions, however, the same activity profile was seen with preparations of TrhC containing a Lys to Glu mutation in the Walker A box (data not shown). This implies that the small amount of observed ATPase activity for these fractions is a due to contaminating proteins, rather than the TrhC species because Lys residue has been shown to be essential for function of TrhC and other related proteins (Berger and Christie, 1993; Rabel *et al.*, 2003; Schneider and Hunke, 1998). A contaminating protein of approximately 60 kDa is visible in the seventh fraction shown in Figure 3-10, which also is the fraction having the highest ATPase activity. This could possible represent a metal ion-binding protein that co-purified with TrhC during affinity chromatography using Ni-NTA resin.

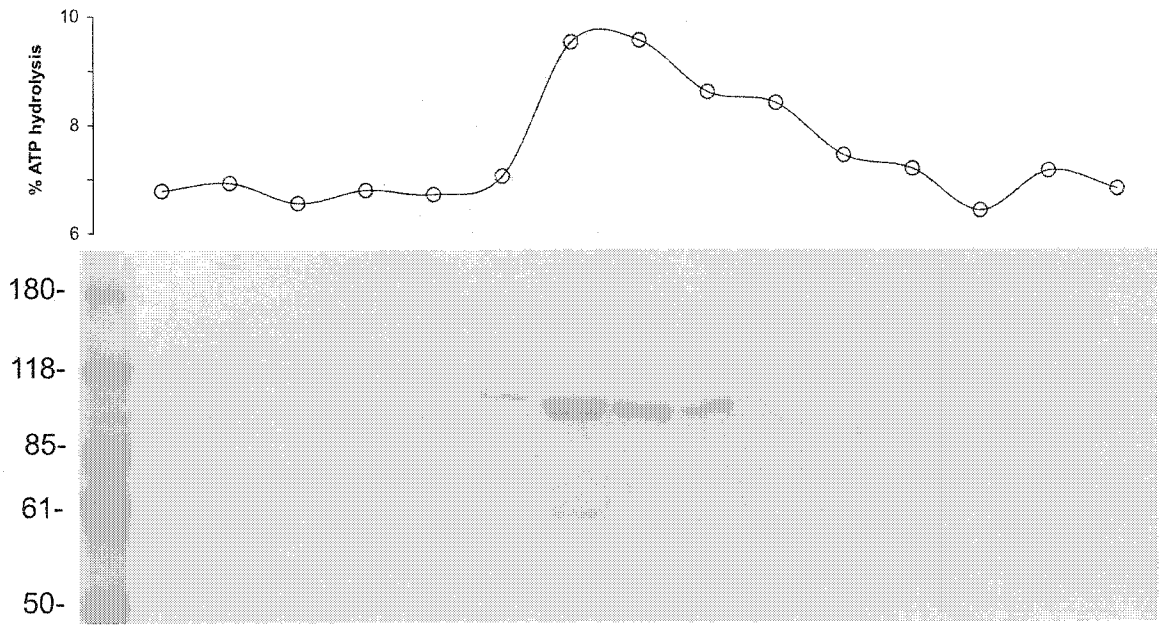
3.4 Discussion

The R27 conjugative transfer system encodes sixteen Mpf proteins and a coupling protein which are presumed to assemble into a transmembrane apparatus (Gilmour *et al.*, 2003; Lawley *et al.*, 2002a; Lawley *et al.*, 2003a). Each of these proteins are required for plasmid transfer and are predicted to localize to the cell envelope; and all Mpf proteins except TrhN and TrhU are required for production of the conjugative pilus structure (Lawley *et al.*, 2003a). In this chapter, membrane-associated protein complexes minimally comprising the Mpf protein TrhC were visualized as discrete foci

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Figure 3-10. Purification and enzymatic analysis of TrhC. Fractions obtained after anion-exchange FPLC were analyzed by SDS-PAGE (lower panel) and an ATPase assay (upper panel). Molecular weight markers are listed on the left. The percent ATP hydrolysis (corrected for background levels of hydrolysis) for each fraction is plotted directly above the respective lane in the coomassie-stained SDS-PAG.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

through a TrhC-GFP fusion. The pDT3048 fusion construct was used to address the issues of transfer protein localization and temperature-sensitive DNA transfer. In addition, the assembly requirements for the fluorescent foci were determined by using a library of R27 transfer mutants, and by creating mutations in the TrhC NTP-binding motifs.

Recent findings that conjugative plasmids reside within discrete subcellular positions (Lawley, 2003; Pogliano *et al.*, 2001) raise the possibility that the proteins responsible for plasmid transfer may also be found at similar positions. However, upon examination of *E. coli* carrying pDT3048, there was no tendency for the fluorescent foci representing TrhC-GFP to arrange at the quarter cell or mid cell positions where the self-transmissible plasmids were previously found. Instead, an average of six GFP foci per bacterium were visualized, primarily at the exterior of the cell, with a modest polar distribution. The slight polar distribution of TrhC-GFP may be attributed to the fact that the older pole (i.e. the pole not derived from most recent septation event) is the most static region of the cell in regards to cell wall synthesis, possibly allowing perpetuation of proteins in this region (Lybarger and Maddock, 2001). Furthermore, the relatively unbiased distribution of TrhC-GFP foci within the bacterial membrane suggests that the TrhC transfer protein complex is unlikely to be anchored or has no mechanism to localize to a specific cellular site and is therefore subject to lateral diffusion within the membrane, unlike the replication machinery (Lemon and Grossman, 2000) and plasmids (Lawley, 2003; Pogliano *et al.*, 2001). Likewise, pili produced by the Mpf complexes were also located randomly about the cell. This is consistent with recent observations that the point of fusion between mating pairs of donor and recipient bacteria was located along the lateral wall or at poles of either cell (Lawley *et al.*, 2002b). When expressed in the absence of Mpf proteins, TrhC was found in the cytoplasm and the bacterial cytoplasmic membrane

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

during cell fractionation experiments. TrhC contains hydrophobic regions that may direct this protein to the bacterial cytoplasmic membrane.

Kumar and colleagues (Kumar *et al.*, 2000) proposed that if a protein were to randomly insert into a membrane then a homogeneous distribution of label would be detected in the membranes, whereas a discrete focus/cluster may represent a specific complex of proteins. In the experimental system used here, GFP foci represent clusters of protein since a lone GFP molecule would be undetectable as a discrete signal. The foci are predicted to represent heterologous complexes of transfer proteins composed partly of TrhC-GFP, rather than homogenous assemblies of TrhC-GFP molecules. The complex nature of a TrhC-GFP focus was revealed by the dependence on twelve other Trh and Tra proteins for assembly. Eleven of the twelve proteins essential for TrhC-GFP foci formation have been classified as Mpf proteins, with the only exception being TraJ. This transfer protein is predicted to have four transmembrane domains by the program TMHMM (Krogh *et al.*, 2001), and is therefore likely to be an inner membrane-associated protein (Figure 2-7). TraJ could possibly interact with other transfer proteins at this cellular location. The protein complex represented as a fluorescent focus may therefore be a subassembly composed of select cell envelope-associated transfer proteins, including both Mpf and non-Mpf proteins.

Distinct protein subcomplexes have also been characterized from the VirB T-DNA transfer apparatus of *A. tumefaciens* that are not composed of the entire complement of VirB proteins, but were proposed to be the core for construction of the remainder of the transfer apparatus (Jakubowski *et al.*, 2003; Krall *et al.*, 2002). These include complexes of VirB6-B8-B9 which can form independently of the other Vir proteins (Jakubowski *et al.*, 2003) and a VirB7-B9-B9-B10 complex that was extracted from the membrane of *A. tumefaciens* (Krall *et al.*, 2002). VirB8 was observed to be essential for assembly of VirB9 and VirB10-associated protein complexes (Kumar *et al.*, 2000). There is no

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

evidence that the twelve transfer proteins required for formation of TrhC-GFP foci participate in direct interactions with TrhC. Instead, these proteins may cumulatively represent a network of interacting proteins, of which TrhC is a member. The assembly of this network of proteins may be ordered, and loss of an initializing component would likely prevent complete assembly, observed in this study as absence of fluorescent foci. Furthermore, none of TraA, TraK, TraG, and TraI, are required for focus formation, but their absence caused a 25-53% reduction in the number of cells containing foci. These proteins may be part of the interaction network but could play an auxiliary role in complex assembly and are not absolutely required.

The precise functional role of TrhC (and TrhC NTP-binding motifs) in the transfer apparatus is undetermined. These results suggest that the function of the transfer system (including DNA transfer and pilus production) is dependent upon the NTP-binding motifs of TrhC, but that the assembly of the apparatus is not. Various mutations affecting the TrhC NTP-binding domains had little effect on the formation of the TrhC-associated subassembly of transfer proteins. It has been suggested that assembly of VirB protein subcomplexes also do not require NTP binding or metabolism by VirB4 (Dang *et al.*, 1999). VirB4 Walker A motif mutants did not diminish the stimulated uptake of plasmid RSF1010 in recipient cells producing a subset of the VirB proteins, including the VirB4 mutants (Dang *et al.*, 1999). Although NTPase activity has not been demonstrated for TrhC or any related proteins *in vitro*, the *in vivo* conditions may support this energy-providing function for translocation of DNA and/or pilin substrates.

Curiously, the transfer of H-plasmids is inhibited at the temperature in which *S. enterica* sv. Typhi infection occurs. Two scenarios have been hypothesized for the thermosensitive mechanism of IncHI plasmid transfer: (i) essential transfer protein(s) are inactivated at the elevated temperature or (ii) a transcriptional or post-transcriptional mechanism results in temperature-dependent synthesis of essential transfer protein(s)

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

(Sherburne *et al.*, 2000). Efforts to isolate an R27 mutant that is temperature-resistant for transfer have been unsuccessful, implying that more than one mechanism is present to thermally regulate plasmid transfer. In this study, evidence of distinct mechanisms underlying this phenomenon were demonstrated. There was differential expression of TrhC-GFP from pDT3048 between 27°C and 37°C, and in contrast to results seen at 27°C, the co-existence of pMWG191 and pDT2969 results in a ubiquitous green label throughout the bacterium at 37°C, not distinct foci. This confirms that R27-encoded components are required for foci formation and are not present at the elevated temperature. This data confirms and extends a previous study examining piliation of pDT1942-containing cells at different temperatures (Maher and Taylor, 1993). Using the transmission electron microscope, no pili were observed on cells grown at 37°C, while cells grown exclusively at 27°C exhibited 1-4 pili. The results of Maher and colleagues indicate that temperature-sensitive DNA transfer is due to a lack of pili at 37°C (Maher and Taylor, 1993). The results presented in this chapter suggest that no pili are produced at 37°C because the Mpf proteins necessary for pilus elaboration and DNA transfer are not present at this temperature. It is likely that multiple Tra2 genes are not expressed at 37°C, since a large proportion of Tra2 proteins are believed to be co-translated from a poly-cistronic mRNA (Rooker *et al.*, 1999). It is possible that a protein involved in activating transfer gene expression is temperature labile, as observed for VirA from the Ti plasmid (Jin *et al.*, 1993).

In addition to the decreased TrhC-GFP production at elevated temperatures, a 2.5 hour lag period was required for the visualization of TrhC-GFP after cultures grown at 37°C were cooled to 27°C. Concomitantly, a 3.5 hour lag period was required to observe H-pili after culturing cells at 37°C and then switching to 27°C (Maher *et al.*, 1993). The long lag period seen in our study might represent the time required to reverse conditions that repress Tra2 transfer gene expression, initiate transcription of

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

transfer genes, translate the resultant mRNA, and then assemble all of the transfer proteins into complexes which can be visualized as discrete fluorescent foci. If TrhC-GFP molecules were visualized immediately after translation (plus the short maturation time of GFPmut2) a much shorter lag time than 2.5 hours would be expected, but it is unlikely that single TrhC-GFP molecules could be detected as discrete foci, thus the foci must represent locations of higher TrhC-GFP density. The long time period required to visualize these foci might primarily consist of the time required to assemble supramolecular complexes. The one hour difference in lag time between the appearance of TrhC-GFP and H-pili is likely due to the period required to assemble the Mpf proteins into a fully functional transfer apparatus and to produce a pilus.

Some cells containing pDT3048 (<1%) still produce GFP foci at 37°C, implying that Tra2 gene repression is not fully comprehensive and a second mechanism to limit plasmid transfer at 37°C must exist. Peter Newnham has shown that plasmid transfer is dramatically and immediately influenced by exposure to 37°C (Newnham, 1995). Since it is unlikely that a transcriptional control mechanism could exert such an effect, it is probable that one or more essential transfer proteins are inactivated or unstable at this temperature. Fullner and Nester (Fullner and Nester, 1996) have shown that while the *vir* genes responsible for T-DNA transfer are expressed maximally at 25°C, the T-DNA transfer apparatus functions with reduced efficiency at temperatures successively greater than 19°C. This can be explained by the observation that the subsistence of VirB10, an essential protein for T-DNA transfer, is temperature-dependent (Banta *et al.*, 1998). These two studies verify that thermosensitivity of T-DNA transfer is a property of the Vir proteins. This model can be extended to R27 where it is possible that one or more of the proteins necessary for H-plasmid transfer is affected by temperature in a manner that prevents conjugation. It is improbable that the pilus is immediately affected

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

by exposure to 37°C, because H-pilus morphology remained unchanged after a one hour pulse at this temperature (Maher *et al.*, 1993).

Both the intracellular position and temperature-dependent assembly profile of protein complexes whose constituents are essential for the process of bacterial conjugation were identified. These complexes may play a role in pilus production, after which recipient bacteria can be brought towards the donor to form intercellular junctions. The complexes present in these junctions may then participate in DNA transmission between bacteria.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Chapter 4

Interaction between the IncHI1 plasmid R27 coupling protein and type IV secretion system

Portions of this chapter have previously been published as:

Gilmour, M.W., Gunton, J.E., Lawley, T.D., and Taylor, D.E. (2003) Interaction between the IncHI1 plasmid R27 coupling protein and type IV secretion system: TraG associates with the coiled-coil mating pair formation protein TrhB. *Mol Microbiol* **49**: 105-116.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

4. Interaction between the IncHI1 plasmid R27 coupling protein and type IV secretion system

4.1 Introduction

During conjugation DNA transfers from one bacterial cell to another; consequently DNA must pass through multiple physical barriers: the cell envelope of both donor and recipient bacteria. This is achieved via the conjugative apparatus, comprised of i) a membrane-associated, multi-protein complex termed the mating pair formation (Mpf) complex responsible for DNA transfer and pilus production, and ii) a cytosolic complex, the relaxosome, which processes plasmid DNA into a single-stranded intermediate for transfer. These two distinct protein complexes are thought to be linked by the inner membrane-associated coupling protein (TraG family), although direct evidence exists only for the relaxosome/coupling protein interaction (Cabezon *et al.*, 1997; Hamilton *et al.*, 2000; Schroder *et al.*, 2002; Szpirer *et al.*, 2000).

The designation of TraG as the R27 coupling protein is based upon several factors, principally the similarity to the model coupling proteins TraD (F), TraG (IncP), TrwB (IncW) and VirD4 (Ti) shown by PSI-BLAST analysis (Lawley *et al.*, 2002a). Additionally, when *traG* is disrupted by transposon insertion, conjugative transfer of R27 ceases entirely, whereas production of H-pili still occurs (Lawley *et al.*, 2002a), implying that TraG is not part of the Mpf complex, but is still required for functional transfer. T-pilus formation on *A. tumefaciens* also occurs in the absence of coupling protein VirD4 (Lai *et al.*, 2000). In Tra1, *traG* is adjacent to *traI* (encoding the R27 relaxase), the characteristic genetic organization of the coupling protein and relaxase in prototypical conjugative systems. The presence of both Walker A and B nucleotide triphosphate binding domains, a hallmark feature of coupling proteins, also correlates with the location of these same motifs in other known coupling proteins (Lessl *et al.*, 1992b; Llosa *et al.*, 1994). These motifs in TraG (IncP) are essential for plasmid RP4 transfer (Balzer

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

et al., 1994), but NTPase activity was not detected for TraD (F), TraG (IncP), HP0524 (*H. pylori*), and TrwB (IncW) (Moncalian *et al.*, 1999; Schroder *et al.*, 2002).

The goal of this study was to identify Mpf protein(s) that interact with the R27 coupling protein TraG. Using both a bacterial two-hybrid analysis and co-purification of protein complexes, an interaction between TraG and TrhB was detected. The ability of TrhB to form a homomultimer was also demonstrated, allowing a model to be proposed which further defines the conjugative pore and mechanism of conjugative plasmid transfer. In this model, TrhB and TraG form a macromolecular structure that spans the inner membrane and periplasm and may connect with an outer membrane protein to allow passage of plasmid DNA from the donor cell. This study will focus on the characterization of the domain of TrhB which interacts with TraG.

4.2 Experimental Procedures

4.2.1 Bacterial strains and growth conditions

E. coli strains used in this study are presented in Table 4-1. Strains were grown with Luria-Bertani media (Lennox formulation). BTH101 was grown at 30°C and DH5α at 37°C, unless otherwise noted. Appropriate antibiotics were added to the following final concentrations: ampicillin (Ap) 100 μg ml^{-1}; kanamycin (Km) 50 μg ml^{-1}; nalidixic acid (Nal) 30 μg ml^{-1}; rifampicin (Rif) 20 μg ml^{-1}; tetracycline (Tc) 10 μg ml^{-1}; and chloramphenicol (Cm) 16 μg ml^{-1}. In the BTH assay, X-gal was added to a final concentration of 40 μg ml^{-1} in agar plates containing Amp and Km, and colonies from each pairwise combination were restreaked then inoculated into 20 ml LB and grown for 16 hours at 28°C for subsequent quantification of β-galactosidase activity in a standard Miller assay (Miller, 1972). Complementation experiments were performed as described (Lawley *et al.*, 2002a; Taylor and Levine, 1980) between donor cells (RG192 + drR27

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 4-1. Plasmids and *E. coli* strains used in this chapter.

Strain or Plasmid	Description	Selective markers[a]	Source
E. coli strain			
BTH101	F⁻ *cya-99 araD139 galE15 galK16 rpsL1 (Str^r) hsdR2 mcrA1 mcrB1*	Str	Karimova *et al.*, 1998
DH5α	λ⁻ ϕ80d*lacZ*ΔM15 Δ(*lacZYA-argF*)*U169 recA endA1 hsdR17*($r_k^- m_k^+$)		Invitrogen
DH10B	F⁻ *mrcA* Δ(*mrr-hsd*RMA-*mrcBC*) ϕ80d*lacZ*ΔM15 Δ*lacX74 deoR recA1 endA1 araD139* Δ(ara,leu) 7697 *galU galK* λ-*rpsL hupG*		Invitrogen
DY330N	W3110 Δ*lacU169 gal490* λ*c1857* Δ(*cro-bioA*); spontaneous Nal^r	Nal	Lawley *et al.*, 2002a
RG192	*ara leu lac*	Rif	Taylor, 1983
Plasmids			
pKT25	p15A origin of replication; encodes $CyaA_{1-224}$	Km	Karimova *et al.*, 1998
pUT18C	ColE1 origin of replication; encodes $CyaA_{225-399}$	Ap	Karimova *et al.*, 1998
pKT25-zip	leucine zipper of GCN1 (BTH positive control)	Km	Karimova *et al.*, 1998
pUT18C-zip	leucine zipper of GCN1 (BTH positive control)	Ap	Karimova *et al.*, 1998
pKT25-traG	2.1kbp *Bam*HI/KpnI-digested PCR product in pKT25	Km	Gilmour *et al.*, 2003
pUT18C-traG	2.1kbp *Bam*HI/*Kpn*I-digested PCR product in pUT18C	Ap	Gilmour *et al.*, 2003
pKT25-trhB	1.4kbp *Bam*HI/*Kpn*I-digested PCR product in pKT25	Km	Gilmour *et al.*, 2003
pUT18C-trhB	1.4kbp *Bam*HI/*Kpn*I-digested PCR product in pUT18C	Ap	Gilmour *et al.*, 2003
pUT18C-$trhB_{1-220}$	0.7kbp *Bam*HI/*Kpn*I-digested PCR product in pUT18C	Ap	Gilmour *et al.*, 2003
pUT18C-$trhB_{1-133}$	0.4kbp *Bam*HI/*Kpn*I-digested PCR product in pUT18C	Ap	Gilmour *et al.*, 2003
pUT18C-$trhB_{204-452}$	0.8kbp *Bam*HI/*Kpn*I-digested PCR product in pUT18C	Ap	Gilmour *et al.*, 2003
pKT25-traH	0.5kbp *Bam*HI/*Kpn*I-digested PCR product in pKT25	Km	Gilmour *et al.*, 2003
pMS119EH	Cloning vector; P_{tac}-*lacI^q*; pMB1 origin of replication	Ap	Strack *et al.*, 1992
pMS119EH-$trhB_{FLAG}$	1.4kbp *Eco*RI/*Bam*HI-digested PCR product in pMS119EH	Ap	Gilmour *et al.*, 2003
pTL42	$TrhB_{His6}$ expression vector (pMS119EH)	Ap	Lawley *et al.*, 2003a
pTL43	TrhB expression vector: 1.4kbp *Eco*RI/*Bam*HI-digested PCR product in pMS119EH	Ap	Gilmour *et al.*, 2003
pDT1942	derepressed transfer mutant of R27	Tc, Km	Maher *et al.*, 1991
pDT2971	pDT1942 with mini::*Tn*10 inserted into *trhB*	Tc, Km, Cm	Rooker *et al.*, 1999
pDT2989	pDT11942 with mini::*Tn*10 inserted into *traG*	Tc, Km, Cm	Lawley *et al.*, 2002a

a. Antibiotic resistances: Ap, ampicillin; Tc, tetracycline; Km, kanamycin; Cm, chloramphenicol; Nal, nalidixic acid; Str, streptomycin

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

transfer mutant + complementation vector) and recipient cells (DY330N). Both donor and recipient strains were grown at 28°C.

4.2.2 DNA manipulations

R27 transfer genes and truncated fragments were cloned into the BTH vectors after PCR amplification with *Taq* polymerase (Invitrogen) and primers listed in Table 4-2 and digestion with restriction enzymes *Bam*HI and *Kpn*I (Invitrogen). After ligation, the products were transformed into DH5α and screened for productive cloning by plasmid isolation and restriction analysis. Cloning into pMS119EH was with *Eco*RI and *Bam*HI (Invitrogen).

4.2.3 Immunoprecipitation of TrhB/B and TrhB/TraG complexes

Following the procedures of Tang *et al.* (Tang *et al.*, 1996) and Ho *et al.* (Ho *et al.*, 2002), 1 ml of overnight culture was used to inoculate 20 ml LB + appropriate antibiotics, grown to mid-log at 30°C, and then induced with 0.4 mM IPTG for one hour. Bacteria were harvested and resuspended in 0.5-1ml of IP buffer (PBS pH 7.4, 0.1% Triton X-100, 100 mM NaCl, 5 mM EDTA and 10% glycerol) + 0.5 mM PMSF; then 10 μg lysozyme was added and the mixture left on ice for 30 minutes. Cells were further lysed by sonication (3 minutes with intermittent 10 second breaks, Fisher model 300 sonicator). Cell debris was cleared by centrifugation at 16 000 rpm for 15 minutes, then 40 μL of ANTI-FLAG M2-agarose affinity gel (Sigma) in a 1:1 slurry with IP buffer was added to the supernatant and rotated at 4°C for 1.5 hours. The affinity gel was then washed three times with 1 ml of IP buffer and once with 1 ml IP buffer + 0.5 mg/ml Angiotensin I (Sigma). Proteins were eluted from the gel by boiling in LSB buffer + DTT and run on a 10% SDS-PAG, transferred to nitrocellulose, and blocked with 10% (w/v) milk, 0.1% Tween 20 in PBS, pH 7.4 (MPBST). Primary antibody was applied [anti-

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 4-2. Primers used to clone *trhB*, *traG* and *traH* (and derivatives) into BTH and expression vectors

Primer	Sequence (5' to 3')[a]
traG (BTH-forward)	TATAGGATCC**G**ACAAAATCAAAAAGAACCAAC
traG (BTH-reverse)	AAGGTACC**CA**ATGCAATTTCCTTAGATATTTA
trhB (BTH-forward)	TATAGGATCC**A**GACATTAAAAAGGCCTGGGAAAATAAAAC
trhB (BTH-reverse)	TATAGGTACC**GC**GCGGCTTGTCAGGTTTGTAG
$trhB_{1-220}$ (BTH-reverse)	TATAGGTACC**AT**GCGAATAGTTCGCGTTGGGGCCTTTTC
$trhB_{1-133}$ (BTH-reverse)	TATAGGTACC**AT**GCCTGCATCGAGATTGCGATTAC
$trhB_{204-452}$ (BTH-forward)	TATAGGATCC**A**GGGCGCGAGGTTGTTGCGGGAAGTG
trhB (EH-forward)	TATAGAATTCAATGGACATTAAAAAGGCCTG
$trhB_{FLAG}$ (EH-reverse)	ATATGGATCC**TTA**CTTGTCATCGTCGTCCTTGTAGTCGCGGCTTGTCAGGTTTGTAG
trhB (EH-reverse)	TATAGGATCCTCAGCGGCTTGTCAGGTTTG
traH (BTH-forward)	TATAGGATCC**G**TCTAAATCAAGCTATTAAAG
traH (BTH-reverse)	TATAGGTACCTCATTGTGCCACCTCTGATTTTG

a. Restriction sites are underlined; bold letters indicated engineered bases to maintain a correct reading frame or an engineered stop codon

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

adenylate cyclase rabbit serum, #L24023, specific for the catalytic domain (D. Ladant)], washed, then secondary antibody (anti-rabbit HRP, Sigma) was applied for one hour. Detection was with SuperSignal West Dura extended substrate (Pierce).

4.2.4 Solubilization/purification of $TrhB_{His6}$ and far-Western analysis

A two litre culture of *E. coli* DH10B containing pTL42 was grown to mid-log at 37°C/200 rpm then induced with 0.4 mM IPTG for one hour at 37°C/200 rpm. Cells were harvested by centrifugation, washed in buffer A (10 mM Tris-HCl pH 7.6 and 5 mM $MgCl_2$), resuspended in 15 ml of buffer A, and then lysed by three passages through a French press (8000 psi). Unlysed cells were removed by two low-speed centrifugations and the membrane fraction collected by ultracentrifugation (Ti70.1 rotor, 35 000 rpm, 4°C, 30 min.) The pellet was washed with buffer A, then resuspended in 10 ml buffer B (50 mM Tris-HCl pH 7.6, 500 mM NaCl, 2% Triton X-100, 10% glycerol) + 10 mM imidazole by gently rotating for 3 hours at 4°C. To clear insoluble material, another low-speed spin was performed (JA25.15 rotor, 8000 rpm, 4°C, 10 min.). The supernatant was applied overnight at 4°C to 2 ml Ni-NTA Superflow resin (Qiagen). Unbound material was removed with three 12 ml washes in buffer B + 10 mM imidazole, and TrhB was eluted with an imidazole gradient of 50-300 mM in buffer B. The eluted protein was confirmed as TrhB by tandem mass spectrometry. TrhB primarily eluted with 200 mM imidazole at ~50 μg/ml.

For far-Western analysis of TrhB/TrhB interactions, protein samples were prepared from cultures containing various expression constructs grown at 28°C/200 rpm until mid-log phase (OD_{600} ~ 0.6), and induced with 0.4 mM IPTG at 28°C/200 rpm for one hour. Cells from 1 ml of culture were harvested by centrifugation and boiled for 5 minutes in 50 μl LSB + DTT, and 15 μL were loaded onto a 10% SDS-PAG. Proteins were visualized by Coomassie stain or transferred to nitrocellulose, blocked for 1 hour with 8% MPBST,

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

and incubated with 1 ml soluble $TrhB_{His6}$ in buffer B + 200 mM imidazole and diluted in 9 ml 5% MPBST overnight at 4°C. Unbound protein was removed by washing 3x10 minutes with 5% MBST, and bound $TrhB_{His6}$ detected by immunoblotting with 100 ng/ml monoclonal Penta-His antibody (Qiagen).

4.2.5 Membrane preparations

500 ml of LB-Lennox with ampicillin was inoculated with 20 ml of an overnight culture of DH10B (pTL42), grown at 28°C/200 rpm until mid-log phase (OD_{600} ~ 0.8), and induced with 0.4 mM IPTG at 28°C/200 rpm for one hour. Cells were harvested by centrifugation (10 min., 6000 rpm, 4°C) and resuspended in 10 ml of ice-cold buffer A with 1 mg of RNase and 100 U of micrococcal nuclease. Cells were lysed by three passages through a French pressure cell at 8000 psi and 1 mg of lysozyme added for 30 minutes on ice. Unlysed cells were removed by centrifugation (2 times; 7 min., 6000 rpm, 4°C), and the supernatant subjected to ultracentrifugation (2 hours, 35000 rpm, Ti70.1 rotor, 4°C) to isolate soluble (cytoplasmic/periplasmic) and insoluble (membrane) fractions. The membrane fraction was resuspended in 55% sucrose (w/w), 10 mM Tris-HCl pH 7.6, 5 mM EDTA for further analysis.

4.2.6 Production of antisera to TrhB and detection of TrhB in cell lysates

A peptide representing the C-terminus of TrhB (CTCqkgatiptnltsr) was synthesized and conjugated to carrier protein KLH (via the above capitalized amino acids) and rabbits were immunized with 200 μg of peptide in complete Freund's adjuvant at week 0 (ProSci Incorporated). Subsequent immunizations at weeks 3, 6 and 10 were with 100 μg of peptide in incomplete Freund's adjuvant. Serum was collected at week 13.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

For detection of TrhB in whole-cell lysates, overnight cultures of RG192 + drR27 and derivatives were grown at 28°C. The optical density of each culture was determined at 600 nm, then the cells were harvested and resuspended in appropriate volumes of PBS (pH 7.4), 0.5% Triton X-100, 10 mM EDTA to equalize cell density (~1 ml). Samples were prepared by boiling in 2x LSB sample buffer with DTT, and 5 μL loaded onto a 10% acrylamide gel. The immunoblot was processed as above, except a 1:2000 dilution of TrhB antisera was used as the primary antibody.

4.2.7 Web-based computer programs

Coils (www.ch.embnet.org/software/COILS_form.html); PSI-BLAST (www.ncbi.nlm.nih.gov/BLAST/); TMHMM (www.cbs.dtu.dk/services/TMHMM-2.0/); and EMBOSS pairwise local alignment (Blosum40 matrix, gap penalty =10, extension penalty =0.5; www.ebi.ac.uk/emboss/align/) were used.

4.3 Results

4.3.1 In vivo detection of TraG and TrhB interactions

A bacterial two-hybrid (BTH) system was employed to characterize the membrane-associated complex of R27 transfer proteins. In this BTH assay, test proteins are fused to one of two adenylate cyclase fragments (AC18 or AC25), and upon interaction of the two investigated proteins within a cyclase-deficient strain of *E. coli*, the cyclase fragments are brought into close proximity leading to the conversion of ATP to cAMP (Karimova *et al.*, 1998). This molecule is a positive activator of catabolic operons, including the *lac* operon, and induction of β-galactosidase enables chromogenic differentiation between interacting and non-interacting clones. This BTH system has been used successfully to study the interactions between other conjugative transfer

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

proteins (Szpirer *et al.*, 2000) and is appropriate for the study of multi-component protein complexes and membrane-associated protein complexes (Karimova *et al.*, 1998).

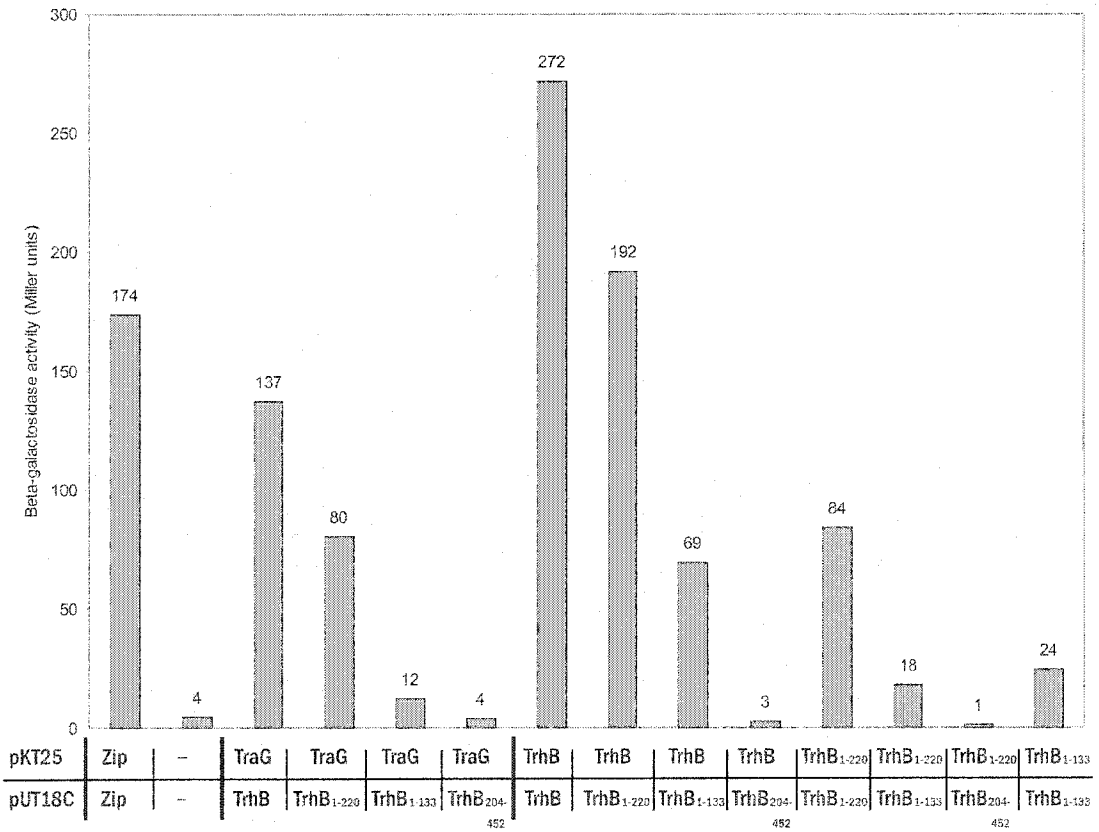
To identify Mpf protein(s) that interact with TraG, the R27 Mpf genes (*trhB, -E, -L, -C, and* -V) were cloned into the bacterial two-hybrid vector pUT18C and co-transformed them with pKT25-*traG* into *E. coli* BTH101 (Table 4-1 and Table 4-2). Each of these fusion clones places a CyaA subunit, AC18 or AC25, on the N-terminus of the transfer protein. Only one productive pairwise combination was observed: $TraG_{AC25}$ with $TrhB_{AC18}$. Within 40 hours after plating, >99% of the BTH101 + pKT25-*traG* + pUT18C-*trhB* colonies turned blue, whereas the BTH positive control (leucine zipper) was also blue, and the negative control (empty BTH vectors) colonies remained white or had faint blue centers (data not shown). Quantification of β-galactosidase activity was performed using a standard Miller assay (Figure 4-1). These data indicate that TraG interacts *in vivo* with TrhB. To my knowledge, this is the first demonstration of a molecular interaction between the coupling protein and an Mpf component.

To determine the domain of TrhB which interacts with TraG, deletion constructs of *trhB* were created (Figure 4-2), cloned into pUT18C, and co-transformed with pKT25-*traG*. Sites for N- and C-terminal truncations of TrhB were selected by comparison with TrhB homologues, and the point of truncation placed in less-conserved regions (discussed below). The three constructs include N-terminal fusions of AC18 to the first 133 aa, the first 220 aa, and the last 249 aa of TrhB ($TrhB_{1-133}$, $TrhB_{1-220}$, and $TrhB_{204-452}$, respectively). After growth on indicator plates, each pairwise TraG/TrhB combination was tested for β-galactosidase activity (Figure 4-1). Only the $TrhB_{1-220}$ peptide fusion construct had a significant level of activity with $TraG_{AC25}$ above the negative control values, whereas $TrhB_{1-133}$ or $TrhB_{204-452}$ co-produced with $TraG_{AC25}$ resulted in low levels

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Figure 4-1. Bacterial two-hybrid analysis of TraG-TrhB and TrhB-TrhB interactions. Full-length and truncated TrhB fusion constructs were co-produced with either TraG or TrhB cyclase fusion constructs in BTH101 and liquid cultures were analyzed for β-galactosidase activity. A representative experiment is shown.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

of cyclase activation. These data indicate that TraG associates with the N-terminal half of TrhB, but not the C-terminus.

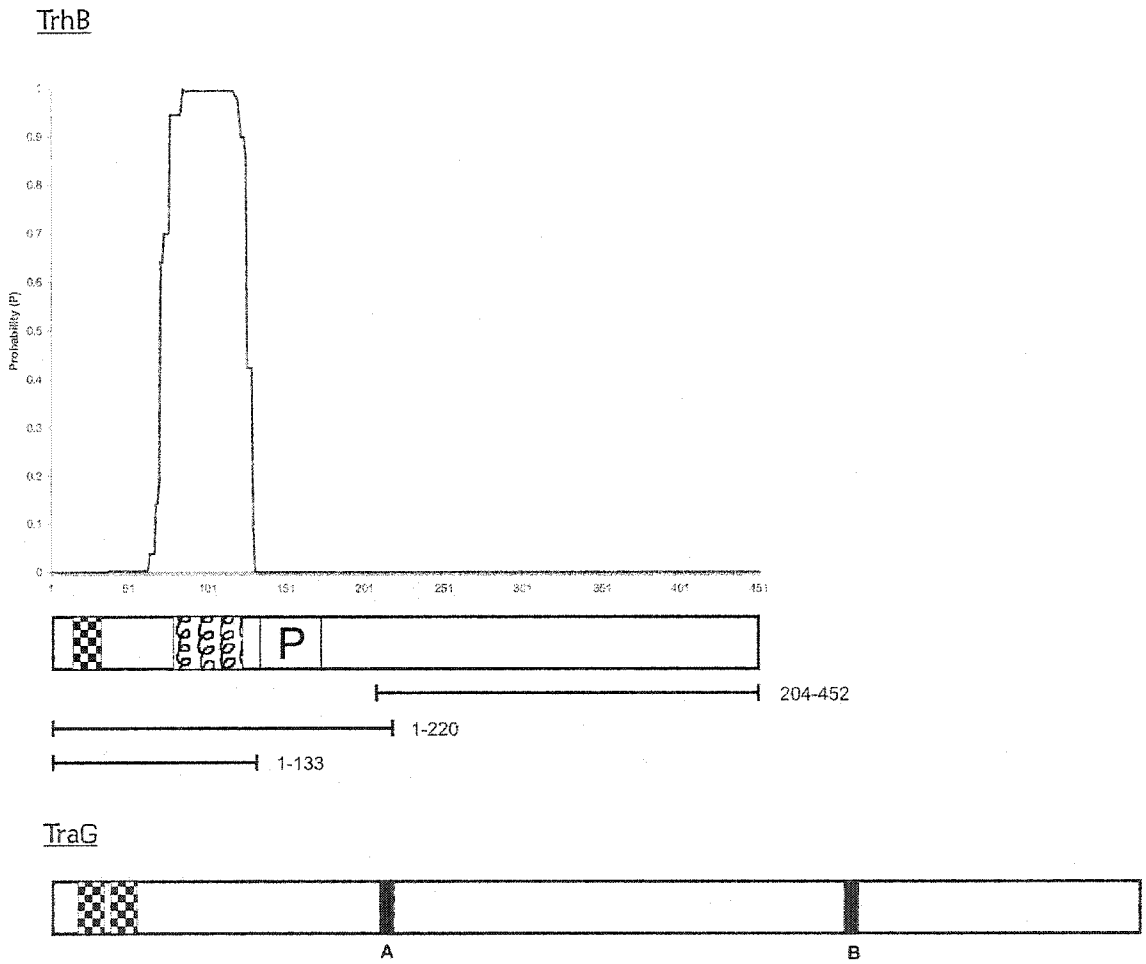
Using the Coils program (Lupas *et al.*, 1991), TrhB was predicted to contain a coiled-coil region from residues 77-124 (Figure 4-2). This domain could aid in the oligomerization of TrhB into a higher order complex, therefore the TrhB and TrhB deletion constructs were tested in the BTH assay (Figure 4-1). A positive interaction was found when full-length TrhB was co-transformed with TrhB, $TrhB_{1-220}$, and $TrhB_{1-133}$, but not with $TrhB_{204-452}$, indicating that the N-terminus of TrhB acts as the homotypic interaction domain. The positive interaction between TrhB/$TrhB_{1-133}$ correlates with the predicted coiled-coil domain. Co-expression of the N-terminal regions $TrhB_{1-220}$/ $TrhB_{1-220}$ also produced a positive BTH signal, however $TrhB_{1-220}$/ $TrhB_{1-133}$ and $TrhB_{1-133}$/ $TrhB_{1-133}$ were much less productive. This may be due to an additive effect elicited by the truncated peptides which, individually, have decreased interaction abilities. Using the two slightly overlapping TrhB N- and C-terminal halves $TrhB_{1-220}$ and $TrhB_{204-452}$, no positive BTH response was detected, and it is therefore unlikely that there is re-association between these segments of TrhB. To confirm that TrhB truncations are stably expressed, and that non-productive BTH responses did not result from a lack of protein production, $TrhB_{204-452}$ was probed for and detected using an antibody specific for the last 13 amino acids of TrhB (Table 4-3). Although this antibody does not recognize N-terminal segments of TrhB, both $TrhB_{1-220}$ and $TrhB_{1-133}$ were both productive in the BTH assay (Figure 4-1 and Table 4-3), implying that they are produced. It cannot be ruled out that decreased levels of protein production resulted in a decreased BTH response.

A functional analysis of transfer protein/adenylate cyclase fusion constructs by using the BTH constructs in a complementation assay of transfer mutants was performed. The BTH vectors pUT18C-*traG* and pUT18C-*trhB* were transformed into RG192 cells

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Figure 4-2. Schematic representation of TrhB and TraG. Predicted transmembrane regions are illustrated as checkered boxes, coiled-coil regions by boxes with spirals, proline-rich domains by a boxed letter P and nucleotide triphosphate-binding domains (Walker A and B) with solid black boxes. Horizontal lines show TrhB truncation constructs used for BTH analysis.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 4-3. Properties of TrhB and TrhB derivatives produced from pUT18C.

trhB-pUT18C clone	Bacterial two-hybrid response[a]		Protein production[b]	Domains present[c]
	traG-pKT25	*trhB*-pKT25		
1-452	+	+	+	CC/Pro
1-220	+	+	N/A	CC/Pro
1-133	-	+	N/A	CC
204-452	-	-	+	-

a. Adenylate cyclase activity in BTH101 containing pKT25 and pUT18C vectors

b. As determined by immunoblot analysis of induced whole cell lysates using anti-TrhB antibody specific for C-terminus (N/A = not applicable)

c. CC = coiled-coil; Pro = proline rich domain (see text)

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

containing the respective R27 transfer mutant and mated with DY330N (Table 4-4). Although full complementation was not achieved with any fusion construct, transfer ability was partially restored with full-length TraG and TrhB constructs, and therefore each was able to support conjugative transfer. TrhB peptides 1-220 or 204-452 were unable to support conjugative transfer, likely due to the loss of required functional domains.

4.3.2 Immunoprecipitation of transfer protein complexes

To confirm the *in vivo* interaction between TraG and TrhB and the homotypic interaction of TrhB, protein complexes were purified from DH5α cells producing epitope-tagged transfer proteins (Figure 4-3). TrhB with a C-terminal FLAG tag ($TrhB_{FLAG}$) and cyclase fusion constructs of either TraG or TrhB was used. The $TrhB_{FLAG}$ construct, which produces TrhB with a C-terminal Asp-Tyr-Lys-Asp-Asp-Asp-Asp-Lys tag, was able to fully complement a *trhB*$^-$ mutant of R27 (Table 4-4). $TrhB_{AC25}$ and $TraG_{AC25}$ were each co-precipitated by $TrhB_{FLAG}$, indicating that both TrhB and TraG interact with TrhB. As a negative control, $TrhB_{FLAG}$ was omitted to ensure that the fusion constructs were not associating with components of the immunoprecipitation reaction (Figure 4-3A). In these reactions, no TrhB or TraG protein was detected with the anti-adenylate cyclase antibody. A protein species of higher molecular weight than $TrhB_{AC25}$ appeared in Coomassie and Sypro Ruby stains of all immunoprecipitation reactions (data not shown), and was detected by the anti-adenylate cyclase antibody in immunoblots (Figure 4-3A, $TrhB_{AC25}$ reactions). The identity of this ~85 kDa species was unknown, however, $TrhB_{AC25}$ was resolved from this species to allow differentiation and identification of $TrhB_{AC25}$ -containing precipitations.

To verify that positive associations were not a result of protein over-expression or non-specific association between the AC25 subunit and $TrhB_{FLAG}$, an

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 4-4. Functional analysis of TrhB and TraG fusion constructs. Complementation experiments were performed with donor cells containing expression vectors and an appropriate R27 transfer mutant.

Complementation Vector	Relative Transfer Frequency[a]
pMS119EH-*trhB*$_{FLAG}$ [b]	4.4
pUT18C-*traG* [c]	6.1×10^{-4}
pUTI8C-*trhB* [b]	5.1×10^{-4}
pUT18C- *trhB*$_{1\text{-}220}$ [b]	0
pUT18C- *trhB*$_{204\text{-}452}$ [b]	0

a: in comparison to transfer frequency of pDT1942 (measured as transconjugants/donor), and where an equal transfer frequency during complementation would be a relative transfer frequency =1

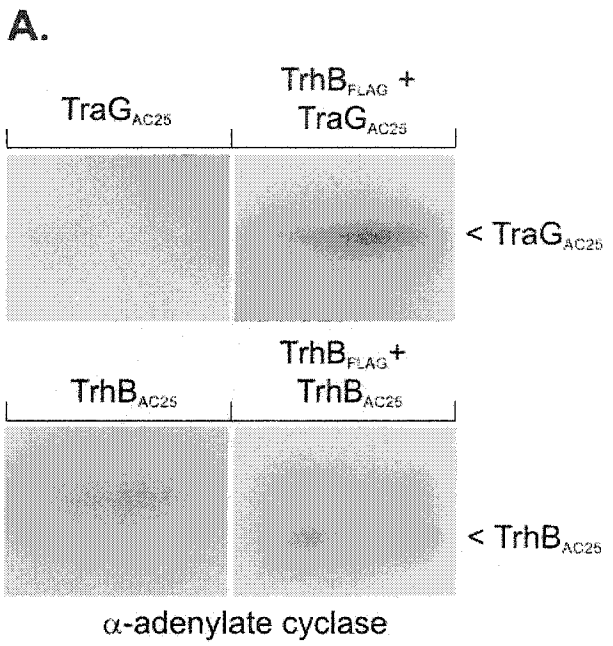
b: complementation of pDT2971

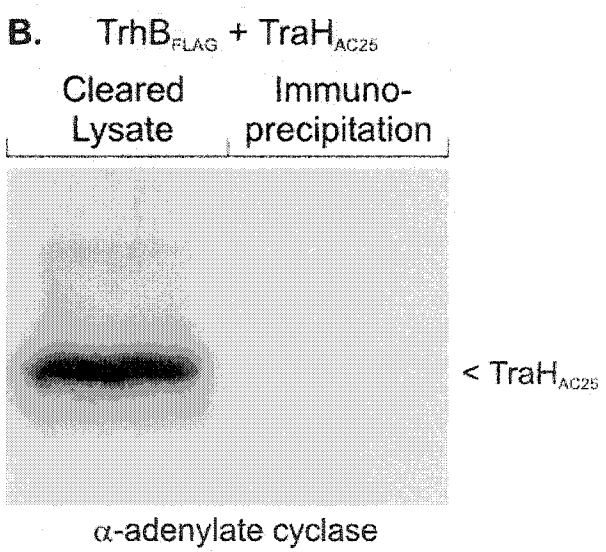
c: complementation of pDT2989

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Figure 4-3. Co-immunoprecipitation of TrhB-TraG and TrhB-TrhB complexes. (A) Cell lysates containing the listed proteins were incubated with M2 anti-FLAG affinity gel, washed, and resolved by SDS-PAGE. After transfer to nitrocellulose, the precipitated proteins were probed with anti-adenylate cyclase antibody. (B) $TraH_{AC25}$ and $TrhB_{FLAG}$ were produced in DH5α, lysed and insoluble material removed by centrifugation. This 'cleared lysate' was incubated with M2 anti-FLAG affinity gel and processed as above (lane marked 'Immunoprecipitation').

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.





Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

immunoprecipitation assay with $TraH_{AC25}$ and $TrhB_{FLAG}$ was performed (Figure 4-3B). TraH is encoded in the Tra1 region of R27 and has been proposed to be a relaxosomal protein (Lawley *et al.*, 2002a). There is no previous evidence of an interaction between this 19.2 kDa transfer protein and TrhB, and these proteins were negative in a BTH assay (data not shown). $TraH_{AC25}$ was stably expressed and was soluble under the same conditions used for the TrhB and TraG assays (Figure 4-3B, cleared lysate), yet was not detected in immunoprecipitated material (Figure 4-3B, immunoprecipitation). These data support our observations of TraG-TrhB and TrhB-TrhB protein interactions.

4.3.3 Far-Western Analysis of TrhB interactions

The TrhB-TrhB protein-protein interaction was further confirmed by far-Western analysis. Cell lysates containing native TrhB (predicted molecular weight of 48 kDa) produced from pMS119EH, or $TrhB_{AC18}$ (predicted molecular weight of 66 kDa) produced from pUT18C, and lysates of cells with pMS119EH or pUT18C were run on a SDS-PAG, transferred to nitrocellulose, and probed with soluble $TrhB_{His6}$. After washing unbound $TrhB_{His6}$, the remaining $TrhB_{His6}$ was detected by Western analysis (Figure 4-4A). It was found that $TrhB_{His6}$ interacts with proteins with molecular weights corresponding to $TrhB_{AC18}$ and TrhB. Each of these proteins were minor species within the probed cell lysate (Figure 4-4B), indicating that $TrhB_{His6}$ interaction was not resulting from non-specific interactions to a predominant species. TraG/TrhB interactions could not be detected using this method, possibly because due to the denaturing conditions during gel electrophoresis.

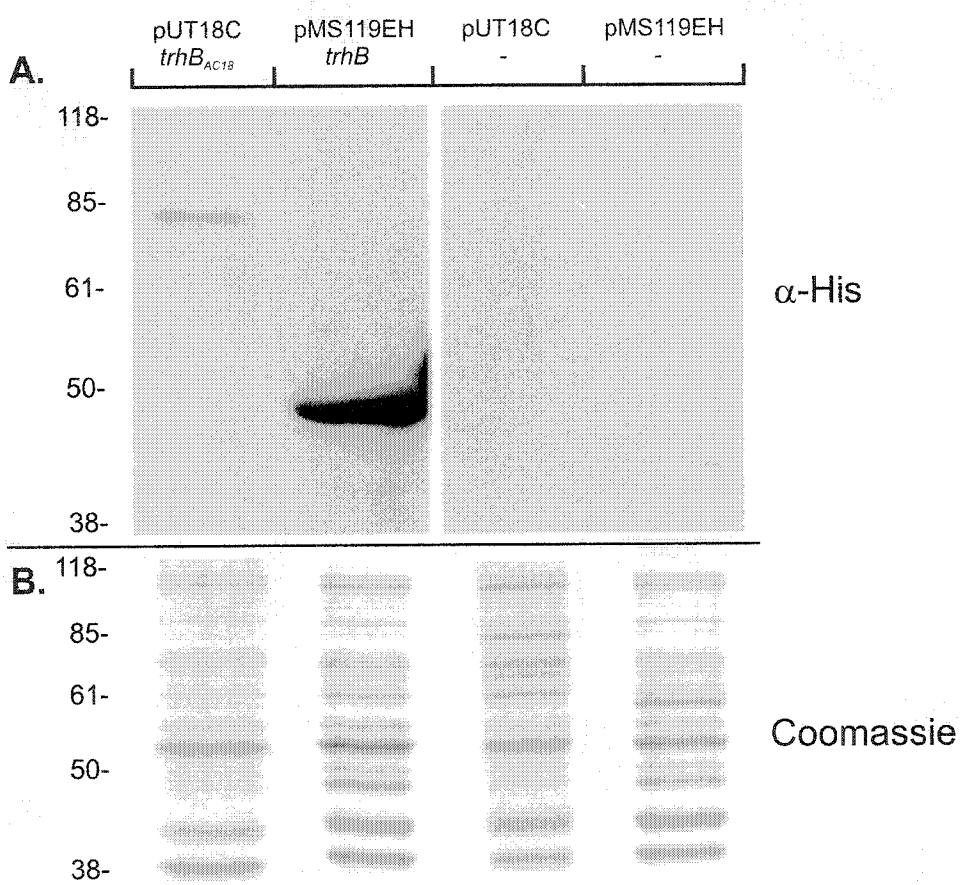
4.3.4 Stability of TrhB in a *traG* mutant

It is common that the stability of a protein is dependent upon the presence of the interacting partner(s). TrhB was probed for in lysates of *E. coli* cells containing either

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Figure 4-4. Far-Western detection of TrhB-TrhB interactions. Cell lysates containing various expression constructs were resolved by 10% SDS-PAGE then (A) immobilized on nitrocellulose, probed with soluble $TrhB_{His6}$ and interactions detected by immunoblotting with Penta-His antibody, or (B) stained with Coomassie blue R-250. Approximate molecular weights (kDa) are listed.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

derepressed R27 (drR27), a *trhB* transposon mutant of drR27 (with and without the $TrhB_{FLAG}$ expression plasmid) or a *traG* transposon mutant of drR27 using polyclonal antiserum raised against a peptide representing the last 13 amino acids of TrhB. After immunoblotting, TrhB was detected in the lysates from drR27 and *traG* R27 mutant-containing cells, but not for the *trhB* R27 mutant (Figure 4-5). Therefore, the presence of TraG does not influence the production/stability of TrhB. This result is supported by an HgaI phage binding assay, wherein the above *traG* R27 mutant was still capable of producing H-pili (Lawley *et al.*, 2002a). Hence, the Mpf apparatus, presumably including TrhB, is still functional in the absence of TraG. $TrhB_{FLAG}$ production was also identified, and there was a detectable difference between in the apparent molecular weight of this epitope-tagged construct and native TrhB (Figure 4-5).

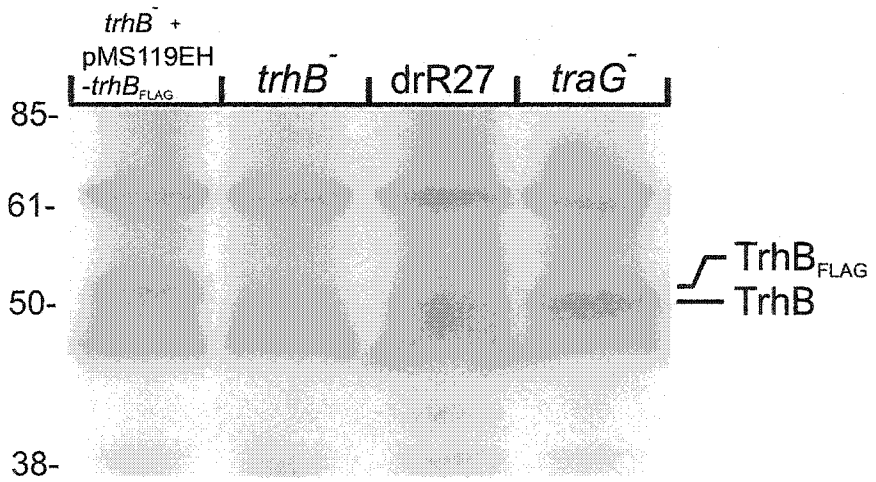
4.3.5 Membrane localization of TrhB

Computational analysis of the TrhB primary amino acid sequence with TMHMM predicted a transmembrane (TM) region at residues 12-34 (Figure 4-6). Subcellular fractionation studies were performed to confirm the predicted membrane localization of TrhB. After cell lysis, ultracentrifugation was used to isolate soluble (cytoplasmic and periplasmic) and insoluble (membrane) fractions. TrhB was found exclusively in the particulate fraction (data not shown), suggesting membrane localization. Attempts to determine the localization of TrhB to a specific membrane fraction using flotation, isopycnic and continuous sucrose gradients or buffered sucrose cushions were unsuccessful (data not shown). To recover TrhB from soluble fractions of cell lysates during $TrhB_{His6}$ purification and $TrhB_{FLAG}$ immunoprecipitation reactions detergent was required at all steps (Figures 4-3 and 4-4), further implying that TrhB is membrane-associated.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Figure 4-5. Production and stability of TrhB. Immunoblot of whole-cell lysates of RG192 containing drR27, or drR27 transfer mutants of *traG* or *trhB*, or *trhB* drR27 transfer mutant with pMS119EH-$trhB_{FLAG}$ (no IPTG induction) probed for TrhB content with TrhB-specific antisera. Approximate molecular weights (kDa) are listed on the left and approximate location of TrhC and $TrhB_{FLAG}$ on the right.

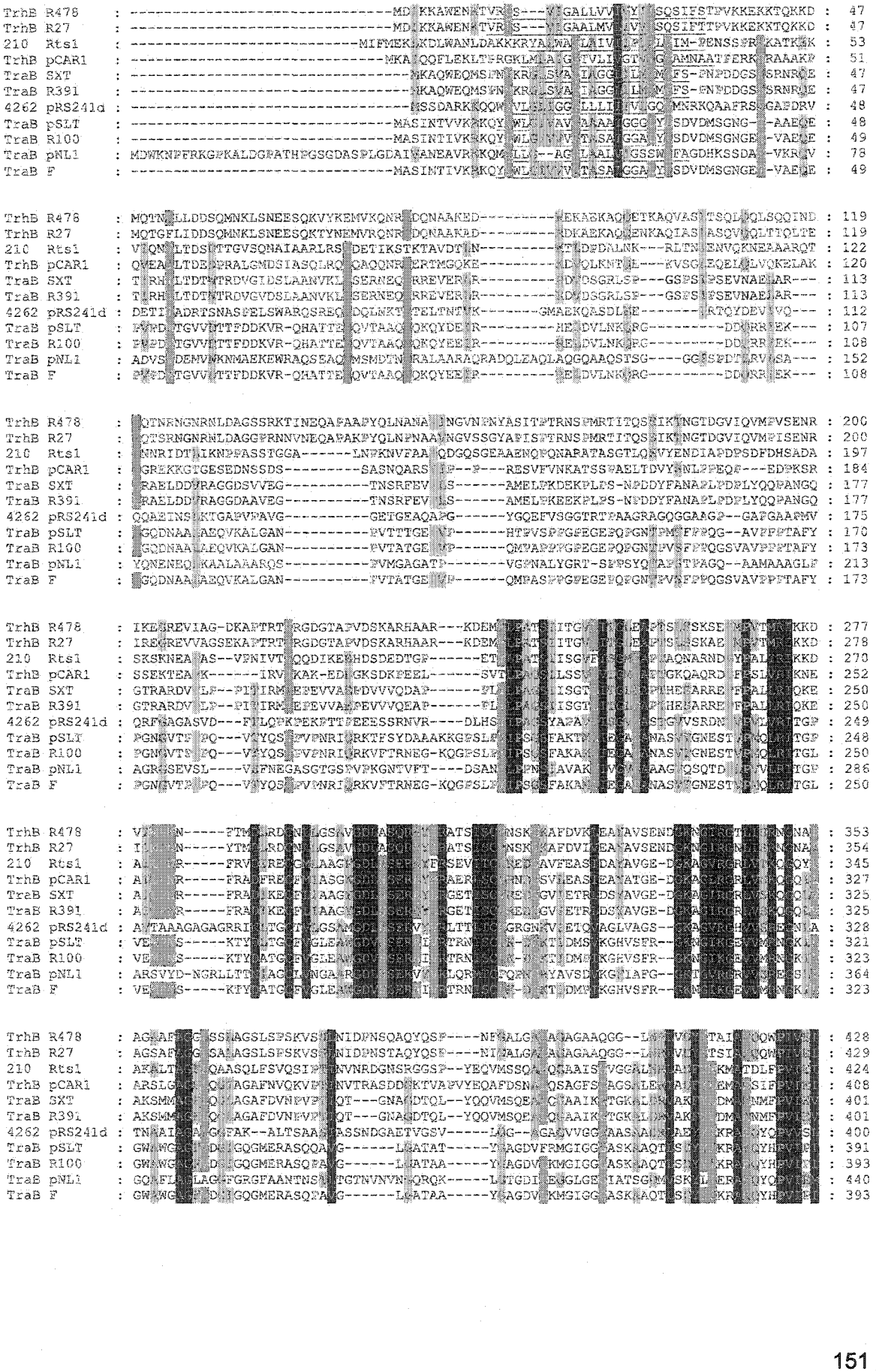
Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.



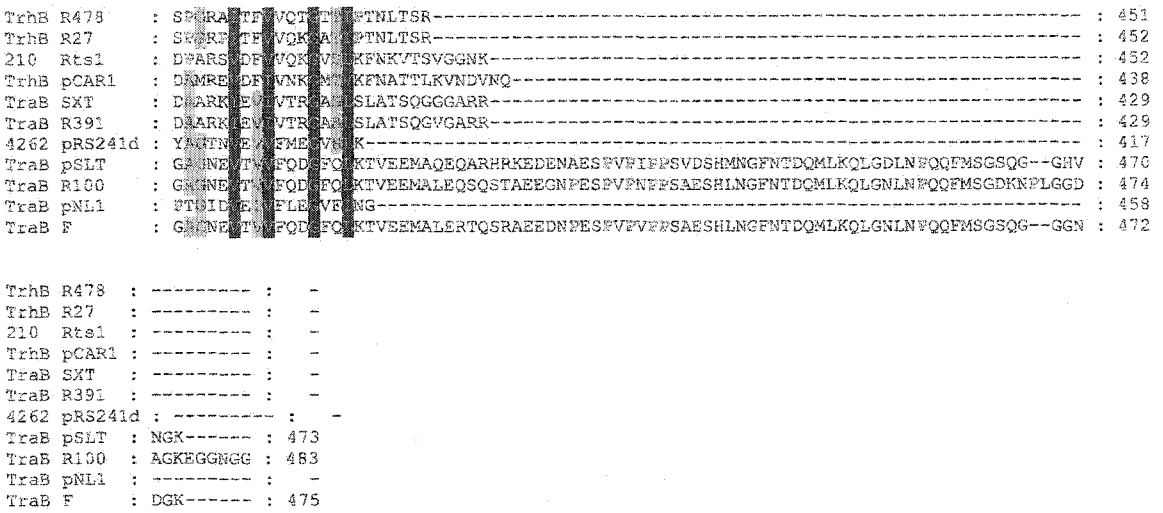
Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Figure 4-6. Multiple sequence alignment and primary/secondary sequence analysis of TrhB and homologues. The alignment was generated using ClustalW in MegAlign (DNAStar) and conserved residues were shaded using GENEDOC. Underlined residues indicate transmembrane helices predicted by TMHMM, red coloured residues indicated coiled-coil domains predicted by Coils, and all proline residues are coloured green. Accession numbers: R478 (CAE51539), R27 (NP_058242), Rts1 (NP_640170), pCAR1 (NP_758672), SXT (AAL59682), R391 (AAM08009), pRS241d (ZP_00008313), pSLT (NP_490567), R100 (NP_052951), pNL1 (NP_049165), F (AAC44179).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

4.4 Discussion

The movement of DNA between bacterial cells necessitates the association of cytoplasmic and membrane-associated proteins. The transmembrane Mpf complex (including 16 proteins in the R27 system) is responsible for production of the conjugative pilus, a structure proposed to contact recipient cells, and formation of the conjugative junction, through which DNA passes. The coupling protein is responsible for the membrane targeting of the relaxosome nucleo-protein complex. This study demonstrates that the coupling protein TraG interacts with the Mpf protein TrhB. The formation of TrhB homomultimers was also observed, although the exact oligomerization state was undefined. The TraG-TrhB protein interaction suggests a mechanism by which the cytosolic DNA molecule is targeted to the transmembrane apparatus for export to recipient bacteria.

TrhB and TraG homologues are found in all known type IV secretion systems, including the conjugative systems (Christie, 2001; Lawley *et al.*, 2003a). It is not surprising that conserved elements are needed to mediate the Mpf complex/coupling protein interaction. TrhB has previously been identified as a mating pair formation (Mpf) protein, since transposon mutagenesis of *trhB* renders R27 transfer deficient and resistant to Hgal, an H-pili specific bacteriophage (Lawley *et al.*, 2003a). This mutant has been complemented for both transfer and pilus production by producing $TrhB_{His6}$ *in trans* (Lawley *et al.*, 2003a). The observed TrhB homotypic interaction could be mediated through the predicted coiled-coil region spanning residues 77-124 (Figure 4-2), as a construct representing the first 133 amino acids of TrhB interacted with full-length TrhB. Coiled-coils are assemblies of α-helices, and this structure has been found extensively in secreted components of T3S systems (Delahay and Frankel, 2002). A coiled-coil structure (Figure 4-6) is also predicted to occur in numerous TrhB homologues, including TraB (F, R391, pSLT, R100, SXT, pNL1) and TrhB (R478 and

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

pCAR1). Although computational predictions for coiled-coil domains are not always accurate, the presence of this feature in a large number of proteins from a single family is indicative of a correct prediction (Lupas, 1996). Furthermore, this region of TrhB is completely free of proline residues, which prohibit the formation of helical structures (Muller *et al.*, 2000).

The membrane topology of TraG and TrhB are unknown. However, it is predicted that TraG is polytopic, with two closely spaced transmembrane (TM) domains at the N-terminus and a short periplasmic loop. This is based upon sequence analysis of TraG and the known membrane topology of other coupling proteins. TMHMM predicted two TM domains at residues 19-41 and 45-64 of TraG (Figures 2-6 and 4-2), likely resulting in cytoplasmic localization for both the short N-terminal segment and the large C-terminal domain containing nucleotide triphosphate-binding motifs. TraG (RP4), TraD (F), and VirD4 (Ti) all have two TM domains at the N terminus separated by a periplasmic domain (Das and Xie, 1998; Lee *et al.*, 1999; Schroder *et al.*, 2002). TrhB is a member of the F-like TraB family of proteins, which are each predicted to contain an N-terminal transmembrane domain and high concentration of proline residues between a coiled-coil domain and the highly conserved C-terminal domain (Figure 4-6). TrhB is predicted to be anchored in the inner membrane by a singular N-terminal TM domain (residues 12-34), and is likely bitopic, with a large C-terminal portion extending into the periplasm (Figure 4-6). The TrhB homologue TraB from the F plasmid (27.2% identity over alignment length = 482, see section 4.2.7) has been predicted to have a large periplasmic domain following the TM domain at residues 13-33 (Frost *et al.*, 1994). TraB was found to be an integral membrane protein, associating with the outer membrane in *tra*$^+$ cells and to the inner membrane in *traV* and *traK* mutant cells (Harris *et al.*, 2001). The VirB10 protein required for T-DNA transfer from *A. tumefaciens* is also similar to TrhB (20.4% identity over alignment length = 460) and was found to be an integral inner

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

membrane protein with the bulk of the protein existing in the periplasm, as revealed by *phoA* fusions (Ward *et al.*, 1990). The co-localization of TraG and TrhB in the inner membrane would in part facilitate an interaction between these two transfer proteins. Neither TraG nor TrhB are predicted to have a cleavable N-terminal signal sequence using the parameters outline by von Heijne (von Heijne, 1986).

Both TrhB and TraB (F) have proline-rich regions (Figure 4-6), indicative of extended structures and protein-protein interaction domains. There is only one proline residue present in the first 134 amino acids (0.7%) of TrhB, whereas the remaining 318 residues include 22 prolines (7%). TraB (F) has a 7% total composition of proline and is comprised of 35% proline from amino acids 135-183 (Anthony *et al.*, 1996; Harris *et al.*, 2001). Similarly, TrhB has a 39 residue segment from 135-173 that is 17.9% proline. Considering that $TrhB_{1-220}$, and not $TrhB_{1-133}$, interacted with TraG, this region may be involved in TraG interaction. Related proteins in the IncP (TrbI, 20.3% identity over alignment length = 518), Ti (VirB10) and IncW (TrwE, 19.2% identity over alignment length = 536) systems all lack predicted coiled-coil domains, yet are also enriched for proline. TrbI is 6.7% proline with a domain from amino acids 88-129 that is 26% proline, VirB10 from pTiC58 is 7.8% proline and 28% proline from 61-114 and TrwE is also 7.8% proline with a domain from 80-137 comprised of 31% proline. Das and Xie found that a LexA fusion of a peptide representing amino acids 47-277 of VirB10 pTiC58 was positive in a yeast two-hybrid analysis with full-length VirB10, whereas a LexA fusion of amino acids 167-377 was negative (Das and Xie, 2000). The inferred homotypic interaction domain of VirB10 also coincides with the proline-rich segment of this protein. TrhB homotypic interactions may solely be due to the coiled-coil segment, however, contributions from the proceeding proline-rich domain may also facilitate this interaction, as indicated by the TrhB homologues lacking the coiled-coil structure. Interaction between TraG and TrhB may also be determined through the TrhB proline-rich domain,

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

since the N-terminal half of TrhB was capable of supporting this interaction. Furthermore, the conserved proline-rich domain present in TrhB-related transfer proteins (TrbI, TrwE and VirB10) may enable binding to the respective coupling protein.

Yeast two-hybrid and fractionation studies of the F plasmid Mpf proteins TraB, TraK and TraV revealed that these proteins form a transmembrane complex extending from the inner membrane to outer membrane (Harris *et al.*, 2001). The R27 Mpf protein TrhK has sequence similarity to TraK, and likewise, R27-encoded TrhV is similar to TraV, the TraK-associated lipoprotein (Harris and Silverman, 2002; Lawley *et al.*, 2003a). Considering the sequence conservation and functional similarities of TrhB, TrhK and TrhV to the F plasmid counterparts, and the similar domain structure of TrhB and TraB, it is plausible that TrhB, TrhK and TrhV enter into a similar transmembrane complex. In this complex TraB may span the periplasm to associate with TraK in a manner similar to the energy-transducing TonB, which has a proline-rich domain that allows extension across the Gram-negative periplasmic space for interaction with outer membrane proteins involved in active transport (Chang *et al.*, 2001; Skare *et al.*, 1993). It was previously found that both TraK and TrhK sequences are similar to members of the outer membrane secretin protein family (Lawley *et al.*, 2003a), and there are preliminary data showing that TraK forms high molecular weight, SDS-resistant complexes (unpublished data). The B/K/V Mpf complex may represent the core transenvelope complex for T4S of H and F plasmids.

Secretins are secretory proteins which form outer membrane-associated ring structures. The secretin subunit PilQ forms a dodecameric ring in the outer membrane *Neisseria meningitidis*, and is required for the secretion of extracellular type IV pilus structures in *N. meningitidis, N. gonorrhoeae,* and *Pseudomonas aeruginosa* (Collins *et al.*, 2001; Drake and Koomey, 1995; Martin *et al.*, 1993). Additional examples of secretins are present in type II (*Klebsiella* PulD) and T3S systems (*Shigella* MxiD,

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Salmonella InvG, *Yersinia* YscC and *Erwinia* HrcC) (Guilvout *et al.*, 1999; Kaniga *et al.*, 1994; Kim *et al.*, 1997; Koster *et al.*, 1997; Schuch and Maurelli, 2001). Notably, a tripartite transmembrane interaction between the inner membrane MxiJ, outer membrane secretin MxiD and MxiM lipoprotein was also observed in the *Shigella* T3S apparatus (Schuch and Maurelli, 2001). This may represent an analogous structure to a transenvelope B/K/V Mpf complex.

Although TraG interacts with TrhB, the coupling protein is not a required component of the transmembrane protein complex for pilus assembly since a R27 *traG* mutant-containing cell produced H-pili (Lawley *et al.*, 2002a). Alternatively, TrhB, TrhK and TrhV were each essential for pilus biogenesis (Lawley *et al.*, 2003a). The data presented in this chapter indicate that in the absence of TraG, TrhB is stably maintained, and thus likely still present in a B/K/V Mpf complex. However, when TraG is present, the B/K/V Mpf complex may be adapted for export of DNA; hence there may be two populations of transport apparatus, those with or without the coupling protein, and each capable of a different function. This may also be true in the T-DNA transfer system, as *virD4* mutants of pTiC58 still produce T-pili (Lai *et al.*, 2000).

These observations allow a model for the macromolecular assembly of transfer proteins into a transmembrane pore for DNA translocation to be proposed (Figure 4-7). Besides interacting with the TraG coupling protein, TrhB may interact with an outer membrane-associated secretin, predicted to be TrhK. Considering the structural data for the IncW coupling protein TrwB (Gomis-Ruth *et al.*, 2001) and the known arrangement of secretins, it is possible that TraG and TrhK each form ring structures in the inner and outer membranes, respectively. If so, TrhB molecules may form a central ring structure spanning the periplasmic space, as TrhB was shown here to minimally form a homodimer and contains coiled-coil and proline-rich features which likely result in an extended conformation. This predicted superstructure would correlate to the stacked

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

ring superstructures having 'cylindrical symmetry' observed for both type III secretion and flagellar motor transmembrane structures (Blocker *et al.*, 2001; Kubori *et al.*, 1998; Thomas *et al.*, 2001).

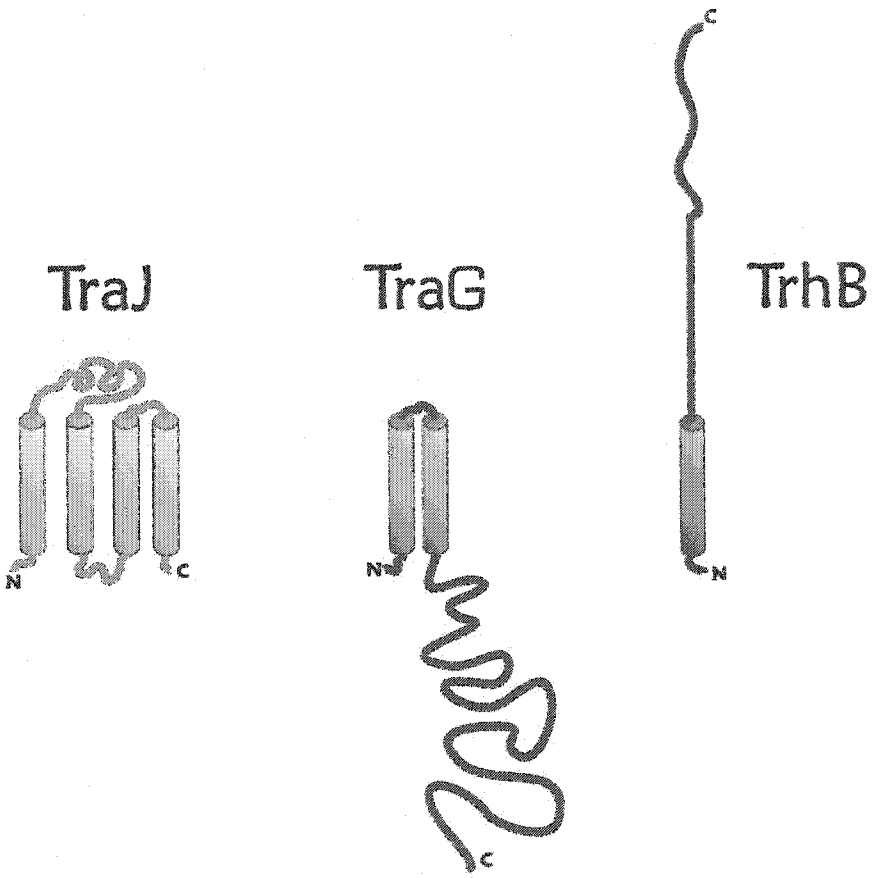
The stacked ring model for the conjugative apparatus necessitates that TrhB is able to span both the inner membrane and periplasmic space (through transmembrane and rigid proline-rich domains, respectively) to interact with the outer membrane secretin TrhK. Deletion constructs of TrhB used in conjugation with TraG in the BTH assay, however, indicated that the proline-rich domain is required for TrhB-TraG interactions (Figure 4-1). The predicted topology for TraG indicates that only a loop of four amino acid residues is present in the periplasmic space, providing a small substrate for interaction with TrhB proline-rich domain. It is more plausible that that TrhB and TraG interact through their inner membrane-spanning domains (as illustrated in Figure 4-7), because these are the only significant segments between these proteins which co-localize to the same subcellular compartment.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

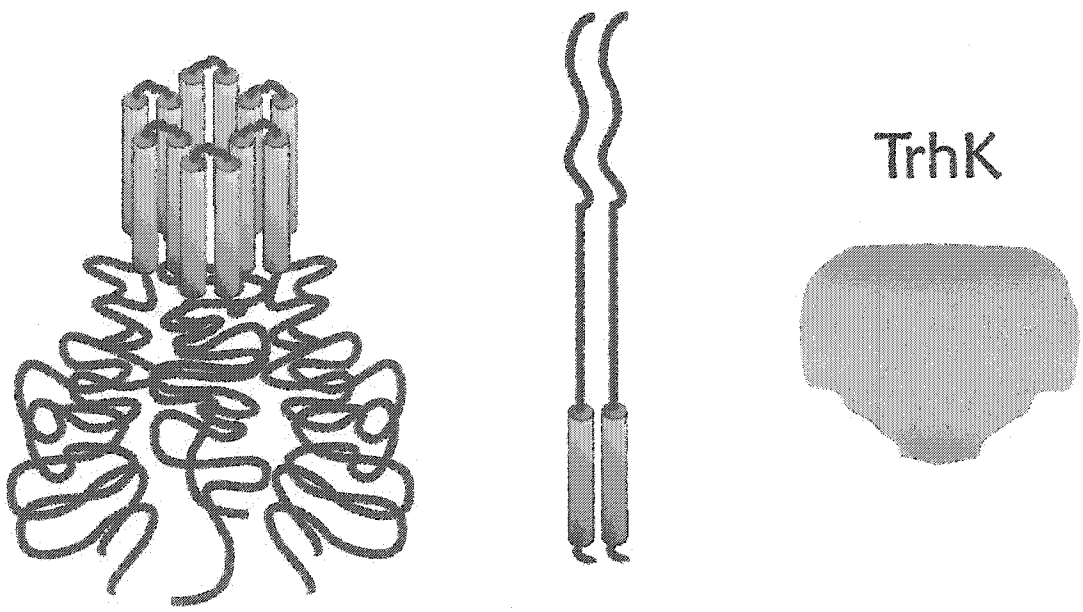
Figure 4-7. Model for the assembly of a pore-like transfer apparatus that traverses the cell envelope. The predicted transmembrane domains of TraJ, TraG, and TrhB monomers are represented as cylinders, with N- and C-termini specified (A). The rigid conformation illustrated for TrhB is predicted from the high proline content in the primary sequence. The predicted multimeric state of TraG and the minimal multimeric state of TrhB are shown (B). TrhK, predicted to be the H-type secretin molecule (Lawley *et al.*, 2003a) is represented as multimeric ring structure, modeled after the three dimensional structure of *Neisseria meningitidis* dodecameric secretin PilQ (Collins *et al.*, 2003). A possible assembly of these proteins into a cell envelope-spanning superstructure is presented (C). Specific interactions are proposed or known to occur between the relaxase-TraG, TraG-TrhB, TrhB-TrhK (see text). The role of TraJ is unknown, but it is a conserved member of the H-type transfer apparatus.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

A.

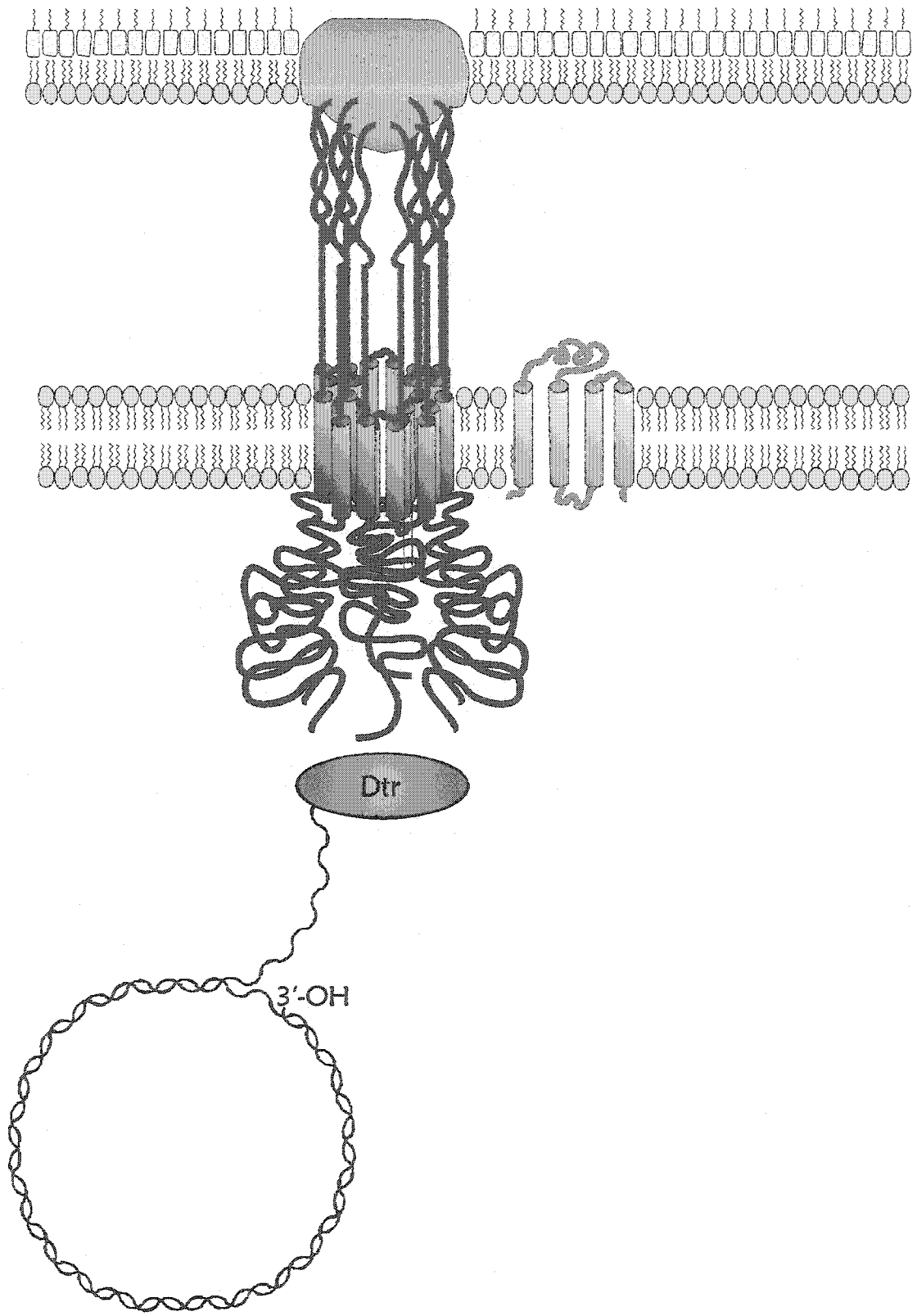


B.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

C.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Chapter 5

General Discussion

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

5. General Discussion

5.1 Genome evolution

Horizontal gene transfer has been suggested to "uproot" the tree of life because of the trans-domain transfer of genetic determinants, a process which may also challenge aspects of the Darwinian theory of evolution (Doolittle, 2000). Rather than being three independent branches, the three domains of life are now realized to resemble an inter-connected web because of transfer events between different genomes, demonstrating that genetic inheritance is not confined to species-specific lineages. As genomes evolve through the mechanistic process of horizontal gene transfer, dramatic phenotypic changes can occur after a single transfer event, however the principles of 'survival of the fittest' still apply. Not all transfer events will be beneficial to the recipient, and selection of transferred genes no doubt still occurs. Genes which are unlikely to be transferred and successfully maintained include the core set of genes required for basic cellular processes (RNA, tRNA, chromosome replication, transcription, and translation proteins), as recipients would already encode such features, and the acquired determinants might be unable to interact with the native complement of core factors (Jain *et al.*, 1999; Lawrence and Hendrickson, 2003). In contrast, those determinants offering novel phenotypic traits (that are able to function alongside resident features) might be selected for horizontal transfer (Lawrence and Roth, 1996). This process is exemplified by the transfer of resistance determinants.

Conjugative resistance plasmids, in addition to being a vector for horizontal gene transfer, also represent distinct genetic units which are subject to processes of acquisition, mutation, and deletion. Through the comparative sequence analysis of IncHI plasmids R27, pHCM1, and R478, a conserved set of determinants for plasmid replication, partitioning, and transfer were revealed. These features are likely conserved

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

because each represent core determinants necessary for vertical and horizontal transmission. In contrast, coding sequences which result in resistance phenotypes are not strictly required for the maintenance of plasmids in the bacterial gene pool. The resistance genes/operons are therefore classified as accessory determinants, and represent exchangeable modules which enhance the phenotypic advantages imparted to the host bacteria.

The conjugative transfer determinants of the IncHI plasmids (or an ancestral sequence) were acquired from different sources. A sequence comparison with numerous other genetic elements demonstrated that the Mpf genes are of a different lineage than the Dtr/coupling determinants. Despite this bifurcated history, an interaction between these different classes of proteins is necessary for the assembly and function of conjugative apparatus. This was revealed by a biochemical characterization of protein-protein interactions and a fluorescent visualization of the conjugative protein apparatus. The ancestral sequences of the H plasmid transfer system were possibly independent functional modules that separately supported pilus production (Mpf) and DNA metabolism (Dtr), but which have converged and co-evolved to establish a singular functional module for conjugative transfer. The adaptation of two ancestrally distinct modules is not only evident in the genetic sequences and organization of H plasmids, but also at the protein interface between these two modules.

5.2 The multifunctional Mpf proteins

Classification of conjugative transfer proteins as Mpf determinants is partly based upon the inability of mutants to support both plasmid transfer and pilus production. Conversely, those mutations which affect only transfer but not assembly of pili are classified as Dtr or coupling determinants. The logic behind this classification scheme is

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

that the Mpf proteins can assemble and function in the absence of the Dtr/coupling determinants. This has been demonstrated through not only the detection of pili on Dtr/coupling mutants (Lawley *et al.*, 2002a), but also by the continued production and stability of Mpf protein TrhB (Figure 4-5), and the visualization of TrhC-associated protein subcomplexes (Figure 3-6) in the absence of relaxase or coupling proteins. The ancestral function of the Mpf module may have been solely pilus production, but was subsequently adapted for conjugative transfer by addition of the Dtr and coupling determinants.

The phenotypic advantages of a bacterial cell producing pili using the Mpf module include the ability to enter into biofilms. Ghigo has demonstrated that cells expressing one of a number of different types of Mpf systems (including those from IncF, I, H2, X, T, N, and W groups) resulted in the formation of biofilms, a process likely initiated by the respective conjugative pili (Ghigo, 2001). This correlates with the observed temperature sensitivity of the H plasmid Mpf system, as pili are produced at lower temperatures, possibly promoting the formation of biofilms under environmental conditions. Therefore an ancestral form of the Mpf module may have existed for the export and assembly of extracellular pili, which were responsible for adhesion to environmental substrate and intimate cell-cell association resulting in biofilm formation. The ability of the ancestral Mpf module to export substrate (i.e. pilin subunits) across the cell envelope and to make contacts with neighbouring cells (via the pilus) may have then been 'exploited' by the Dtr and coupling determinants for conjugative transfer of cytoplasmic DNA molecules.

5.3 H-type Dtr/coupling determinants

The TraJ transfer determinant was previously thought to be unique to IncHI plasmids, and was classified as a Dtr component (Lawley *et al.*, 2002a). After PSI-BLAST analysis of the R478-encoded TraJ sequence, sixteen potential orthologues were identified in

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

both plasmid and chromosomal genomes (Table A-1). Although there is modest sequence identity between these sixteen sequences and TraJ (13-31% local identity), each are ~220 residues in length, each contain four predicted transmembrane domains, and each have conserved subdomains, including a tandem of positive residues between the third and fourth transmembrane domains (Figure 2-7). The precise cellular function of TraJ is unknown, however the discovery of related sequences in numerous other genomes suggest that TraJ is an important Dtr determinant.

The IncHI Dtr and coupling determinants TraI and TraG were not considered unique to the IncHI complex because they contain conserved motifs found in the respective IncP and IncW-encoded Dtr and coupling determinants, such as relaxase and Walker NTP-binding motifs (Lawley *et al.*, 2002a; Sherburne *et al.*, 2000). An up-to-date PSI-BLAST analysis of R478-encoded relaxase and coupling proteins identified numerous coding sequences which have sequence identity to TraI or TraG outside of these motifs (Figures 2-5 and 2-6), therefore more closely related orthologues than the P and W determinants exist. Notably, eleven of the genomes which encode IncHI TraI and TraG-related sequences, also encode a TraJ orthologue (Table 2-2). This indicates that *traI*, *traG*, and *traJ* are genetically linked in numerous genomes, and that they might have been inherited from a common ancestor. For the purpose of this discussion, these three genes will collectively be designated the H-type Dtr/coupling determinants, representing a novel phylogenetic group of transfer sequences.

The Mpf module of the IncHI plasmids is related to the F-type Mpf module, yet the H-type Dtr/coupling determinants are distinct from the F-type Dtr/coupling sequences. In particular, the Mpf genes of R478 are closely related to the sequences encoded on pRS241d, pNL1, pSLT, and *V. vulnificus* (Figure 2-4), but these F-type genomes do not include H-type *traI-G-J* sequences. These observations suggest that the IncHI plasmid

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

transfer system evolved from two different lineages which converged to form a novel module that is functional for plasmid transfer.

Just as the IncHI Mpf module might have had a different cellular function prior to being adapted for conjugative transfer, the function of an ancestral H-type *traI-J-G* module (in the absence of Mpf functions) may have been related to a general DNA processing reaction. The conjugative relaxases such as TraI can participate in the most basic DNA metabolic functions: that of cutting and rejoining and therefore serving as both endonuclease and ligase. These are each essential features of both DNA recombination and replication (i.e. during lagging strand synthesis) reactions, and furthermore, type I topoisomerase enzymes similarly nick one strand of a DNA duplex and serve as a ligase (Cavalier-Smith, 2002). The coupling protein TraG has also been predicted to have a role in DNA metabolism, as revealed by the structure of IncW coupling protein TrwB (Gomis-Ruth *et al.*, 2001). The homohexameric structure of TrwB resembles that of DNA ring helicases which act as pumps for single DNA strands passing through the central channel, thereby serving as a motor for DNA translocation. TraJ contains no known functional motifs, but may be required for the coordinated function of TraI and TraG.

5.4 Functional adaptation of the Mpf and Dtr determinants

The versatile role of the coupling protein is not only indicated by its predicted ability to serve as an inner membrane pore for transfer of the single-stranded transfer intermediated, but by many clear examples of chimeric transfer systems comprised of non-cognate coupling proteins (Table 5-1). Some transferable elements encode only one of the two conjugative transfer modules, and mobilizable plasmids such as RSF1010 have only the Dtr module, and lack the Mpf module and a coupling protein. To support conjugative transfer, the remaining determinants must be supplied by a non-

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 5-1. Chimeric functional groups capable of mobilizing DNA. The plasmid source or lineage of the different classes (Mpf, coupling, Dtr/Mob) of conjugal transfer proteins are listed. For example, in this study R478 has H-type Dtr and coupling determinants, but F-type Mpf determinants. All other examples represent experimental systems where a functional apparatus has been reconstituted from different lineages of transfer determinants.

Mpf	Coupling	Dtr/Mob	Reference
F	H	H	This study
N	W	W	Llosa *et al.*, 2003
W	N	N	Llosa *et al.*, 2003
X	W	W	Llosa *et al.*, 2003
X	N	N	Llosa *et al.*, 2003
P	P	RSF1010	Lessl *et al.*, 1993
W	P	RSF1010	Cabezon *et al.*, 1994
W	P	ColE1	Cabezon *et al.*, 1994
W or F	F[a]	RSF1010	Sastre *et al.*, 1998
W or F	F	ColEI	Sastre *et al.*, 1998
Ti	Ti	RSF1010	Beijersbergen *et al.*, 1992
P	P	pBHR1	Szpirer *et al.*, 2000
P	P	pBGR1	Seubert *et al.*, 2003a
F, I, N, P, W	CloDF13	CloDF13	Cabezon *et al.*, 1997
X	CloDF13	CloDF13	Strauch *et al.*, 2003

a. An F plasmid TraD construct with a 140 residue C-terminal truncation was used

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

cognate system. Therefore, the RSF1010 Dtr components must interact with an unrelated coupling protein (Table 5-1). Other mobilizable plasmids such as CloDF13 encode both the Dtr and coupling determinants, but lack the Mpf module. In this instance, the CloDF13 coupling protein must interact with an unrelated Mpf module (Table 5-1). In addition, interactions between TrhB-related proteins and non-cognate coupling proteins have also been demonstrated (Table 5-1), which also confirms the results presented in Chapter 4 (Llosa *et al.*, 2003). These observations verify that interactions between transfer determinants from different lineages occur, and the coupling protein plays a central role in adaptation of the different transfer modules.

Select T4S effector translocation systems may also depend upon coupling proteins for substrate recruitment in a fashion similar to relaxase-DNA recruitment during conjugative transfer. Each of the *A. tumefaciens* Vir, *H. pylori* Cag, and *L. pneumophila* Dot/Icm systems require a coupling protein, whereas the T4S systems encoded by *Bordetella*, *Brucella*, and *Bartonella* spp. apparently function without a coupling protein (Cascales and Christie, 2003). It is possible that in these systems in which the coupling protein is not required for substrate translocation that the nature of the substrate does not require a coupling protein to interact with the transenvelope type IV secretion apparatus.

5.5 A transenvelope structure for macromolecular transport

The term 'conjugative apparatus' might be hyperbole considering that a transenvelope structure has never been visualized through traditional microscopy. Notably, the IncP relaxosome has been visualized under *in vivo* conditions (Pansegrau *et al.*, 1990) and a membrane fraction (of intermediate density between the inner and outer membrane fractions) containing the IncP Mpf proteins was observed (Grahn *et al.*, 2000). This latter observation supports the hypothesis that conjugal transfer proteins

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

indeed form a transenvelope structure, and this thesis has provided additional cytological evidence (Chapter 3). The visualization of fluorescent TrhC-GFP foci in live bacterial cells represents an indirect visualization for a macromolecular assembly of transfer proteins. These studies indicated that a large network of proteins is responsible for foci formation, and that ATPase function of TrhC is dispensable for complex assembly.

The Mpf, coupling, and Dtr modules are individually capable of distinct functions, such as DNA processing and pilus production; however, the cellular function of the conjugative apparatus is dependent upon the interaction and coordinated operation of these modules. Using the TrhC-GFP fusion as a label for the conjugative apparatus, mutations in Mpf, Dtr and coupling determinants each affected the assembly of this machine. TrhC is classified as an Mpf protein, yet non-Mpf proteins such as TraJ were critical for assembly of the TrhC-GFP foci. In addition, the interaction between TraG and the Mpf protein TrhB may account for the observed influence of TraG on TrhC-associated complex assembly, of which TrhB is a member. These observations highlight the relationship of the different structures that compose the transfer apparatus: distinct subassemblies and complexes may be observed, but coordination of all the elements as a singular unit may ultimately be required for function.

5.6 Future Directions

The three-dimensional structure of individual conjugative transfer proteins has been obtained (Gomis-Ruth *et al.*, 2001; Guasch *et al.*, 2003), and has in each instance provided a wealth of insight into the mechanism of DNA processing and transport. The data presented in this thesis indicates that the different transfer protein modules are dependent upon each other for assembly and function of the conjugative apparatus. It might be possible to purify and obtain structural information for heteromeric protein

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

complexes, and ultimately identify the architecture of the conjugative apparatus. Of particular interest would be the precise association of the coupling proteins and TrhB homologues at the inner membrane, and if the TrhB homologues extend through the periplasm to interact with an outer membrane-associated secretin-type transfer protein.

A detailed characterization of the conjugative apparatus might also aid in determination of the true substrate of conjugative transfer and the key proteins which make up the transenvelope pore. It is plausible that one or both of the transfer DNA strand or covalently attached relaxase are transferred to recipient cells. The majority of T4S systems seem geared for the export of protein substrates, including the *A. tumefaciens* system which transfers the VirD2 relaxase in association with the T-DNA. Furthermore, the precise roles of the numerous ATPase transfer proteins encoded by each conjugative system during substrate transport are still unknown. The ATP-binding motifs of TrhC were not required for assembly of TrhC-GFP-associated complexes, however, function during pilus assembly and/or DNA translocation remain possibilities.

An interesting paradox concerning the evolution of conjugative plasmids is that while being a vehicle for horizontal transfer of genetic information, they are also subject to acquisition of horizontally-transferred coding sequences. This is indicated by the large number of genes and operons encoded on the sequenced IncHI plasmids which were attributed to different origins. In particular, the IncHI transfer determinants have been acquired from at least two lineages. The ancestral history of the prototypic plasmid systems might be further revealed using the growing amount of sequence data, for both whole genomes and individual plasmids. The observation of chromosomal sequences which are related to plasmid-transfer determinants encoding DNA-processing and membrane-targeting features is intriguing and might suggest that a widespread and uncharacterized mechanism for DNA transport is encoded in multiple microbial genera. In addition, an increasing number of T4S effector translocation systems will undoubtedly

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

be uncovered in the genome data from both pathogenic and non-pathogenic bacteria. The relationship between these transport systems and true conjugative systems may also be discovered, suggesting a mechanism for the evolution of these analogous protein machines.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Bibliography

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Bibliography

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

Anderson, E.S., and Smith, H.R. (1972) Chloramphenicol resistance in the typhoid bacillus. *Br Med J* **3**: 329-331.

Anthony, K.G., Kathir, P., Moore, D., Ippen-Ihler, K., and Frost, L.S. (1996) Analysis of the *traLEKBP* sequence and the TraP protein from three F-like plasmids: F, R100-1 and ColB2. *J Bacteriol* **178**: 3194-3200.

Atmakuri, K., Ding, Z., and Christie, P.J. (2003) VirE2, a type IV secretion substrate, interacts with the VirD4 transfer protein at cell poles of *Agrobacterium tumefaciens*. *Mol Microbiol* **49**: 1699-1713.

Bachman, B.J. (1972) Pedigrees of some mutant types of *Escherichia coli* K-12. *Bacteriol Rev* **36**: 525-557.

Bacon, D.J., Alm, R.A., Hu, L., Hickey, T.E., Ewing, C.P., Batchelor, R.A., Trust, T.J., and Guerry, P. (2002) DNA sequence and mutational analyses of the pVir plasmid of *Campylobacter jejuni* 81-176. *Infect Immun* **70**: 6242-6250.

Balzer, D., Pansegrau, W., and Lanka, E. (1994) Essential motifs of relaxase (TraI) and TraG proteins involved in conjugative transfer of plasmid RP4. *J Bacteriol* **176**: 4285-4295.

Banta, L.M., Bohne, J., Lovejoy, S.D., and Dostal, K. (1998) Stability of the *Agrobacterium tumefaciens* VirB10 protein is modulated by growth temperature and periplasmic osmoadaption. *J Bacteriol* **180**: 6597-6606.

Baron, C., O'Callaghan, D., and Lanka, E. (2002) Bacterial secrets of secretion: EuroConference on the biology of type IV secretion processes. *Mol Microbiol* **43**: 1359-1365.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S.R., Griffiths-Jones, S., Howe, K.L., Marshall, M., and Sonnhammer, E.L. (2002) The Pfam protein families database. *Nucleic Acids Res* **30**: 276-280.

Beaber, J.W., Hochhut, B., and Waldor, M.K. (2002) Genomic and functional analyses of SXT, an integrating antibiotic resistance gene transfer element derived from *Vibrio cholerae*. *J Bacteriol* **184**: 4259-4269.

Becker, E.C., and Meyer, R.J. (2000) Recognition of oriT for DNA processing at termination of a round of conjugal transfer. *J Mol Biol* **300**: 1067-1077.

Beijersbergen, A., Dulk-Ras, A., Schilperoort, R., and Hooykaas, P. (1992) Conjugative transfer by the virulence system of *Agrobacterium tumefaciens*. *Science* **256**: 1324-1326.

Berger, B.R., and Christie, P.J. (1993) The *Agrobacterium tumefaciens virB4* gene product is an essential virulence protein requiring an intact nucleoside triphosphate-binding domain. *J Bacteriol* **175**: 1723-1734.

Birge, E.A. (1994) *Bacterial and bacteriophage genetics.* New York, N.Y.: Springer-Verlag.

Blocker, A., Jouihri, N., Larquet, E., Gounon, P., Ebel, F., Parsot, C., Sansonetti, P., and Allaoui, A. (2001) Structure and composition of the *Shigella flexneri* "needle complex", a part of its type III secreton. *Mol Microbiol* **39**: 652-663.

Boltner, D., MacMahon, C., Pembroke, J.T., Strike, P., and Osborn, A.M. (2002) R391: a Conjugative Integrating Mosaic Comprised of Phage, Plasmid, and Transposon Elements. *J Bacteriol* **184**: 5158-5169.

Bradley, D.E. (1980) Determination of pili by conjugative bacterial drug resistance plasmids of incompatibility groups B, C, H, J, K, M, V, and X. *J Bacteriol* **141**: 828-837.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Brown, N.L., Barrett, S.R., Camakaris, J., Lee, B.T., and Rouch, D.A. (1995) Molecular genetics and transport analysis of the copper-resistance determinant (pco) from *Escherichia coli* plasmid pRJ1004. *Mol Microbiol* **17**: 1153-1166.

Cabezon, E., Lanka, E., and de la Cruz, F. (1994) Requirements for mobilization of plasmids RSF1010 and ColE1 by the IncW plasmid R388: trwB and RP4 traG are interchangeable. *J Bacteriol* **176**: 4455-4458.

Cabezon, E., Sastre, J.I., and de la Cruz, F. (1997) Genetic evidence of a coupling role for the TraG protein family in bacterial conjugation. *Mol Gen Genet* **254**: 400-406.

Cascales, E., and Christie, P.J. (2003) The versatile bacterial type IV secretion systems. *Nat Microbiol Rev* **1**: 137-149.

Casjens, S. (2003) Prophages and bacterial genomics: what have we learned so far? *Mol Microbiol* **49**: 277-300.

Cavalier-Smith, T. (2002) Origins of the machinery of recombination and sex. *Heredity* **88**: 125-141.

Cha, J.S., and Cooksey, D.A. (1991) Copper resistance in *Pseudomonas syringae* mediated by periplasmic and outer membrane proteins. *Proc Natl Acad Sci U S A* **88**: 8915-8919.

Chang, C., Mooser, A., Pluckthun, A., and Wlodawer, A. (2001) Crystal structure of the dimeric C-terminal domain of TonB reveals a novel fold. *J Biol Chem* **276**: 27535-27540.

Christie, P.J. (2001) Type IV secretion: intercellular transfer of macromolecules by systems ancestrally related to conjugation machines. *Mol Microbiol* **40**: 294-305.

Collins, R.F., Davidsen, L., Derrick, J.P., Ford, R.C., and Tonjum, T. (2001) Analysis of the PilQ secretin from *Neisseria meningitidis* by transmission electron microscopy reveals a dodecameric quaternary structure. *J Bacteriol* **183**: 3825-3832.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Collins, R.F., Ford, R.C., Kitmitto, A., Olsen, R.O., Tonjum, T., and Derrick, J.P. (2003) Three-dimensional structure of the *Neisseria meningitidis* secretin PilQ determined from negative-stain transmission electron microscopy. *J Bacteriol* **185**: 2611-2617.

Cormack, B.P., Valdivia, R.H., and Falkow, S. (1996) FACS-optimized mutants of the green fluorescent protein (GFP). *Gene* **173**: 33-38.

Crosa, J.H. (1980) Three origins of replication are active in vivo in the R plasmid RSF1040. *J Biol Chem* **255**: 11075-11077.

Curtiss, R., 3rd (1969) Bacterial conjugation. *Annu Rev Microbiol* **23**: 69-136.

Dang, T.A., and Christie, P.J. (1997) The VirB4 ATPase of *Agrobacterium tumefaciens* is a cytoplasmic membrane protein exposed at the periplasmic surface. *J Bacteriol* **179**: 453-462.

Dang, T.A., Zhou, X.R., Graf, B., and Christie, P.J. (1999) Dimerization of the *Agrobacterium tumefaciens* VirB4 ATPase and the effect of ATP-binding cassette mutations on the assembly and function of the T-DNA transporter. *Mol Microbiol* **32**: 1239-1253.

Das, A., and Xie, Y.H. (1998) Construction of transposon Tn3phoA: its application in defining the membrane topology of the *Agrobacterium tumefaciens* DNA transfer proteins. *Mol Microbiol* **27**: 405-414.

Das, A., and Xie, Y.H. (2000) The *Agrobacterium* T-DNA transport pore proteins VirB8, VirB9, and VirB10 interact with one another. *J Bacteriol* **182**: 758-763.

Datz, M., Janetzki-Mittmann, C., Franke, S., Gunzer, F., Schmidt, H., and Karch, H. (1996) Analysis of the enterohemorrhagic *Escherichia coli* O157 DNA region containing lambdoid phage gene p and Shiga-like toxin structural genes. *Appl Environ Microbiol* **62**: 791-797.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Davies, J. (1994) Inactivation of antibiotics and the dissemination of resistance genes. *Science* **264**: 375-382.

del Solar, G., Giraldo, R., Ruiz-Echevarria, M.J., Espinosa, M., and Diaz-Orejas, R. (1998) Replication and control of circular bacterial plasmids. *Microbiol Mol Biol Rev* **62**: 434-464.

Delahay, R.M., and Frankel, G. (2002) Coiled-coil proteins associated with type III secretion systems: a versatile domain revisited. *Mol Microbiol* **45**: 905-916.

Delcher, A.L., Harmon, D., Kasif, S., White, O., and Salzberg, S.L. (1999) Improved microbial gene identification with GLIMMER. *Nucleic Acids Res* **27**: 4636-4641.

Deng, W., Liou, S.R., Plunkett, G., 3rd, Mayhew, G.F., Rose, D.J., Burland, V., Kodoyianni, V., Schwartz, D.C., and Blattner, F.R. (2003) Comparative genomics of *Salmonella enterica* serovar Typhi strains Ty2 and CT18. *J Bacteriol* **185**: 2330-2337.

Dillard, J.P., and Seifert, H.S. (2001) A variable genetic island specific for *Neisseria gonorrhoeae* is involved in providing DNA for natural transformation and is found more often in disseminated infection isolates. *Mol Microbiol* **41**: 263-277.

Ding, Z., Atmakuri, K., and Christie, P.J. (2003) The outs and ins of bacterial type IV secretion substrates. *Trends Microbiol* **11**: 527-535.

Doolittle, W.F. (2000) Uprooting the tree of life. *Sci Am* **282**: 90-95.

Doolittle, W.F., Boucher, Y., Nesbo, C.L., Douady, C.J., Andersson, J.O., and Roger, A.J. (2003) How big is the iceberg of which organellar genes in nuclear genomes are but the tip? *Philos Trans R Soc Lond B Biol Sci* **358**: 39-57; discussion 57-38.

Drake, S.L., and Koomey, M. (1995) The product of the pilQ gene is essential for the biogenesis of type IV pili in *Neisseria gonorrhoeae*. *Mol Microbiol* **18**: 975-986.

Dubnau, D. (1999) DNA uptake in bacteria. *Annu Rev Microbiol* **53**: 217-244.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Durrenberger, M.B., Villiger, W., and Bachi, T. (1991) Conjugational junctions: morphology of specific contacts in conjugating *Escherichia coli* bacteria. *J Struct Biol* **107**: 146-156.

Falquet, L., Pagni, M., Bucher, P., Hulo, N., Sigrist, C.J., Hofmann, K., and Bairoch, A. (2002) The PROSITE database, its status in 2002. *Nucleic Acids Res* **30**: 235-238.

Farizo, K.M., Huang, T., and Burns, D.L. (2000) Importance of holotoxin assembly in Ptl-mediated secretion of pertussis toxin from *Bordetella pertussis*. *Infect Immun* **68**: 4049-4054.

Faruque, S.M., and Mekalanos, J.J. (2003) Pathogenicity islands and phages in *Vibrio cholerae* evolution. *Trends Microbiol* **11**: 505-510.

Feilmeier, B.J., Iseminger, G., Schroeder, D., Webber, H., and Phillips, G.J. (2000) Green fluorescent protein functions as a reporter for protein localization in *Escherichia coli*. *J Bacteriol* **182**: 4068-4076.

Frishman, D., Mironov, A., Mewes, H.W., and Gelfand, M. (1998) Combining diverse evidence for gene recognition in completely sequenced bacterial genomes. *Nucleic Acids Res* **26**: 2941-2947.

Frost, L.S., Ippen-Ihler, K., and Skurray, R.A. (1994) Analysis of the sequence and gene products of the transfer region of the F sex factor. *Microbiol Rev* **58**: 162-210.

Fullner, K.J., Stephens, K.M., and Nester, E.W. (1994) An essential virulence protein of *Agrobacterium tumefaciens*, VirB4, requires an intact mononucleotide binding domain to function in transfer of T-DNA. *Mol Gen Genet* **245**: 704-715.

Fullner, K.J., and Nester, E.W. (1996) Temperature affects the T-DNA transfer machinery of *Agrobacterium tumefaciens*. *J Bacteriol* **178**: 1498-1504.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Gabant, P., Newnham, P., Taylor, D., and Couturier, M. (1993) Isolation and location on the R27 map of two replicons and an incompatibility determinant specific for IncHI1 plasmids. *J Bacteriol* **175**: 7697-7701.

Gatti, D., Mitra, B., and Rosen, B.P. (2000) *Escherichia coli* soft metal ion-translocating ATPases. *J Biol Chem* **275**: 34009-34012.

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

Ghigo, J.M. (2001) Natural conjugative plasmids induce bacterial biofilm development. *Nature* **412**: 442-445.

Gilmour, M.W., Lawley, T.D., Rooker, M.M., Newnham, P.J., and Taylor, D.E. (2001) Cellular location and temperature-dependent assembly of IncHI1 plasmid R27-encoded TrhC-associated conjugative transfer protein complexes. *Mol Microbiol* **42**: 705-715.

Gilmour, M.W., Gunton, J.E., Lawley, T.D., and Taylor, D.E. (2003) Interaction between the IncHI1 plasmid R27 coupling protein and type IV secretion system: TraG associates with the coiled-coil mating pair formation protein TrhB. *Mol Microbiol* **49**: 105-116.

Gomis-Ruth, F.X., Moncalian, G., Perez-Luque, R., Gonzalez, A., Cabezon, E., de la Cruz, F., and Coll, M. (2001) The bacterial conjugation protein TrwB resembles ring helicases and F1-ATPase. *Nature* **409**: 637-641.

Gomis-Ruth, F.X., de la Cruz, F., and Coll, M. (2002a) Structure and role of coupling proteins in conjugal DNA transfer. *Res Microbiol* **153**: 199-204.

Gomis-Ruth, F.X., Moncalian, G., de la Cruz, F., and Coll, M. (2002b) Conjugative plasmid protein TrwB, an integral membrane type IV secretion system coupling protein. Detailed structural features and mapping of the active site cleft. *J Biol Chem* **277**: 7556-7566.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Goncharoff, P., Saadi, S., Chang, C.H., Saltman, L.H., and Figurski, D.H. (1991) Structural, molecular, and genetic analysis of the kilA operon of broad-host-range plasmid RK2. *J Bacteriol* **173**: 3463-3477.

Gordon, G.S., Sitnikov, D., Webb, C.D., Teleman, A., Straight, A., Losick, R., Murray, A.W., and Wright, A. (1997) Chromosome and low copy plasmid segregation in *E. coli*: visual evidence for distinct mechanisms. *Cell* **90**: 1113-1121.

Grahn, A.M., Haase, J., Lanka, E., and Bamford, D.H. (1997) Assembly of a functional phage PRD1 receptor depends on 11 genes of the IncP plasmid mating pair formation complex. *J Bacteriol* **179**: 4733-4740.

Grahn, A.M., Haase, J., Bamford, D.H., and Lanka, E. (2000) Components of the RP4 conjugative transfer apparatus form an envelope structure bridging inner and outer membranes of donor cells: implications for related macromolecule transport systems. *J Bacteriol* **182**: 1564-1574.

Griffith, F. (1928) The significance of pneumococcal types. *J. Hyg* **27**: 113-159.

Grindley, N.D., Humphreys, G.O., and Anderson, E.S. (1973) Molecular studies of R factor compatibility groups. *J Bacteriol* **115**: 387-398.

Guasch, A., Lucas, M., Moncalian, G., Cabezas, M., Perez-Luque, R., Gomis-Ruth, F.X., De La Cruz, F., and Coll, M. (2003) Recognition and processing of the origin of transfer DNA by conjugative relaxase TrwC. *Nat Struct Biol* **10**: 1002-1010.

Guilvout, I., Hardie, K.R., Sauvonnet, N., and Pugsley, A.P. (1999) Genetic dissection of the outer membrane secretin PulD: are there distinct domains for multimerization and secretion specificity? *J Bacteriol* **181**: 7212-7220.

Gupta, A., and Silver, S. (1998) Silver as a biocide: will resistance become a problem? *Nat Biotechnol* **16**: 888.

Gupta, A., Phung, L.T., Taylor, D.E., and Silver, S. (2001) Diversity of silver resistance genes in IncH incompatibility group plasmids. *Microbiology* **147**: 3393-3402.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Haase, J., Lurz, R., Grahn, A.M., Bamford, D.H., and Lanka, E. (1995) Bacterial conjugation mediated by plasmid RP4: RSF1010 mobilization, donor-specific phage propagation, and pilus production require the same Tra2 core components of a proposed DNA transport complex. *J Bacteriol* **177**: 4779-4791.

Hamilton, C.M., Lee, H., Li, P.L., Cook, D.M., Piper, K.R., von Bodman, S.B., Lanka, E., Ream, W., and Farrand, S.K. (2000) TraG from RP4 and TraG and VirD4 from Ti plasmids confer relaxosome specificity to the conjugal transfer system of pTiC58. *J Bacteriol* **182**: 1541-1548.

Hamilton, H.L., Schwartz, K.J., and Dillard, J.P. (2001) Insertion-duplication mutagenesis of *Neisseria*: use in characterization of DNA transfer genes in the gonococcal genetic island. *J Bacteriol* **183**: 4718-4726.

Harnett, N., McLeod, S., AuYong, Y., Wan, J., Alexander, S., Khakhria, R., and Krishnan, C. (1998) Molecular characterization of multiresistant strains of *Salmonella typhi* from South Asia isolated in Ontario, Canada. *Can J Microbiol* **44**: 356-363.

Harris, R.L., Hombs, V., and Silverman, P.M. (2001) Evidence that F-plasmid proteins TraV, TraK and TraB assemble into an envelope-spanning structure in *Escherichia coli*. *Mol Microbiol* **42**: 757-766.

Harris, R.L., and Silverman, P.M. (2002) Roles of internal cysteines in the function, localization, and reactivity of the TraV outer membrane lipoprotein encoded by the F plasmid. *J Bacteriol* **184**: 3126-3129.

Hedges, R.W., Rodriguez-Lemoine, V., and Datta, N. (1975) R factors from *Serratia marcescens*. *J Gen Microbiol* **86**: 88-92.

Ho, Y., Gruhler, A., Heilbut, A., Bader, G.D., Moore, L., Adams, S.L., Millar, A., Taylor, P., Bennett, K., Boutilier, K., Yang, L., Wolting, C., Donaldson, I., Schandorff, S., Shewnarane, J., Vo, M., Taggart, J., Goudreault, M., Muskat, B., Alfarano, C.,

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Dewar, D., Lin, Z., Michalickova, K., Willems, A.R., Sassi, H., Nielsen, P.A., Rasmussen, K.J., Andersen, J.R., Johansen, L.E., Hansen, L.H., Jespersen, H., Podtelejnikov, A., Nielsen, E., Crawford, J., Poulsen, V., Sorensen, B.D., Matthiesen, J., Hendrickson, R.C., Gleeson, F., Pawson, T., Moran, M.F., Durocher, D., Mann, M., Hogue, C.W., Figeys, D., and Tyers, M. (2002) Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* **415**: 180-183.

Hofreuter, D., Odenbreit, S., and Haas, R. (2001) Natural transformation competence in *Helicobacter pylori* is mediated by the basic components of a type IV secretion system. *Mol Microbiol* **41**: 379-391.

Horiuchi, T., and Hidaka, M. (1988) Core sequence of two separable terminus sites of the R6K plasmid that exhibit polar inhibition of replication is a 20 bp inverted repeat. *Cell* **54**: 515-523.

Hou, Y., and Taylor, D.E. (1994) Incidence of tellurite resistance determinants among plasmids of different incompatibility groups. *Plasmid* **32**: 306-311.

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

Jain, R., Rivera, M.C., and Lake, J.A. (1999) Horizontal gene transfer among genomes: the complexity hypothesis. *Proc Natl Acad Sci U S A* **96**: 3801-3806.

Jakubowski, S.J., Krishnamoorthy, V., and Christie, P.J. (2003) *Agrobacterium tumefaciens* VirB6 protein participates in formation of VirB7 and VirB9 complexes required for type IV secretion. *J Bacteriol* **185**: 2867-2878.

Jin, S., Song, Y.N., Deng, W.Y., Gordon, M.P., and Nester, E.W. (1993) The regulatory VirA protein of *Agrobacterium tumefaciens* does not function at elevated temperatures. *J Bacteriol* **175**: 6830-6835.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Kaniga, K., Bossio, J.C., and Galan, J.E. (1994) The *Salmonella typhimurium* invasion genes invF and invG encode homologues of the AraC and PulD family of proteins. *Mol Microbiol* **13**: 555-568.

Karimova, G., Pidoux, J., Ullmann, A., and Ladant, D. (1998) A bacterial two-hybrid system based on a reconstituted signal transduction pathway. *Proc Natl Acad Sci U S A* **95**: 5752-5756.

Kariuki, S., Gilks, C., Revathi, G., and Hart, C.A. (2000) Genotypic analysis of multidrug-resistant *Salmonella enterica* Serovar typhi, Kenya. *Emerg Infect Dis* **6**: 649-651.

Kim, J.F., Wei, Z.M., and Beer, S.V. (1997) The hrpA and hrpC operons of *Erwinia amylovora* encode components of a type III pathway that secretes harpin. *J Bacteriol* **179**: 1690-1697.

Koonin, E.V. (2003) Horizontal gene transfer: the path to maturity. *Mol Microbiol* **50**: 725-727.

Koster, M., Bitter, W., de Cock, H., Allaoui, A., Cornelis, G.R., and Tommassen, J. (1997) The outer membrane component, YscC, of the Yop secretion machinery of *Yersinia enterocolitica* forms a ring-shaped multimeric complex. *Mol Microbiol* **26**: 789-797.

Krall, L., Wiedemann, U., Unsin, G., Weiss, S., Domke, N., and Baron, C. (2002) Detergent extraction identifies different VirB protein subassemblies of the type IV secretion machinery in the membranes of *Agrobacterium tumefaciens*. *Proc Natl Acad Sci U S A* **99**: 11405-11410.

Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E.L. (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* **305**: 567-580.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Kubori, T., Matsushima, Y., Nakamura, D., Uralil, J., Lara-Tejero, M., Sukhan, A., Galan, J.E., and Aizawa, S.I. (1998) Supramolecular structure of the *Salmonella typhimurium* type III protein secretion system. *Science* **280**: 602-605.

Kulaeva, O.I., Koonin, E.V., Wootton, J.C., Levine, A.S., and Woodgate, R. (1998) Unusual insertion element polymorphisms in the promoter and terminator regions of the *mucAB*-like genes of R471a and R446b. *Mutat Res* **397**: 247-262.

Kumar, R.B., Xie, Y.H., and Das, A. (2000) Subcellular localization of the *Agrobacterium tumefaciens* T-DNA transport pore proteins: VirB8 is essential for the assembly of the transport pore. *Mol Microbiol* **36**: 608-617.

Kumar, R.B., and Das, A. (2002) Polar location and functional domains of the *Agrobacterium tumefaciens* DNA transfer protein VirD4. *Mol Microbiol* **43**: 1523-1532.

Lai, E.M., Chesnokova, O., Banta, L.M., and Kado, C.I. (2000) Genetic and environmental factors affecting T-pilin export and T-pilus biogenesis in relation to flagellation of *Agrobacterium tumefaciens*. *J Bacteriol* **182**: 3705-3716.

Lanka, E., and Wilkins, B.M. (1995) DNA processing reactions in bacterial conjugation. *Annu Rev Biochem* **64**: 141-169.

Lawley, T.D., Burland, V., and Taylor, D.E. (2000) Analysis of the complete nucleotide sequence of the tetracycline-resistance transposon Tn*10*. *Plasmid* **43**: 235-239.

Lawley, T.D., Gilmour, M.W., Gunton, J.E., Standeven, L.J., and Taylor, D.E. (2002a) Functional and mutational analysis of conjugative transfer region 1 (Tra1) from the IncHI1 plasmid R27. *J Bacteriol* **184**: 2173-2180.

Lawley, T.D., Gordon, G.S., Wright, A., and Taylor, D.E. (2002b) Bacterial conjugative transfer: visualization of successful mating pairs and plasmid establishment in live *Escherichia coli*. *Mol Microbiol* **44**: 947-956.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Lawley, T.D. (2003) The life cycle of the conjugative plasmid. In *Biological Science* Edmonton: University of Alberta, pp. 1-168.

Lawley, T.D., Gilmour, M.W., J.E., G., Tracz, D.M., and Taylor, D.E. (2003a) Functional and mutational analysis of conjugative transfer Region 2 (Tra2) from the IncHI1 plasmid R27. *J Bacteriol* **185**: 581-591.

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

Lawley, T.D., and Taylor, D.E. (2003) Characterization of the double-partitioning modules of R27: correlating plasmid stability with plasmid localization. *J Bacteriol* **185**: 3060-3067.

Lawrence, J.G., and Roth, J.R. (1996) Selfish operons: horizontal transfer may drive the evolution of gene clusters. *Genetics* **143**: 1843-1860.

Lawrence, J.G. (2002) Gene transfer in bacteria: speciation without species? *Theor Popul Biol* **61**: 449-460.

Lawrence, J.G., and Hendrickson, H. (2003) Lateral gene transfer: when will adolescence end? *Mol Microbiol* **50**: 739-749.

Lee, E.H., Kornberg, A., Hidaka, M., Kobayashi, T., and Horiuchi, T. (1989) *Escherichia coli* replication termination protein impedes the action of helicases. *Proc Natl Acad Sci U S A* **86**: 9104-9108.

Lee, M.H., Kosuk, N., Bailey, J., Traxler, B., and Manoil, C. (1999) Analysis of F factor TraD membrane topology by use of gene fusions and trypsin-sensitive insertions. *J Bacteriol* **181**: 6108-6113.

Lemon, K.P., and Grossman, A.D. (2000) Movement of replicating DNA through a stationary replisome. *Mol Cell* **6**: 1321-1330.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Lessl, M., Balzer, D., Lurz, R., Waters, V.L., Guiney, D.G., and Lanka, E. (1992a) Dissection of IncP conjugative plasmid transfer: definition of the transfer region Tra2 by mobilization of the Tra1 region in trans. *J Bacteriol* **174**: 2493-2500.

Lessl, M., Pansegrau, W., and Lanka, E. (1992b) Relationship of DNA-transfer-systems: essential transfer factors of plasmids RP4, Ti and F share common sequences. *Nucleic Acids Res* **20**: 6099-6100.

Lessl, M., Balzer, D., Weyrauch, K., and Lanka, E. (1993) The mating pair formation system of plasmid RP4 defined by RSF1010 mobilization and donor-specific phage propagation. *J Bacteriol* **175**: 6415-6425.

Li, P.L., and Farrand, S.K. (2000) The replicator of the nopaline-type Ti plasmid pTiC58 is a member of the *repABC* family and is influenced by the TraR-dependent quorum-sensing regulatory system. *J Bacteriol* **182**: 179-188.

Liu, Z., and Binns, A.N. (2003) Functional Subsets of the VirB Type IV Transport Complex Proteins Involved in the Capacity of *Agrobacterium tumefaciens* To Serve as a Recipient in virB-Mediated Conjugal Transfer of Plasmid RSF1010. *J Bacteriol* **185**: 3259-3269.

Llosa, M., Bolland, S., and de la Cruz, F. (1994) Genetic organization of the conjugal DNA processing region of the IncW plasmid R388. *J Mol Biol* **235**: 448-464.

Llosa, M., Zunzunegui, S., and de la Cruz, F. (2003) Conjugative coupling proteins interact with cognate and heterologous VirB10-like proteins while exhibiting specificity for cognate relaxosomes. *Proc Natl Acad Sci U S A* **100**: 10465-10470.

Lobry, J.R., and Sueoka, N. (2002) Asymmetric directional mutation pressures in bacteria. *Genome Biol* **3**: RESEARCH0058.

Lupas, A., Van Dyke, M., and Stock, J. (1991) Predicting coiled coils from protein sequences. *Science* **252**: 1162-1164.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Lupas, A. (1996) Coiled coils: new structures and new functions. *Trends Biochem Sci* **21**: 375-382.

Lybarger, S.R., and Maddock, J.R. (2001) Polarity in action: asymmetric protein localization in bacteria. *J Bacteriol* **183**: 3261-3267.

Maeda, K., Nojiri, H., Shintani, M., Yoshida, T., Habe, H., and Omori, T. (2003) Complete nucleotide sequence of carbazole/dioxin-degrading plasmid pCAR1 in *Pseudomonas resinovorans* strain CA10 indicates its mosaicity and the presence of large catabolic transposon Tn4676. *J Mol Biol* **326**: 21-33.

Maher, D., Sherburne, R., and Taylor, D.E. (1991) Bacteriophages for incompatibility group H plasmids: morphological and growth characteristics. *Plasmid* **26**: 141-146.

Maher, D., Sherburne, R., and Taylor, D.E. (1993) H-pilus assembly kinetics determined by electron microscopy. *J Bacteriol* **175**: 2175-2183.

Maher, D., and Taylor, D.E. (1993) Host range and transfer efficiency of incompatibility group HI plasmids. *Can J Microbiol* **39**: 581-587.

Majewski, J., Zawadzki, P., Pickerill, P., Cohan, F.M., and Dowson, C.G. (2000) Barriers to genetic exchange between bacterial species: *Streptococcus pneumoniae* transformation. *J Bacteriol* **182**: 1016-1023.

Marchler-Bauer, A., Anderson, J.B., DeWeese-Scott, C., Fedorova, N.D., Geer, L.Y., He, S., Hurwitz, D.I., Jackson, J.D., Jacobs, A.R., Lanczycki, C.J., Liebert, C.A., Liu, C., Madej, T., Marchler, G.H., Mazumder, R., Nikolskaya, A.N., Panchenko, A.R., Rao, B.S., Shoemaker, B.A., Simonyan, V., Song, J.S., Thiessen, P.A., Vasudevan, S., Wang, Y., Yamashita, R.A., Yin, J.J., and Bryant, S.H. (2003) CDD: a curated Entrez database of conserved domain alignments. *Nucleic Acids Res* **31**: 383-387.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Margolin, W. (2000) Green fluorescent protein as a reporter for macromolecular localization in bacterial cells. *Methods* **20**: 62-72.

Martin, P.R., Hobbs, M., Free, P.D., Jeske, Y., and Mattick, J.S. (1993) Characterization of pilQ, a new gene required for the biogenesis of type 4 fimbriae in *Pseudomonas aeruginosa*. *Mol Microbiol* **9**: 857-868.

McHugh, G.L., Moellering, R.C., Hopkins, C.C., and Swartz, M.N. (1975) *Salmonella typhimurium* resistant to silver nitrate, chloramphenicol, and ampicillin. *Lancet* **1**: 235-240.

Medeiros, A.A., and O'Brien, T.F. (1969) Contribution of R factors to the antibiotic resistance of hospital isolates of Serratia. *Antimicrobial Agents Chemother* **8**: 30-35.

Mergeay, M., Nies, D., Schlegel, H.G., Gerits, J., Charles, P., and Van Gijsegem, F. (1985) *Alcaligenes eutrophus* CH34 is a facultative chemolithotroph with plasmid-bound resistance to heavy metals. *J Bacteriol* **162**: 328-334.

Miller, J.H. (1972) *Experiments in Molecular Genetics.* Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Mirza, S., Kariuki, S., Mamun, K.Z., Beeching, N.J., and Hart, C.A. (2000) Analysis of plasmid and chromosomal DNA of multidrug-resistant *Salmonella enterica* serovar typhi from Asia. *J Clin Microbiol* **38**: 1449-1452.

Miyamoto, H., Yoshida, S.I., Taniguchi, H., and Shuman, H.A. (2003) Virulence Conversion of *Legionella pneumophila* by Conjugal Transfer of Chromosomal DNA. *J Bacteriol* **185**: 6712-6718.

Moncalian, G., Cabezon, E., Alkorta, I., Valle, M., Moro, F., Valpuesta, J.M., Goni, F.M., and de La Cruz, F. (1999) Characterization of ATP and DNA binding activities of TrwB, the coupling protein essential in plasmid R388 conjugation. *J Biol Chem* **274**: 36117-36124.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Muller, K.M., Arndt, K.M., and Alber, T. (2000) Protein fusions to coiled-coil domains. *Methods Enzymol* **328**: 261-282.

Munson, G.P., Lam, D.L., Outten, F.W., and O'Halloran, T.V. (2000) Identification of a copper-responsive two-component system on the chromosome of *Escherichia coli* K-12. *J Bacteriol* **182**: 5864-5871.

Murata, T., Ohnishi, M., Ara, T., Kaneko, J., Han, C.G., Li, Y.F., Takashima, K., Nojima, H., Nakayama, K., Kaji, A., Kamio, Y., Miki, T., Mori, H., Ohtsubo, E., Terawaki, Y., and Hayashi, T. (2002) Complete nucleotide sequence of plasmid Rts1: implications for evolution of large plasmid genomes. *J Bacteriol* **184**: 3194-3202.

Nagai, H., Kagan, J.C., Zhu, X., Kahn, R.A., and Roy, C.R. (2002) A bacterial guanine nucleotide exchange factor activates ARF on *Legionella* phagosomes. *Science* **295**: 679-682.

Nagai, H., and Roy, C.R. (2003) Show me the substrates: modulation of host cell function by type IV secretion systems. *Cell Microbiol* **5**: 373-383.

Nakamura, Y., Gojobori, T., and Ikemura, T. (2000) Codon usage tabulated from international DNA sequence databases: status for the year 2000. *Nucleic Acids Res* **28**: 292.

Newnham, P.J., and Taylor, D.E. (1994) Molecular analysis of RepHI1A, a minimal replicon of the IncHI1 plasmid R27. *Mol Microbiol* **11**: 757-768.

Newnham, P.J. (1995) Transfer and replication systems of the IncHI1 plasmid R27. In *Department of Medical Microbiology and Immunology* Edmonton: Universtiy of Alberta, pp. 223.

Neyt, C., Iriarte, M., Thi, V.H., and Cornelis, G.R. (1997) Virulence and arsenic resistance in Yersiniae. *J Bacteriol* **179**: 612-619.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Nielsen, H., Engelbrecht, J., Brunak, S., and von Heijne, G. (1997) A neural network method for identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Int J Neural Syst* **8**: 581-599.

Nies, D.H. (1995) The cobalt, zinc, and cadmium efflux system CzcABC from *Alcaligenes eutrophus* functions as a cation-proton antiporter in *Escherichia coli. J Bacteriol* **177**: 2707-2712.

Nunez, B., and De La Cruz, F. (2001) Two atypical mobilization proteins are involved in plasmid CloDF13 relaxation. *Mol Microbiol* **39**: 1088-1099.

O'Callaghan, D., Cazevieille, C., Allardet-Servent, A., Boschiroli, M.L., Bourg, G., Foulongne, V., Frutos, P., Kulakov, Y., and Ramuz, M. (1999) A homologue of the *Agrobacterium tumefaciens* VirB and *Bordetella pertussis* Ptl type IV secretion systems is essential for intracellular survival of *Brucella suis. Mol Microbiol* **33**: 1210-1220.

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

Ou, J.T., and Anderson, T.F. (1970) Role of pili in bacterial conjugation. *J Bacteriol* **102**: 648-654.

Page, D.T., Whelan, K.F., and Colleran, E. (1999) Mapping studies and genetic analysis of transfer genes of the multi-resistant IncHI2 plasmid, R478. *FEMS Microbiol Lett* **179**: 21-29.

Page, D.T., Whelan, K.F., and Colleran, E. (2001) Characterization of two autoreplicative regions of the IncHI2 plasmid R478: RepHI2A and RepHI1A((R478)). *Microbiology* **147**: 1591-1598.

Pansegrau, W., Balzer, D., Kruft, V., Lurz, R., and Lanka, E. (1990) In vitro assembly of relaxosomes at the transfer origin of plasmid RP4. *Proc Natl Acad Sci U S A* **87**: 6555-6559.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Pansegrau, W., and Lanka, E. (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**: 13068-13076.

Paranchych, W. (1975) Attachment, ejection and penetration stages of the RNA phage infectious process. In *RNA phages*. Zinder, N.D. (ed). Cold Spring Harbour: Cold Spring Harour Laboratory.

Parkhill, J., Achtman, M., James, K.D., Bentley, S.D., Churcher, C., Klee, S.R., Morelli, G., Basham, D., Brown, D., Chillingworth, T., Davies, R.M., Davis, P., Devlin, K., Feltwell, T., Hamlin, N., Holroyd, S., Jagels, K., Leather, S., Moule, S., Mungall, K., Quail, M.A., Rajandream, M.A., Rutherford, K.M., Simmonds, M., Skelton, J., Whitehead, S., Spratt, B.G., and Barrell, B.G. (2000) Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491. *Nature* **404**: 502-506.

Parkhill, J., Dougan, G., James, K.D., Thomson, N.R., Pickard, D., Wain, J., Churcher, C., Mungall, K.L., Bentley, S.D., Holden, M.T., Sebaihia, M., Baker, S., Basham, D., Brooks, K., Chillingworth, T., Connerton, P., Cronin, A., Davis, P., Davies, R.M., Dowd, L., White, N., Farrar, J., Feltwell, T., Hamlin, N., Haque, A., Hien, T.T., Holroyd, S., Jagels, K., Krogh, A., Larsen, T.S., Leather, S., Moule, S., O'Gaora, P., Parry, C., Quail, M., Rutherford, K., Simmonds, M., Skelton, J., Stevens, K., Whitehead, S., and Barrell, B.G. (2001) Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi CT18. *Nature* **413**: 848-852.

Parkhill, J., Sebaihia, M., Preston, A., Murphy, L.D., Thomson, N., Harris, D.E., Holden, M.T., Churcher, C.M., Bentley, S.D., Mungall, K.L., Cerdeno-Tarraga, A.M., Temple, L., James, K., Harris, B., Quail, M.A., Achtman, M., Atkin, R., Baker, S., Basham, D., Bason, N., Cherevach, I., Chillingworth, T., Collins, M., Cronin, A.,

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Davis, P., Doggett, J., Feltwell, T., Goble, A., Hamlin, N., Hauser, H., Holroyd, S., Jagels, K., Leather, S., Moule, S., Norberczak, H., O'Neil, S., Ormond, D., Price, C., Rabbinowitsch, E., Rutter, S., Sanders, M., Saunders, D., Seeger, K., Sharp, S., Simmonds, M., Skelton, J., Squares, R., Squares, S., Stevens, K., Unwin, L., Whitehead, S., Barrell, B.G., and Maskell, D.J. (2003) Comparative analysis of the genome sequences of *Bordetella pertussis*, *Bordetella parapertussis* and Bordetella *bronchiseptica*. *Nat Genet* **35**: 32-40.

Pearson, W.R., and Lipman, D.J. (1988) Improved tools for biological sequence comparison. *Proc Natl Acad Sci U S A* **85**: 2444-2448.

Perna, N.T., Plunkett, G., 3rd, Burland, V., Mau, B., Glasner, J.D., Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J., Kirkpatrick, H.A., Posfai, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L., Grotbeck, E.J., Davis, N.W., Lim, A., Dimalanta, E.T., Potamousis, K.D., Apodaca, J., Anantharaman, T.S., Lin, J., Yen, G., Schwartz, D.C., Welch, R.A., and Blattner, F.R. (2001) Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. *Nature* **409**: 529-533.

Peters, J.E., and Craig, N.L. (2001) Tn*7*: smarter than we thought. *Nat Rev Mol Cell Biol* **2**: 806-814.

Pogliano, J., Ho, T.Q., Zhong, Z., and Helinski, D.R. (2001) Multicopy plasmids are clustered and localized in *Escherichia coli*. *Proc Natl Acad Sci U S A* **98**: 4486-4491.

Rabel, C., Grahn, A.M., Lurz, R., and Lanka, E. (2003) The VirB4 family of proposed traffic nucleoside triphosphatases: common motifs in plasmid RP4 TrbE are essential for conjugation and phage adsorption. *J Bacteriol* **185**: 1045-1058.

Rodriguez-Lemoine, V., Jacob, A.E., Hedges, R.W., and Datta, N. (1975) Thermosensitive production of their transfer systems by group S plasmids. *J Gen Microbiol* **86**: 111-114.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Rohde, M., Puls, J., Buhrdorf, R., Fischer, W., and Haas, R. (2003) A novel sheathed surface organelle of the *Helicobacter pylori* cag type IV secretion system. *Mol Microbiol* **49**: 219-234.

Romine, M.F., Stillwell, L.C., Wong, K.K., Thurston, S.J., Sisk, E.C., Sensen, C., Gaasterland, T., Fredrickson, J.K., and Saffer, J.D. (1999) Complete sequence of a 184-kilobase catabolic plasmid from *Sphingomonas aromaticivorans* F199. *J Bacteriol* **181**: 1585-1602.

Rooker, M.M., Sherburne, C., Lawley, T.D., and Taylor, D.E. (1999) Characterization of the Tra2 region of the IncHI1 plasmid R27. *Plasmid* **41**: 226-239.

Roussel, A.F., and Chabbert, Y.A. (1978) Taxonomy and epidemiology of gram-negative bacterial plasmids studied by DNA-DNA filter hybridization in formamide. *J Gen Microbiol* **104**: 269-276.

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.A., and Barrell, B. (2000) Artemis: sequence visualization and annotation. *Bioinformatics* **16**: 944-945.

Ryan, D., and Colleran, E. (2002) Arsenical resistance in the IncHI2 plasmids. *Plasmid* **47**: 234-240.

Samuels, A.L., Lanka, E., and Davies, J.E. (2000) Conjugative junctions in RP4-mediated mating of *Escherichia coli*. *J Bacteriol* **182**: 2709-2715.

Sastre, J.I., Cabezon, E., and de la Cruz, F. (1998) The carboxyl terminus of protein TraD adds specificity and efficiency to F-plasmid conjugative transfer. *J Bacteriol* **180**: 6039-6042.

Schandel, K.A., Muller, M.M., and Webster, R.E. (1992) Localization of TraC, a protein involved in assembly of the F conjugative pilus. *J Bacteriol* **174**: 3800-3806.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Schmiederer, M., Arcenas, R., Widen, R., Valkov, N., and Anderson, B. (2001) Intracellular induction of the *Bartonella henselae* virB operon by human endothelial cells. *Infect Immun* **69**: 6495-6502.

Schneider, E., and Hunke, S. (1998) ATP-binding-cassette (ABC) transport systems: functional and structural aspects of the ATP-hydrolyzing subunits/domains. *FEMS Microbiol Rev* **22**: 1-20.

Schroder, G., Krause, S., Zechner, E.L., Traxler, B., Yeo, H.J., Lurz, R., Waksman, G., and Lanka, E. (2002) TraG-like proteins of DNA transfer systems and of the *Helicobacter pylori* type IV secretion system: inner membrane gate for exported substrates? *J Bacteriol* **184**: 2767-2779.

Schuch, R., and Maurelli, A.T. (2001) MxiM and MxiJ, base elements of the Mxi-Spa type III secretion system of *Shigella*, interact with and stabilize the MxiD secretin in the cell envelope. *J Bacteriol* **183**: 6991-6998.

Segal, E.D., Cha, J., Lo, J., Falkow, S., and Tompkins, L.S. (1999) Altered states: involvement of phosphorylated CagA in the induction of host cellular growth changes by *Helicobacter pylori*. *Proc Natl Acad Sci U S A* **96**: 14559-14564.

Seubert, A., Falch, C., Birtles, R.J., Schulein, R., and Dehio, C. (2003a) Characterization of the cryptic plasmid pBGR1 from *Bartonella grahamii* and construction of a versatile *Escherichia coli-Bartonella* spp. shuttle cloning vector. *Plasmid* **49**: 44-52.

Seubert, A., Hiestand, R., de la Cruz, F., and Dehio, C. (2003b) A bacterial conjugation machinery recruited for pathogenesis. *Mol Microbiol* **49**: 1253-1266.

Shanahan, P.M., Jesudason, M.V., Thomson, C.J., and Amyes, S.G. (1998) Molecular analysis of and identification of antibiotic resistance genes in clinical isolates of *Salmonella typhi* from India. *J Clin Microbiol* **36**: 1595-1600.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Shanahan, P.M., Karamat, K.A., Thomson, C.J., and Amyes, S.G. (2000) Characterization of multi-drug resistant *Salmonella typhi* isolated from Pakistan. *Epidemiol Infect* **124**: 9-16.

Sherburne, C.K., Lawley, T.D., Gilmour, M.W., Blattner, F.R., Burland, V., Grotbeck, E., Rose, D.J., and Taylor, D.E. (2000) The complete DNA sequence and analysis of R27, a large IncHI plasmid from *Salmonella typhi* that is temperature sensitive for transfer. *Nucleic Acids Res* **28**: 2177-2186.

Silver, S., and Phung, L.T. (1996) Bacterial heavy metal resistance: new surprises. *Annu Rev Microbiol* **50**: 753-789.

Skare, J.T., Ahmer, B.M., Seachord, C.L., Darveau, R.P., and Postle, K. (1993) Energy transduction between membranes. TonB, a cytoplasmic membrane protein, can be chemically cross-linked in vivo to the outer membrane receptor FepA. *J Biol Chem* **268**: 16302-16308.

Smith, H.R., Grindley, N.D., Humphreys, G.O., and Anderson, E.S. (1973) Interactions of group H resistance factors with the F factor. *J Bacteriol* **115**: 623-628.

Sonea, S., and Mathieu, L.G. (2001) Evolution of the genomic systems of prokaryotes and its momentous consequences. *Int Microbiol* **4**: 67-71.

Strack, B., Lessl, M., Calendar, R., and Lanka, E. (1992) A common sequence motif, -E-G-Y-A-T-A-, identified within the primase domains of plasmid-encoded I- and P-type DNA primases and the alpha protein of the *Escherichia coli* satellite phage P4. *J Biol Chem* **267**: 13062-13072.

Strauch, E., Goelz, G., Knabner, D., Konietzny, A., Lanka, E., and Appel, B. (2003) A cryptic plasmid of *Yersinia enterocolitica* encodes a conjugative transfer system related to the regions of CloDF13 Mob and IncX Pil. *Microbiology* **149**: 2829-2845.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Summers, A.O. (1992) Untwist and shout: a heavy metal-responsive transcriptional regulator. *J Bacteriol* **174**: 3097-3101.

Szpirer, C.Y., Faelen, M., and Couturier, M. (2000) Interaction between the RP4 coupling protein TraG and the pBHR1 mobilization protein Mob. *Mol Microbiol* **37**: 1283-1292.

Tanaka, J., Suzuki, T., Mimuro, H., and Sasakawa, C. (2003) Structural definition on the surface of *Helicobacter pylori* type IV secretion apparatus. *Cell Microbiol* **5**: 395-404.

Tang, H., Braun, T.F., and Blair, D.F. (1996) Motility protein complexes in the bacterial flagellar motor. *J Mol Biol* **261**: 209-221.

Taylor, D.E., and Grant, R.B. (1976) Inhibition of bacteriophage lambda, T1, and T7 development by R plasmids of the H incompatibility group. *Antimicrob Agents Chemother* **10**: 762-764.

Taylor, D.E., and Grant, R.B. (1977a) R plasmids of the S incompatibility group belong to the H2 incompatibility group. *Antimicrob Agents Chemother* **12**: 431-434.

Taylor, D.E., and Grant, R.B. (1977b) Incompatibility and bacteriophage inhibition properties of N-1, a plasmid belonging to the H2 incompatibility group. *Mol Gen Genet* **153**: 5-10.

Taylor, D.E., and Levine, J.G. (1980) Studies of temperature-sensitive transfer and maintenance of H incompatibility group plasmids. *J Gen Microbiol* **116**: 475-484.

Taylor, D.E. (1983) Transfer-defective and tetracycline-sensitive mutants of the incompatibility group HI plasmid R27 generated by insertion of transposon 7. *Plasmid* **9**: 227-239.

Taylor, D.E., Hedges, R.W., and Bergquist, P.L. (1985) Molecular homology and incompatibility relationships between F and IncH1 plasmids. *J Gen Microbiol* **131**: 1523-1530.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Taylor, D.E. (1989) General properties of resistance plasmids. In *Handbook of Experimental Pharmacology*. Vol. 91. Bryan, L.E. (ed). Berlin: Springer-Verlag, pp. 325-357.

Taylor, D.E. (1999) Bacterial tellurite resistance. *Trends Microbiol* **7**: 111-115.

Taylor, D.E., Newnham, P.J., Sherburne, C., Lawley, T.D., and Rooker, M.M. (1999) Sequencing and characterization of *Salmonella typhi* plasmid R27 (incompatibility group HI1) *trhC*, a transfer gene encoding a potential nucleoside triphosphate-binding domain. *Plasmid* **41**: 207-218.

Taylor, D.E., Rooker, M., Keelan, M., Ng, L.K., Martin, I., Perna, N.T., Burland, N.T., and Blattner, F.R. (2002) Genomic variability of O islands encoding tellurite resistance in enterohemorrhagic *Escherichia coli* O157:H7 isolates. *J Bacteriol* **184**: 4690-4698.

Tetaz, T.J., and Luke, R.K. (1983) Plasmid-controlled resistance to copper in *Escherichia coli*. *J Bacteriol* **154**: 1263-1268.

Thomas, C.M. (2000) Paradigms of plasmid organization. *Mol Microbiol* **37**: 485-491.

Thomas, D., Morgan, D.G., and DeRosier, D.J. (2001) Structures of bacterial flagellar motors from two FliF-FliG gene fusion mutants. *J Bacteriol* **183**: 6404-6412.

Tian, Q.B., Hayashi, T., Murata, T., and Terawaki, Y. (1996) Gene product identification and promoter analysis of hig locus of plasmid Rts1. *Biochem Biophys Res Commun* **225**: 679-684.

Toptchieva, A., Sisson, G., Bryden, L.J., Taylor, D.E., and Hoffman, P.S. (2003) An inducible tellurite-resistance operon in *Proteus mirabilis*. *Microbiology* **149**: 1285-1295.

Turner, R.J., Weiner, J.H., and Taylor, D.E. (1994) In vivo complementation and site-specific mutagenesis of the tellurite resistance determinant kilAtelAB from IncP alpha plasmid RK2Ter. *Microbiology* **140**: 1319-1326.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Turner, R.J., Weiner, J.H., and Taylor, D.E. (1999) Tellurite-mediated thiol oxidation in *Escherichia coli*. *Microbiology* **145**: 2549-2557.

Vergunst, A.C., Schrammeijer, B., den Dulk-Ras, A., de Vlaam, C.M., Regensburg-Tuink, T.J., and Hooykaas, P.J. (2000) VirB/D4-dependent protein translocation from *Agrobacterium* into plant cells. *Science* **290**: 979-982.

von Heijne, G. (1986) A new method for predicting signal sequence cleavage site. *Nucleic Acids Res* **14**: 4683-4690.

Wain, J., Diem Nga, L.T., Kidgell, C., James, K., Fortune, S., Song Diep, T., Ali, T., P, O.G., Parry, C., Parkhill, J., Farrar, J., White, N.J., and Dougan, G. (2003) Molecular analysis of incHI1 antimicrobial resistance plasmids from *Salmonella* serovar Typhi strains associated with typhoid fever. *Antimicrob Agents Chemother* **47**: 2732-2739.

Waldor, M.K., and Mekalanos, J.J. (1996) Lysogenic conversion by a filamentous phage encoding cholera toxin. *Science* **272**: 1910-1914.

Ward, D.V., Draper, O., Zupan, J.R., and Zambryski, P.C. (2002) Peptide linkage mapping of the *Agrobacterium tumefaciens vir*-encoded type IV secretion system reveals protein subassemblies. *Proc Natl Acad Sci U S A* **99**: 11493-11500.

Ward, J.E., Dale, E.M., Nester, E.W., and Binns, A.N. (1990) Identification of a virB10 protein aggregate in the inner membrane of *Agrobacterium tumefaciens*. *J Bacteriol* **172**: 5200-5210.

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

Whelan, K.F., Maher, D., Colleran, E., and Taylor, D.E. (1994) Genetic and nucleotide sequence analysis of the gene *htdA*, which regulates conjugal transfer of IncHI plasmids. *J Bacteriol* **176**: 2242-2251.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Whelan, K.F., Colleran, E., and Taylor, D.E. (1995) Phage inhibition, colicin resistance, and tellurite resistance are encoded by a single cluster of genes on the IncHI2 plasmid R478. *J Bacteriol* **177**: 5016-5027.

Whelan, K.F., Sherburne, R.K., and Taylor, D.E. (1997) Characterization of a region of the IncHI2 plasmid R478 which protects *Escherichia coli* from toxic effects specified by components of the tellurite, phage, and colicin resistance cluster. *J Bacteriol* **179**: 63-71.

Whiteley, M., and Taylor, D.E. (1983) Identification of DNA homologies among H incompatibility group plasmids by restriction enzyme digestion and Southern transfer hybridization. *Antimicrob Agents Chemother* **24**: 194-200.

Willetts, N. (1981) Sites and systems for conjugal DNA transfer in bacteria. In *Molecular Biology, Pathogenicity and Ecology of Bacterial Plasmids*. Levy, S.B., Clowes, R.C. and Koenig, E.L. (eds). New York: Plenum Press, pp. 207-215.

Woese, C.R., Kandler, O., and Wheelis, M.L. (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci U S A* **87**: 4576-4579.

Yu, D., Ellis, H.M., Lee, E.C., Jenkins, N.A., Copeland, N.G., and Court, D.L. (2000) An efficient recombination system for chromosome engineering in *Escherichia coli*. *Proc Natl Acad Sci U S A* **97**: 5978-5983.

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

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Appendix

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table A-1. Predicted coding sequences of plasmid R478. Each amino acid sequence was compared against the non-redundant protein database at the National Center for Biotechnology Information (NCBI) using BLAST or PSI-BLAST and similar sequences from R27, pHCM1, and any other genetic element are reported.

Gene/Protein	Co-ordinates[a]	Amino acids	pHCM1 / R27 Homologue	% Identity/Range (E value)[b]	Other Homologues	%Identity/Range (E value)[b]	Accession number
SMR0001 RepHIA	1-876	291	HCM1.64 (RepA2) R0036 (RepHIA) HCM1.54 (RepA) R0050 (RepHIB)	86%/289 (e-134) 86%/289 (e-134) 36%/288 (e-41) 36%/288 (e-41)	replication protein A (Yersnia pMT1) RepA (pHCM2) RepA (Vir plasmid, S. cholerasesuis) RepA (pSLT) RepFIB (F plasmid) RepA (pO157) RepA (P1 phage) RepA (Rts1) RepA2 (pS268a, Serratia sp.)	42%/286 (2e-61) 42%/286 (3e-61) 42%/280 (e-60) 42%/280 (2e-60) 43%/273 (9e-60) 44%/267 (3e-59) 44%/269 (2e-59) 41%/278 (4e-52) 41%/115 (4e-19)	AAC13203 NP_569603 NP_073241 NP_490513 AAA71886 NP_052630 BVBPRA NP_639962 AAL84746
SMR0004 TrhA	2025-2378	117	HCM1.67 (TrhA) R0034 (TrhA)	65.8%/117 (3.9e-28) 65.8%/117 (3.9e-28)	TraA (Vibreo cholera, SXT) TraA (R391) TrhA (pCAR1, Pseudo. resinovorans) 20631 (Vibrio vulnificus)	18%/86 (0.71) 18%/86 23%78 18.5%/146	AAL59683 AAM08062 NP_758680 NP_762584
SMR0005 TrhL	2396-2746	116	HCM1.69 (TrhL) R0033 (TrhL)	71.6%/116 (4.9e-36) 71.6%/116 (4.9e-36)	Rsph4259 TrhL (pCAR1) 20632 (Vibrio vulnificus)	16.8%/137 17.2%/122 16.6/145	ZP_00008310 NP_758669 NP_762585
SMR0006 TrhE	2758-3546	262	HCM1.70 (TrhE) R0032 (TrhE)	74.3%/261 (2.6e-76) 74.3%/261 (2.6e-76)	TrhE (pCAR1) TraE (SXT) TraE (R391) ORF208 (Rts1) TraE (pED208, S. typhimurium) TraE (R100) TraE (pSLT) Rsph4260 20633 (Vibrio vulnificus)	26%/170 (2e-10) 26%/167 (1e-14) 25%/115 (e-5) 25%/117 (9e-5) 26%/173 (3e-4) 25%/147 (0.018) 22%/171 17.9%/162 18.2%/292	NP_758670 AAL59679 AAM08060 NP_640168 AAM90705 NP_052949 NP_490565 ZP_00008311 NP_762586
SMR0007 TrhK	3546-4817	423	HCM1.71 R0031 (TrhK)	79.7%/423 (1.6e-100) 79.7%/423 (1.6e-100)	ORF209 (Rts1) TraK (R391) TrhK (pCAR1) s048 (SXT) RscC (P. flurorscens) HrcC (P. syringae) Rsph4261 20634 (Vibrio vulnificus)	24%/376 (8e-11) 26%/253 (1e-06) 25%/258 (6e-8) 26%/253 (8e-7) 25%/124 (0.013) 32%/68 (0.15) 25.7%/113 19.5%/159	NP_640169 AAM08021 NP_758671 AAL59718 AAK81929 AAL34756 ZP_00008312 NP_762587
SMR0008 HtdO	4819-5280	153	HCM1.72 R0030	45.1%/144 (4.6e-18) 45.1%/144 (4.6e-18)	-	-	-
SMR0009 TrhB	5270-6625	451	HCM1.73 (TrhB) R0029 (TrhB)	86.5%/452 (1.5e-128) 86.5%/452 (1.5e-128)	ORF210 (Rts1) TrhB (pCAR1) TraB (SXT) TraB (R391) Rsph4262 (Rhodobacter sphaeroides) TraB (pSLT) TraB (R100) TraB (pED208)	29%/462 (4e-42) 27%/464 (2e-36) 39% /210 (3e-36) 39%/210 (3e-36) 31%/228 (2e-18) 31%/207 (5e-18) 30%/207 (5e-17) 24%/453 (6e-14)	NP_640170 NP_758672 AAL59682 AAM08009 ZP_00008313 NP_490567 NP_052951 AAM90707

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

					TraB (pNL1) 20635 (Vibrio vulnificus)	25%/427 (4e-11) 23.2%/478	NP_049165 NP_762588
SMR0010 HtdV	6633-7115	160	HCM1.74 R0028	53.3%/152 (1.2e-25) 53.3%/152 (1.2e-25)			
SMR0011 HtdT	7129-7986	285	HCM1.75 R0027	77%/285 (e-125) 82%/250 (e-121)	DsbC (R391) disulphide bond isomerase (SXT) ORF212 (Rts1) DsbC (pCAR1)	26%/195 (9e-12) 26%/195 (9e-12) 22%/240 (0.014) 23%/197	AAM08034 AAL59715 NP_640172 NP_758681
SMR0012 TrhV	7996-8946	316	HCM1.76 (TrhV) R0025 (TrhV)	80%/316 (2.3e-89) 80%/316 (2.3e-89)	TraV (SXT) HtdD (R391) ORF211 (Rts1) TrhV (pCAR1) Rsph4263 20636 (Vibrio vulnificus)	54%/48 (0.019) and 23%/214 (1.6) 54%/48 (0.020) and 32%/67 (0.074) 36%/66 (0.52) and 36%/74 (0.98) 32%/61 (0.078) 16.1%/273 20.9%/148	AAL59672 AAM08037 NP_640171 NP_758673 ZP_00008314 NP_762589
SMR0013 TrhC	8955-11636	893	HCM1.77 (TrhC) R0024 (TrhC)	83.9%/893 (0) 83.9%/893 (0)	TrhC (pCAR1) TraC (SXT) TraC (R391) ORF213 (Rts1) Rsph4264 TraC (pSLT) TraC (F) TraC (R100) TraC (pNL1) TrwK (R388) 20637 (Vibrio vulnificus) 20638 (Vibrio vulnificus)	28%/840 (6e-85) 27%/868 (7e-90) 27%/868 (7e-90) 26%/870 (3e-75) 24%/575 (9e-40) 22%/901 (5e-29) 22.9%/900 (2.7e-9) 22%/902 (5e-29) 22%532 (8e-25) 21%/178 (7e-6) 18.6%/451 24.1%/382	NP_758682 AAL59681 AAM08001 NP_640173 ZP_00008315 AAL23492 AAB61935 NP_052960 NP_049162 CAC78982 NP_762590 NP_762591
SMR0014 ParA	15475-16728	417	HCM1.86 R0020 (ParA)	89.4%/416 (1.6e-143)	VCA1115 (Vibreo cholera) ParA (pMT-1, Y. pestis) ParA (phage P7) ParA (phage P1) ParA (pSLT)	27.6%/395 (5.6e-23) 28%/234 (2e-10) 28%/232 (2e-10) 24%/264 (2e-10) 28%/222 (6e-10)	AAF97006 AAC82736 S06099 BVECPA NP_490542
SMR0015 ParB	16725-17729	334	HCM1.87 R0019 (ParB)	72.1%/330 (1.8 e-88) 72.1%/330 (1.8 e-88)	ParB (P1) ParB (Shigella flexneri pCP301) ParB (pSLT) ParB (Rts1) ParB (S. cholerasuis vir plasmid) ParB (Yersinia pestis pMT1) ParB (P7) SopB (pO157 and F)	28%/303 (6e-21) 28%/296 (2e-20) 28%/297 (2e-19) 27%/289 (3e-18) 27%/262 (1e-15) 26%/295 (8e-15) 27%/291 (3e-16) 24%/161 (6e-5)	AAA99231 AAL72312 AAL23465 BAB93856 BAB20544 CAB55249 S06100 NP_052641
SMR0016 TrhZ	c17791-18180	129	HCM1.88c R0018 (TrhZ)	80.5%/128 (2.3e-38) 80.5%/128 (2.3e-38)	-	-	-
SMR0017	c18177-18698	173	R0017 HCM1.89C	57.7%/168 (2.9e-37) 58.3%/168 (1.8e-37)	-	-	-
SMR0018	c18691-19527	278	HCM1.90c R0016	45.1%/237 (3.7e-35) 45.1%/237 (3.7e-35)	-	-	-

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

SMR0019 TrhO	c19524-20471	315	HCM1.91c R0015 (TrhO)	71.7%/314 (1.2e-95) 71.7%/314 (1.2e-95)	Ser/Thr phosphatase (PphB, pSLT) ORF75 (pCAR1) * and numerous phosphatases (1G5B)	25%/120 (0.72) 34%/67 (0.019)	AAL21787 NP_758617
SMR0020 ParM	20835-21872	345	HCM1.92 R0014 (ParM)	80.2%/344 (7.7e-101) 80.2%/344 (7.7e-101)	Z2979 (E. coli O157:H7) StbA (pB171, EPEC) StbA (Shigella flexneri pCP301) ParM (R1) ParA (pCTXM3, Citrobacter freundii)	36%/317 (2e-45) 37%/324 (4e-44) 35%/330(4e-40) 31%/316 (2e-35) 31%/327 (e-28)	AAG56911 BAA84903 AAL72301 PDB id: **1MWK** NP_774986
SMR0021 ParR	21884-22522	212	HCM1.93 R0013 (ParR)	41.4%/222 (1.2e-19) 41.4%/222 (1.2e-19)	-	-	-
SMR0022 HtdA	22925-23377	150	HCM1.94 (HtdA) R0012 (HtdA)	84.6%/149 (2e-51) 84.6%/149 (2e-51)	-	-	-
SMR0023 HtdF	23367-23792	141	HCM1.95 (HtdF) R0011 (HtdF)	50.7%/142 (2e-24) 50.7%/142 (2e-24)	-	-	-
SMR0024 HtdK	23801-24355	184	HCM1.96 (HtdK) R0010 (HtdK)	51.3%/187 (2.6e-34) 51.3%/187 (2.6e-34)	ORF131 (pCAR1)	37%/45 (9.1) (90-134 vs. 1276-1321)	NP_758674
SMR0025	24475-28779	1434	HCM1.97 R0009	83%/1428 (0) 83%/1428 (0)	ORF136 (pCAR1) ORF131 (pCAR1)	27%/421 (e-29) (993-1405 vs. 48-457) 21%/402 (3e-19) (511-910 vs. 1585-1961)	NP_758679 NP_758674
SMR0026 TrhP	29057-29569	170	HCM1.98 (TrhF) R0008 TrhP	77.6%/179 (2.8e-52) 77.6%/179 (2.8e-52)	ORF215 signal peptidase (Rts1) TrsF (SXT) TrhF (R391) PA1303 (Pseudomonas aeruginosa) TrhF (pCAR1) Signal peptidase I (B. cereus)	33%/129 (5e-9) 24%/160 (4e-7) 24%/160 (4e-6) 26%/118 (2e-5) 29%/127 (2e-8) 25%/174 (9e-5)	Np_640175 AAL59670 AAM08054 AAG04692 NP_758683 NP_830922
SMR0027 TrhW	29556-31067	503	HCM1.99 (TrhW) R0007 (TrhW)	80.5%/502 (1.9e-147) 80.5%/502 (1.9e-147)	TrhW (pCAR1) TraW (R391) TraW (SXT) ORF216 (Rts1) TraW (pSLT) TraW (F, R100) Rsph4268 Reut5572 (Ralstonia metallidurans) TraW (pNL1) 20640 (Vibrio vulnificus) 20644 (Vibrio vulnificus)	25%/364 (9e-30) 27%/295 (1e-25) 28%/295 (4e-25) 29%/240 (1e-20) 23%/180 (2e-9) 25%/193 (8e-9) 26%/180 (2e-8) 21%/189 (0.001) 21%/184 (9e-8) 22%/198 (2e-6) 20%/176 (0.047)	NP_758684 AAM08011 AAL59671 BAB93778 AAL23494 AAA24689 ZP_00008319 ZP_00026548 NP_049159 NP_762593 NP_762597
SMR0028 TrhU	31206-32267	353	HCM1.100 (TrhU) R0006 (TrhU)	97%/334 (5.6e-145) 97%/334 (5.6e-145)	TraU (SXT) TraU (R391) ORF217 (Rts1) TraU (pNL1) Rsph4269 TraU (pSLT) TrhU (pCAR1) TraU (F, R100) Reut5571 20641 (Vibrio vulnificus)	36%/347 (1e-57) 36%/345 (2e-56) 31%/359 (5e-44) 33%/340 (1e-40) 30%/326 (1e-39) 31%/346 (2e-38) 32%/356 (3e-38) 32%/320 (1e-35) 27%/351 (7e-26) 30%/301 (8e-40)	AAL59673 AAM08014 BAB93779 AAD03953 ZP_00008320 AAL23495 NP_758685 AAA98088 ZP_00026547 NP_762594
SMR0029 TrhN	32286-35474	1062	HCM1.101 (TrhN) R0005 (TrhN)	87.6%/1062 (0) 87.6%/1062 (0)	ORF24 (Rts1) S005 (S. typhimurium) TraN (SXT)	24%/856 (4e-47) 25%/988 (5e-42) 24%/497 (8-25);	BAB93806 AAK02035 AAL59755

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

					TraN (R391)	29%/232 (9e-17) 24%/501 (3e-24); 28%/345 (6e-18)	AAM07995
					TraN (pNL1)	30%/88 (5e-5)	AAD03951
					TraN (pSLT)	27%/112 (0.008); 23%/134 (0.15)	AAL23498
					TraN (R-100)	25%/112 (0.032)	AAB61941
					Rsph4271	27%/111 (0.043); 24%/193 (0.63)	ZP_00008322
					20645 (Vibrio vulnificus)	32%/67 (5e-4)	NP_762598
SMR0030	c35503-36375	290	HCM1.106c R0004	75.9%/290 (1.2e-89) 75.9%/290 (1.2e-89)	-	-	-
SMR0031 TrhI	36555-38339	594	HCM1.107 (TrhI) R0003 (TrhI)	82.4%/592 (9.3e-199) 82.4%/592 (9.3e-199)	Reut_p_5684	26%/586 (2e-38)	ZP_00026658
					DNA Helicase II (V. cholera)	26%/641 (4e-35)	AAF93366
					DNA Helicase II (S. marcescens)	25%/638 (2e-32)	AAC46278
					UvrD (S. typhimurium LT2)	25%/643 (1e-29)	AAL22795
					DNA Helicase II (S. typhi)	25%/643 (1e-29)	CAD07941
SMR0034	38734-39606	290	HCM1.108 R0002	60.1%/291 (1.2e-60) 60.1%/291 (1.2e-60)	-	-	-
SMR0035	39664-40041	125	HCM1.109 R0001	87.2%/125 (3e-41) 87.2%/125 (3e-41)	-	-	-
SMR0036	40102-41079	325	HCM1.110 R0210	56.9%/327 (5.1e-71) 56.9%/327 (5.1e-71)	-	-	-
SMR0037	41137-42195	352	HCM1.111 R0209	75.7%/350 (8.2e-99) 75.7%/350 (8.2e-99)	s067 (SXT)	30%/276 (8e-24)	AAL59708
					ORF70 (R391)	29%/275 (1e-23)	AAM08012
					ORF101 (pCAR1)	29%/238 (6e-17)	NP758644
					Ddes0370	31%/132 (2e-10)	ZP_00128764
					ORF227 (Rts1)	27%/224 (3e-8)	BAB93789
					Rsph2275	28%/179 (7e-8)	ZP_00006359
					Reut163	25%/159 (3e-6)	ZP_00021263
					NorQ (Paracoccus denitrificans)	27%/153 (1e-6)	AAA68972
					NorQ (Rhodobacter sphaeroides)	28%/175 (1e-5)	AAC45373
SMR0038	42327-42500	57	HCM1.112 R0208	52.7%/55 (1.2e-7) 52.7%/55 (1.2e-7)	-	-	-
SMR0039	42564-43691	375	HCM1.114 R0207	79.1%/378 (1.4e-106) 79.1%/378 (1.4e-106)	s068 (SXT)	18%/234 (5e-4)	AAL59707
					ORF72 (R391)	18%/234 (0.001)	AAM08016
					Reut_p_5247	15%/361 (0.004)	ZP_00026235
					ORF228 (Rts1)	21%/251 (0.043)	BAB93790
					Ddes0369	21%/209 (0.71)	ZP_00128763
					ORF102 (pCAR1)	19%/233 (0.28)	NP_758645
SMR0040	43769-45775	668	HCM1.116 R0206	62.6%/672 (3.5e-139) 62.6%/672 (3.5e-139)	ORF74 (R391)	22%/631	AAM08006
					ORF103 (pCAR1)	20%/537 (9e-11)	NP_758646
					sO70 (SXT)	22%/330	AAL59705
					CobT (Y. pestis pMT1)	21%/301	AAC13217
					pHCM2.0017 (pHCM2)	18%706	CAD09884
					ORF233 (Rts1)	21%/233	BAB93795
					NorD (R. sphaeroides)	20%/194	AAC45374
					NorD (Paracoccus denitrificans)	18%/193	AAA68973
					Saro_p_2284 (N. aromaticivorans)	17%/597	ZP_00095256
					vWF type A domain (B. anthracis)	14%/574	BA_1206
SMR0041	45848-46069	73	HCM1.116a R0205	65.8%/73 (5e-18) 65.8%/73 (5e-18)	-	-	-

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

SMR0042	46183-47415	410	HCM1.117 R0204	78.5%/410 (9.6e-140) 78.5%/410 (9.6e-140)	ORF78 (R391) s072 (SXT) ORF146 (pCAR1) Reut_p_5650 Magn_p_1471 (M. magnetotactium) ID498 (B. japonicum) Sma0911 (pSymA, S. meliloti) Vir. assoc. protein E (B.melitensis) DNA primase STM0901 Saro_p_3633 P4 DNA primase (STY4832) ORF229 (pNL1)	30%/368 (5e-27) 30%/372 (1e-25) 29%/370 (4e-19) 24%/314 (7e-9) 25%/258 (2e-8) 32%/187 (2e-7) 29%/177 (2e-6) 28%/224 (3e-5) 23%/299 24%/180 17%/287 19%/178	AAM08013 AAL59703 NP_758689 ZP_00026624 ZP_00048890 AAG60914 AAK65155 AAL52878 AAL19837 ZP_00096593 CAD06954 AAD03876
SMR0043	47685-48206	173	HCM1.118 R0203	69%/171 (2.9e-50) 69%/171 (2.9e-50)	-	-	-
SMR0044	48285-49163	292	HCM1.119 R0202	51.45/292 (3.2e-59) 51.45/292 (3.2e-59)	-	-	-
SMR0045	49233-50168	311	HCM1.120 R0201	52.4%/311 (9.5e-66) 52.4%/311 (9.5e-66)	-	-	-
SMR0046	50237-51157	306	HCM1.121 R0200 HCM1.120 R0201	55.3%/300 (1.8e-59) 55.3%/300 (1.8e-59) 23%/95 (0.73) 23%/95 (0.73)	-	-	-
SMR0047	51224-52162	312	HCM1.122 R0199 HCM1.123 R0198 R0194	47.4%/310 (5.1e-52) 47.4%/310 (5.1e-52) 26%/308 (8e-23) 27%/181 (e-8) 28%/125 (3e-7)	-	-	-
SMR0048	52328-53290	320	R0198 R0194 R0199 HCM1.123 HCM1.122	66.6%/183 (8.6e-47) 53%/137 (3e-31) 26%/311 (3e-25) 61%/318 (1e-102) 27%/311 (2e-26)	-	-	-
SMR0049	53524-53829	101	HCM1.124 R0193	50%/80 (8.3e-13) 50%/80 (8.3e-13)	-	-	-
SMR0050	53874-54464	196	R0192 HCM1.125	62%/156 (6e-50) 62.8%/196 (1.5e-47)	-	-	-
SMR0051	54472-54729	85	HCM1.126 R0191	64.7%/85 (6.3e-19) 64.7%/85 (6.3e-19)	-	-	-
SMR0052	54803-55339	178	HCM1.127 R0190	57.9%/178 (2.1e-46) 57.9%/178 (2.1e-46)	-	-	-
SMR0053	55356-55802	148	HCM1.128 R0189	57%/135 (e-34) 58%/135 (3e-35)	-	-	-
SMR0055 Tus	55893-56855	320	HCM1.131 R0186	60.0%/316 (1.9e-71) 60.0%/316 (1.9e-71)	Tus (ORF15, Rts1) Tus (E. coli O157:H7) **1ECR** Tus (S. flexneri) Tus (S. typhimurium LT2) STY1652 (S. typhi CT18) Tus (Y. pestis)	27%/154 (2e-4) 18%/149 18%/149 18%/190 17%/190 17%/174	BAB93578 NP_288045 AAN43216 AAL20390 CAD01897 AAF68950
SMR0056	56865-57770	301	HCM1.132 R0185	78.5%/302 (2.3e-101) 78.5%/302 (2.3e-101)	-	-	-
SMR0057	57784-58050	88	HCM1.133	73.9%/88 (5.2e-27)	-	-	-

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

			R0184	73.9%/88 (5.2e-27)			
SMR0058	c58072-59232	386	HCM1.134c R0183	64.8%/386 (1.7e-101) 64.8%/386 (1.7e-101)	CRE recombinase	27%/368 (4e-19)	P06956
SMR0059				Doubtful CDS	-	-	-
SMR0060 Hha	59417-59629	70	R0182 (Hha) HCM1.135	91%/61 (2e-23) 90%/70 (7e-27)	Hha (E. coli) Haemolysin expression modulation protein STY0516 YdfA (R721) Hha (S. flexneri) Haem. modul. (E. coli O157:H7) Hha/YmoA (Y. pestis) Hmo (S. flexneri, pWR501) Haem. modul. (L7004, pO157) STY1661 (S. typhi CT18)	50%/66 (9e-11) 50%.72 (6e-10) 52%/67 (9e-10) 50%/66 (9e-10) 50%/66 (9e-10) 58%/66 (1e-8) 42%/69 (5e-8) 42%/69 (2e-7) 44%/72 (5e-6)	S15270 CAD04958 BAB12630 AAN42060 NP_286201 NP_668376 AAK18587 AAC70072 CAD01906
SMR0061	59990-61072	360	HCM1.141 R0175	73.9%/360 (1.7e-110) 73.9%/360 (1.7e-110)	-	-	-
SMR0062	c61238-62737	499	-	-	putative protein kinase (y3587, Y. pestis) Hso1308 (Haemphilus somnus) Hr0976 (Thermosynechococcus elongatus) Npun_p_3785 (Nostoc punctiforme) Tery_p_3358 (Trichodesmium erythraecum)	86%/499 (0) 59%/496 (0) 30%/416 (1e-109) 23%/372 (7e-84) 20%/375 (5e-62)	NP_670884 ZP_00123302 NP_681766 ZP_00109337 ZP_0074073
SMR0063	c62763-64403	546	-	-	y3586 (Y.pestis) Hsom1307 (H. somnus) Tery_p_1177 (T. erythraecum) alr2823 (Nostoc sp.) ECs2880 (E. coli O157:H7) SF2136 (S. flexneri) tll0386 (T. elongatus)	47%/559 (1e-139) 36%/357 (1e-43) 26%/267 (0.057) 28%/213 (0.44) 23%/216 24%/216 29%/88 (1.5)	NP_670883 ZP_00123301 ZP_00072001 NP_486863 NP_310907 NP_707964 NP_681176
SMR0064 TerY3	c64400-65440	346	-	-	y3585 (Y. pestis) Hsom1306 (H.somnus) y3581 (Y. pestis) Hsom1304 (H. somnus) TerY (R478) y3584 (Y. pestis) Hsom1305 (H. somnus) tlr0974 (T. elongatus) Ecs2881 (E. coli O157:H7) Tery_p_3358 (T. erythraecum) alr2822 (Nostoc sp.) Npun_p_3783 (N. punctiforme) YegL (S. flexneri)	79%/286 (1e-140) 43%/351 (8e-73) 40%/196 (1e-35) 41%/178 (2e-32) 39%/181 (4e-29) 36%/185 (6e-25) 30%/199 (1e-23) 32%/200 (2e-19) 28%/157 (9e-11) 22%/313 (3e-10) 26%/199 (2e-9) 26%/199 (6e-9) 25%/144 (5e-6)	NP_670882 ZP_00123300 NP_670878 ZP_00123298 AAL44737 NP_670881 ZP_00123299 NP_681764 NP_310908 ZP_00074073 NP_486862 ZP_00109335 NP_707965
SMR0065 TerY2	c65525-66163	212	-	-	y3584 (Y. pestis) Hsom1305 (H. somnus) Hsom1304 (H. somnus) y3581 (Y. pestis) TerY (R478) y3585 (Y. pestis) Hsom1306 (H.somnus) tlr0974 (T. elongatus)	82%/212 (4e-96) 51%/211 (2e-59) 50%/216 (5e-50) 49%/190 (2e-45) 46%/202 (3e-37) 39%/194 (2e-31) 36%/197 (1e-23) 30%/223 (2e-18)	NP_670881 ZP_00123299 ZP_00123298 NP_670878 AAL44737 NP_670882 ZP_00123300 NP_681764

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

					Tery_p_3358 (T. erythraecum) alr2822 (Nostoc sp.) Tery_p_1176 Npun_p_3783 (N. punctiforme) Ecs2881 (E. coli O157:H7) Chlor_p_2225 (Chloroflexus aurantiacus) YegL (S. flexneri) YvcC (Lactococcus lactis)	30%/198 (7e-15) 29%/199 (3e-11) 29%/213 (8e-11) 29%/199 (2e-10) 35%/121 (7e-10) 31%/98 (1e-4) 32%/121 (2e-4) 26%/156 (7e-4)	ZP_00074073 NP_486862 ZP_00072000 ZP_00109335 NP_310908 ZP_00019225 NP_707965 NP_268200
SMR0066 TerX	c6616-66804	213	-	-	y3583 (Y. pestis) many TerD and TerE homologues	86%/213 (1e-104)	NP_670879
SMR0067 TerY1	c66827- 67420	197	-	-	y3581 (Y. pestis) - also many from TerY2/3 including: Chlo_p_2225 (C. aurantiacus)	85%/197 (2e-81) 30%/168 (1e-5)	NP_670878 ZP_00019225
SMR0068 TerW	c67928-68395	155	-	-	TerW (ECs1343, E. coli O157:H7) y0549 (Y. pestis)	99%/155 (9e-6) 51%/153 (2e-28)	NP_309370 NP_667886
SMR0069	c68413-69621	402	-	-	ECs1344 (E. coli O157:H7)	98%/402 (0)	NP_309371
SMR0070	c69632-70588	318	-	-	ECs1345 (E. coli O157:H7) y0550 (Y. pestis) Psyr4495 (Pseudomonas syringae) Rrub_p_1529 (Rhodospirillum rubrum) DR2219 (Deinococcus radiodurans) SCC8A.21C (Streptomyces coelicolor) Rsph_p_3499 (R. sphaeroides)	98%/318 (1e-179) 65%/310 (1e-115) 52%/298 (3e-72) 41%313 (4e-54) 43%/300 (1e-53) 33%/353 (1e-31) 20%/305	NP_309372 NP_667887 ZP_00128167 ZP_00014514 NP_295941 NP_626610 ZP_00007562
SMR0071	c70588-71667	359	-	-	Z1606 (E. coli O157:H7) y0551 (Y. pestis) Ecs1346 (E. coli O157:H7) Psyr4494 (P. syringae) yp063 (Rhizobium etli, p42d) Rrub_p_1528 (R. rubrum) ECs1347 (E. coli O157:H7) SCC8A.20C (S. coelicolor) DR2218 (D. radiodurans)	98%/359 (0) 70%/353 (1e-147) 98%/224 (1e-136) 51%/352 (2e-97) 44%/356 (2e-75) 44%/356 (7e-63) 98%/106 (2e-54) 36%/346 (9e-44) 35%/335 (2e-33)	NP_287110 NP_667888 NP_309373 ZP_00128166 NP_659965 ZP_00014513 NP_309374 NP_626609 NP_295940
SMR0072	c71669-72442	257	-	-	ECs1348 (E. coli O157:H7) y0552 (Y. pestis) yp064 (Rhizobium etli, p42d) Psyr4493 (P. syringae) DR2215 (D. radiodurans) Rrub_p_1527 (R. rubrum) SCCA8.19C (S. coelicolor)	**28%**/257 (1e-143) 66%/254 (1e-101) 41%/252 (5e-39) 35%/247 (6e-34) 37%/255 (1e-26) 33%/240 (2e-25) 21%/247	NP_309375 NP_667889 NP_659966 ZP_00128165 NP_295937 ZP_00014512 NP_626608
SMR0073	c72435-73577	380	-	-	ECs1349 (E. coli O157:H7) y0553 (Y. pestis) Psyr4492 (P. syringae) yp062 (Rhizobium etli, p42d) ORF3 (Proteus mirabilis) DR2214 (D. radiodurans) Rrub_p_1526 (R. rubrum) SCCA8.20c (S. coelicolor) Rsph_p_2082 (R. sphaeroides)	98%/380 (0) 68%/372 (1e-145) 53%/361 (1e-103) 44%/371 (3e-78) 67%/198 (2e-76) 39%/368 (2e-53) 42%/216 (4e-34) 31%/349 (3e-16) 28%/77	NP_39376 NP_667890 ZP_00128164 NP_659964 AAD47283 NP_295936 ZP_00014511 NP_626609 ZP_00006173
SMR0074	c73587-74576	329	-	-	ECs1350 (E. coli O157:H7)	98%/327 (0)	NP_309377

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

					y0554 (Y. pestis) Psyr4490 (P. syringae) Rrub_p_1531 (R. rubrum) DR2213 (D. radiodurans) SCA7.21c (S. coelicolor) Npun_p_5173 (N. punctiforme) Reut_p_4570 (R. metallidurans) Magn_p_2486 (M. magnetotactium) SMb21227 carbamoyl phosphate synthase (S. meliloti, pSymB)	54%/320 (1e-98) 40%/285 (1e-54) 29%/287 (7e-23) 28%/292 (3e-14) 28%/288 (5e-14) 20%/252 (0.022) 26%/171 (0.055) 22%/181 (1.2) 21%/238	NP_667891 ZP_00128162 ZP_00014516 NP_295935 NP_630746 ZP_00110708 ZP_00025569 ZP_00049855 NP_437271
SMR0075 TerZ	74967-75548	193	-	-	TerZ ECs1351 (E. coli O157:H7) TerZ (Y. pestis) TerZ (P. mirabilis) Hsom1311 (H. somnus) Psyr4497 (P. syringae) Rrub_p_1534 (R. rubrum) TerZ DR2224 (D. radiodurans) SCC54.25c (S. coelicolor) Npun_p_5790 (N. punctiforme) TerD BA_0974 (B. anthracis) YceD (B. subtilus) Tery_p_2097 (T. erythraecum) Rrub_p_1533 (R. rubrum) TerD (Y. pestis) Hsom1310 (H. somnus) cAMP binding protein (Dictyostelium discoideum)	98%/193 (1e-95) 74%/197 (1e-77) 73%/194 (2e-72) 51%/187 (1e-48) 51%/193 (2e-48) 54%/184 (2e-46) 50%/188 (9e-44) 41%/186 (4e-31) 44%/184 (2e-30) 37%/193 (4e-30) 42%/185 (5e-30) 39%/183 (7e-28) 40%/185 (6e-27) 40%/185 (5e-26) 38%/180 (9e-26) 38%/188 (6e-25)	NP_309378 NP_667892 AAD47284 ZP_00123305 ZP_00128169 ZP_00014519 NP_295946 NP_626229 ZP_00111320 NP_654334 NP_388172 ZP_00072867 ZP_00014518 NP_667896 ZP_00123304 A49752
SMR0076 TerA	75548-76705	385	-	-	TerA (E. coli O157:H7) TerA (Ecs1352, E. coli O157:H7) TerA (Y. pestis) TerA (pMJ606, Alcaligenes sp.) TerA (P. mirabilis) Psyr4498 (P. syringae) Rrub_p_1535 (R. rubrum) TerA DR2223 (D. radiodurans) TerA (pMER610, IncHI2) TerF (pTE53) SCC8A.23c (S. coelicolor) Npun_p_5789 (N. punctiforme)	99%/385 (0) 99%/341 (0) 68%/394 (1e-152) 75%/343 (1e-148) 65%/389 (1e-142) 49%/394 (6e-95) 38%/415 (3e-66) 37%/403 (7e-52) 74%/128 (2e-51) 44%/211 (5e-33) 36%/219 (3e-27) 36%/193 (4e-24)	NP_286707 NP_309379 NP_667893 AAA98289 AAD47285 ZP_00128170 ZP_00014520 NP_29545 2009362A CAB43000 NP_626612 ZP_00111319
SMR0077 TerB	76728-77183	151	-	-	TerB (Ecs1353, E. coli O157:H7) TerB (pTE53) TerB (P. mirabilis) TerB (Y. pestis) TerB (pMJ606, Alcaligenes sp.) Avin_p_141 (Azotobacter vinelandii) Rrub_p_1536 (R. rubrum) Psyr4499 (P. syringae) TerB DR2220 (D. radiodurans) SCH63.14 (S. coelicolor)	100%/151 99%/151 (5e-71) 81%/151 (8e-58) 71%/151 (1e-52) 92%/118 (2e-51) 43%/147 (5e-31) 43%/151 (1e-29) 42%/147 (2e-26) 325/151 (6e-15) 26%/151 (3e-10)	NP_309380 CAB42996 AAD47286 NP_667894 AAA98290 ZP_00088494 ZP_00014521 ZP_00128171 NP_295942 NP_627957
SMR0078 TerC	77206-78246	346	-	-	TerC (Ecs1354, E. coli O157:H7) TerC (pMJ606, Alcaligenes sp.) TerC (P. mirabilis)	100%/346 90%/346 (1e-158) 88%/338 (1e-152)	NP_309381 AAA98291 AAD47287

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

					TerC (Y. pestis) Psyr4500 (P. syringae) Rrub_p_1537 (R. rubrum) Hsom1302 (H. somnus) Chut_p_754 (Cytophaga hutchinsonii) Au1861 (Agrobacterium tumefaciens) TerC CC3682 (Caulobacter crescentus) Rrub_p_240 (R. rubrum) Gmet_p_2691 (Geobacter metallireducens) PA2549 (P. aeruginosa) VCA0546 (V. cholera) Chlo_p_3802 (C. aurantiacus) NMA0259 (Neisseria meningitidis) Magn_p_4803 (M. magnetotactium) Avin_p_1079 (A. vinelandii) lots more.....	83%/313 (1e-133) 69%/344 (1e-120) 53%/340 (8e-85) 51%/337 (2e-81) 37%/326 (2e-45) 34%/335 (5e-42) 36%/325 (2e-40) 35%/334 (5e-40) 34%/321 (9e-40) 34%/339 (1e-39) 34%/333 (6e-39) 34%/322 (4e-38) 34%/336 (4e-37) 33/%/304 (2e-36) 35%/295 (2e-36)	NP_667895 ZP_00128172 ZP_00014522 ZP_00123296 ZP_00117385 NP_332541 NP_422476 ZP_00013246 ZP_00081928 NP_251239 NP_232936 ZP_00020757 NP_283099 ZP_00052083 ZP_00089409
SMR0079 TerD	78295-78873	192	-	-	TerD (E. coli O157:H7) TerD (pMJ606) TerD (P. mirabilis) TerD (Y. pestis) Psyr4501 (P. syringae) Hsom1309 * and many more like above	100%/192 (8e-97) 95%/192 (6e-93) 92%/192 (e-92) 86%/192 (4e-88) 68%/190 (3e-69) 67%/191 (3e-68)	NP_309382 AAA98292 AAD47288 NP_667896 ZP_00128173 ZP_00123303
SMR0080 TerE	78942-79517	191	-	-	TerE (E. coli O157:H7) TerE (P. mirabilis) Psyr4502 (P. syringae) Hsom1310 * same conservations as above	98%/191 (3e-95) 88%/191 (7e-89) 69%/191 (2e-65) 61%/191 (7e-64)	NP_309383 AAD47289 ZP_00128174 ZP_00123304
SMR0081 TerF	79946-81187	413	-	-	PSPO0946 YcbR (B. subtilis) TerF (E. coli O157:H7) TerF (D. radiodurans) TerD (D. radiodurans) TerA (P. mirabilis)	43%/423 (e-81) 32%/221 (6e-28) 94%/102 (9e-49) 39%/197 (4e-23) 37%/189 (2e-21) 34%/215 (5e-25)	NP_790785 NP_388143 NP_309385 NP_285380 NP_295943 AAD47285
SMR0082	81278-81733	151	-	-	-	-	-
SMR0083	c81974-82165	63	-	-	-	-	-
SMR0084	c82257-82598	113	-	-	-	-	-
SMR0086	83212-83397	61	-	-	Ddes0780 HsdM (X. axonoposis pv. citri)	85%/27 (3e-7) 81%/16 (0.26)	ZP_00129169 NP_643209
SMR0087	83584-83838	84	HCM1.104c	100%/84 (2.4e-30)			
SMR0088	83841-85880	679	HCM1.103	99.3%/679 (0)	ATP-dependent DNA Helicase of many others	~30% over whole or first half	
SMR0089	c85877-86917	346	HCM1.102	99.1%/328 (1.1e-138)			
SMR0090	87745-88176	143	R0172 HCM1.145	95.4%/130 (2e-52) 94.6%/130 (3.7e-52)			
SMR0091	88185-88466	93	R0171	72.2%/90 (4e-25)			
SMR0092	88835-89491	218	-	-	Mmc10909 (Magnetococcus sp.)	32%/103 (0.87)	ZP_00043194
SMR0093	89694-90191	165	-	-	CPE1290 (Clostridium perfringens)	43%/71 (0.073)	NP_562206
SMR0094	90196- 91584	462	-	-	CPE1289 (C. perfingens)	24%/466 (5e-16)	NP_562205

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

					Cj0569 (Campylobacter jejuni) Cj0568 (C. jejuni)	29%/224 (2e-15) 33%/75 (0.74)	NP_281752 NP_281751
SMR0095	91753-92469	238	HCM1.213c HCM1.162c Transposase for IS26	100%/234 (e-126) 65%/233 (2e-77)	TnIS26 (S. typhimurium pSEM) TnpIS15 (R391) ORF165 (Rts1) TnpA (IncN R46) Reut0471 (R. mettalidurans)	99%/238 (e-128) 99%/238 (e-128) 100%/234 (e-126) 100%/234 (1e-126) ?/? (1e-36)	CAB71034 CAD27661 NP_640126 NP_511230 ZP_00021558
SMR0096	92524-92841	105	-	-	Integrase (pseudogene) – 100% to C-term of numerous IntI1 integrases (including from In2 –R100/pNR)		
SMR0096a	92810-93049	80	-	-	Transposon fragment (pseudogene)		
SMR0096b	93194-93340	49	-	-	Recominase/resolvase (pseudogene)		
SMR0097	93392-94108	238	= SMR0095 100%				
SMR0098	94161-95099	312	-	-	Class III aminotransferase (pseudogene) (PDB avail.)		
SMR0099	95096-95647	183					
SMR0100	95780-96532	250	-	-	Glutamine binding/transport (PDB)		
SMR0101	96585-97241	218	-	-	Glutamine permeases		
SMR0102	97244-97330	29	-	-	fragment of Glu ATP-bind component		
SMR0103	c97628-98344	238	= SMR0095/97				
SMR0104 AdhC	c98486-99604	372	-	-	Alcohol dehydrogenase class III (PDB) STM1627 XAC0734 AdhC (E. coli O157:H7) ORF10 (Rts1)	 94%/371 (0) 88%/369 (0) 81%/369 (1e-171) 80%/369 (1e-169)	 NP_460586 NP_641287 NP_286099 NP_639971
SMR0105	c99650-99925	91	-	-	STM1628 ORF9 (Rts1) YaiN (E. coli O157:H7) XAC0733 XAC04059 STY3168 Rsph_p_2551 Saro0778	87%/91 (1e-32) 57%/91 (2e-19) 57%/91 (2e-18) 53%/91 (5e-15) 49%/91 (4e-12) 40%/91 (3e-10) 43%/91 (9e-10) 35%/91 (8e-8)	NP_460587 NP_639970 NP_286100 NP_641086 NP_644358 NP_457418 ZP_0006630 ZP_00093770
SMR0106	100130-100459	109	-	-	-	-	-
SMR0107	100620-101114	164	HCM1.207	99%/125 (2e-67)	ybjA (R100) SMb21092 Bcep5068 XCC0457 XAC0474	99%/125 (2e-67) 31%/108 (8e-10) 30%/104 (4e-8) 31%/101 (1e-7) 36%/79 (1e-7)	NP_052902 NP_437211.1 ZP_00032220 NP_635851.1 NP_640829.1
SMR0108 Cat	101315-101974	219	HCM1.206	99%/219 (1e-120)	cat (R100) and MANY other cat	99%/219 (1e-120)	NP_0529031
SMR0109	c102196-102912	238	= SMR95/97/103				
SMR0110 AphA	c103018-103833	271	HCM1.223 StrA	27%/308	Identical to numerous aminoglycoside phosphotransferases		
SMR0112	103972-104688	238	=SMR95/97/103/ 109				
SMR0113	104743-105249	168			TnpA (psedudogene via IS26 insert.)		
SMR0114	105304-105567	88	HCM1.162c	100%/88 (2.7e-34)	TnpA fragment (pseduogene)		
SMR0115	c105627-105671	15	-	-	TniA transposase fragment from IS21 (pseudogene)	100%/15	

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

SMR0115a	c105710-105775	21	HCM1.227c	100%/21	pseudogene		
SMR0116	c105808-106020	70	HCM1.161	100%/70 (4e-25)	Urf2Y (Mercury resist. bacteria)	100%/70 (4e-25)	AAD34401
					Reut5580	68%/70 (3e-20)	ZP_00026556
					TnpM (Tn21 modulator, R100)	71%/69 (1e-26)	NP_052899.1
SMR0117	c106086-106322	78	HCM1.160	98%/78 (7e-32)	Urf-1 (MerE, Alcaligenes sp.)	100%/78 (3e-32)	CAA70192.1
MerE			HCM1.229c	86%/78 (4e-26)	Mer operon ORF1 (pDU1358)	91%/78 (3e-30)	D29010
					Reut5579	83%/77 (5e-27)	ZP_00026555
					Urf1 (R100)	86%/78 (4e-26)	NP_052887.1
					MerE (pSB102)	75%/78 (3e-23)	NP_361076.1
SMR0118	c106319-106705	128	HCM1.159	100%/121 (3e-49)	MerD (Alcaligenes sp.)	99%/121 (2e-48)	CAA70191.1
MerD			HCM1.230c	82%/121 (5e-32)	MerD (pSB102)	88%/126 (1e-47)	NP_361075.1
			HCM1.230c (MerR)	29%/105 (.005)	MerD (pDU1358)	88%/121 (2e-43)	C29010
					Reut5578	89%/128 (3e-43)	ZP_00026554
					MerD (R100)	82%/121 (5e-32)	NP_052886.1
					Rsph1730	35%/105 (.002)	ZP_00005826
					Rsph3193	33%/108 (.002)	ZP_00007262
SMR0119	c106702-108387	561	HCM1.158	100%/561 (0)	MerA (Klebsiella pneumoniae; 12604550)	100%/561 (0)	AAM44226
MerA			HCM1.231c	84%/565 (0)	MerA (pKLH256 Enterobacter; 9159519)	99%/561 (0)	CAA70231.1
					Reut5577	93%/561 (0)	ZP_00026553
					MerA (Tn501)	93%/561 (0)	NP_085424
					MerA (pSB102)	89%/561 (0)	NP_361073
					MerA (Alcaligenes sp.)	88%/558 (0)	CAA70190.1
					MerA (R100)	84%/565 (0)	NP_052885.1
					MerA (Tn21)	84%/565 (0)	AAC33905.1
					Reut4216	79%/559 (0)	ZP_00025223
					MerA (pKLH201); 9159519	100%/366 (0)	CAA71183
					MerA (R391)	56%/364 (1e-163)	AAM08005
					Mer red fragment (pDU1358); 2666393	90%/343 (1e-162)	D33858
					Serratia marcescens Mer. reductatase	90%/342 (1e-161)	P08662_1
SMR0120	c108426-108851	141	HCM1.157	100%/141 (4e-63)	Mer C (Klebsiella pneumoniae)	100%/141 (4e-63)	AAM44225
MerC			HCM1.232c	72%/140 (7e-39)	MerC (pKLH256)	100%/141 (4e-63)	CAA70230.1
					MerC (pSB102)	75%/141 (9e-40)	NP_361072
					MerC (R100)	74%/140 (5e-40)	NP_052884
					MerC (Alcalgenes sp.); 9159519	68%/141 (4e-39)	CAA70189.1
					MerC (R391)	59%/132 (6e-34)	AAM08059.1
SMR0121	c108879-109154	91	HCM1.153	100%/91 (3e-26)	MerP (Klebiella pneumoniae)	100%/91 (3e-26)	AAM44224
MerP			HCM1.233c	82%/91 (7e-21)	MerP (pKLH256)	100%/91 (3e-26)	CAA70229.1
					MerP (pDU1358)	83%/91 (4e-22)	AAA98223
					MerP (R100; PDB:1AFJ)	83%/91 (5e-21)	NP_052883
					Reut5576	85%/91 (8e-21)	ZP_00026552
					MerP (RP1 Tn5053; 8594337)	84%/90 (8e-20)	S70144
					MerP (Alcaligenes sp.)	85%/91 (1e-19)	CAA70197.1
					MerP (pSB102)	82%/91 (7e-19)	NP_361070
					Reut4215	67%/95 (1e-14)	ZP_00025222
					MerP (R391)	54%/90 (1e-13)	AAM08084.1
					Reut3236	50%/60 (5e-7)	ZP_00024264
SMR0122	c109170-109535	121	HCM1.152	100%/121 (2e-53)	MerT (pKLH210)	100%120 (7e-55)	CAA70239.1
MerT			HCM1.234c	93%/115 (2e-47)	MerT (Klebsiella pneumoniae)	100%120 (7e-55)	AAM44223
					MerT (pKLH256)	100%120 (3e-54)	CAA70228
					MerT (Alcaligenes sp.)	99%/115 (5e-49)	CAA70196
					MerT (pSB102)	95%/115 (4e-48)	NP_361069
					Reut5575	89%/119 (1e-47)	ZP_00026551

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

					MerT (R100) MertT (pDU1358) Reut4214 Reut3237 MerT (R391)	93%/115 (2e-47) 96%/115 (2e-45) 81%/115 (3e-38) 55%/110 (2e-23) 52%/100 (8e-20)	NP_052882.1 AAA98222.1 ZP_00025221 ZP_00024265 AAM08065.1
SMR0123 MerR	109607-110062	151	HCM1.151 HCM1.225	99%/151 (1e-79) 90%/142 (5e-65)	MerR (Alcaligenes sp.) MerR (pKLH201) MerR (pKLH256) MerR (Klebsiella pneumoniae) Tn501 repressor; 11292750 Reut0554 Reut5574 MerR (pSB102) MerR (Tn5053) MerR (R100) MerR (pDU1358) Reut4213 Reut3238 MerR (R391) SMa0281 (pSymA) Rsph3193 MerR (Exiguobacterium sp. pKLH3; 9534232)	100%/151 (4e-80) 100%/151 (4e-80) 100%/151 (4e-80) 99%/151 (8e-80) 92%/142 (2e-68) 92%/142 (2e-67) 92%/142 (2e-67) 89%/142 (5e-67) 90%/142 (4e-65) 90%/142 (5e-65) 92%/133 (7e-64) 74%/127 (4e-48) 52%/122 (8e-28) 45%/124 (2e-25) 46%/128 (3e-21) 39%/128 (3e-20) 38%/125 (1e-17)	CAA70195 CAA70240.1 CAA70237.1 AAM44222.1 NP_085422.1 ZP_00021641 ZP_00026550 NP_361068.1 S32798 NP_052881.1 AAA98221.1 ZP_00025200 ZP_00024266 AAM08061.1 NP_435397 ZP_00007262 CAA67818
SMR0124	110128-110736	202	HCM1.177 HCM1.175 R0166 R0168	53%/194 (2e-53) 36%/99 (4e-4) 59%/118 (1e-35) 36%/99 (4e-4)	ORF5 (E. coli O6:K15:H31; 12379716) Reut5304	29%/143 (2e-7) 33%/77 (0.056)	CAD33715.1 ZP_00026287
SMR0125	110798-111001	67	-	-	-	-	-
SMR0126 H-NS	c111343-111747	134	HCM1.178ac R0164 (H-NS)	81%/134 (1e-37) 81%/134 (2e-37)	Sfh (Shigella) H-NS (Erwinia) StpA (E. coli) H-NS (E. coli O157:H7, 1HNR, 1LR1_A) StpA (S. Typhi) H-NS (Yersinia) H-NS (S. Typhi) Reut0246 KorB (IncN R46 regulator) CR1 (Citrobacter LEE regul., 11553577)	79%/134 (1e-37) 60%/135 (2e-33) 62%/134 (8e-33) 62%/134 (8e-33) 59%/134 (5e-30) 60%/136 (3e-29) 63%/138 (5e-26) 34%/103 (8e-5) 43%/53 (0.003) 20%/123	AAN38340 CAC44357 CAA49146 NP_289219 NP_457198 CAC24849 NP_455749.1 ZP_00021345 NP_511185.1 AAL06349
SMR0127	c111925-112284	119	-	-	-	-	-
SMR0128	c112521-112976	151	HCM1.176 R0167	24%/131 24%/131	ORF21 (Citrobacter freundii pCTX-M3, broad host-range conju. plasmid)	31%/135 (3e-10)	NP_774991.1
SMR0129	c113036-113701	221	HCM1.180c R0161	57%/218 (9e-68) 57%/218 (9e-68)	ORF7 (pCTX-M3) ORF163 (Rts1)	40%/210 (1e-37) 36%/214 (2e-29)	NP_774973.1 NP_640124.1
SMR0130	c113759-114139	126	-	-	-	-	-
SMR0132	114891-115733	280	-	-	Tn7 TnsA (1400257, 2156235; 1F1Z_A, 1F1Z_B)	39%/192 (4e-27) (last 80 aa don't match)	NP_065320.1
SMR0133	115765-117843	692	-	-	Chte1624 (Clostridium thermocellum) Npun0417 Tn7 TnsB Reut3107 Npun4141	28%/667 (2e-56) 25%/669 (2e-45) 25%/695 (2e-42) 27%/459 (3e-29) 33%/313 (7e-37)	ZP_00061216 ZP_00106048 NP_065319.1 ZP_00024138 ZP_00109790
SMR0134	117843-119291	482	-	-	Chte1623	34%/373 (1e-51)	ZP_00061215

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

					Npun4242 Tn7 TnsC	31%/377 (3e-51) 30%/383 (1e-33)	ZP_00109791 NP_065318.1
SMR0135	119323-120888	521	-	-	Npun4245 Chte1622 Tn7 TnsD	28%/177 (9e-8) 28%/110 (2e-4) 23%/156 (0.095); 21%/206 (0.21)	ZP_00109794 ZP_00061214 NP_065317.1
SMR0136	120900-121826	308	-	-	-	-	-
SMR0138	122179-122478	99	-	-	HipB (R. soanacearum megaplasmid) Bcep7822 TM1330 (Thermotoga maritime)	44%/81 (4e-8) 36%/68 (2e-4) 26%/80 (7e-4)	NP_522813.1 ZP_00034923 NP_229132.1
SMR0139	123042-124868	608	-	-	ORF024 (Pediococcus pMD136) ORF04 (pMD136) ORFT (ICESt1) S024 (S. typhimurium) Atu5023 (pAT) SMC (R. sphaeroides) msi092 (M. loti symbiosis island, 12003951)	36%/556 (4e-86) 40%//213 (2e-34) 21%/566 (2e-8) 29%/173 (1e-7) 25%/398 (3e-7) 47%/44 (7e-4) 26%/209(0.002); 30%/85 (0.023)	NP_037556.1 NP_052600.1 CAC67530.1 AAK02041 NP_535400 CAD66603 CAD31497.1
SMR0140	c125037-125387	116	-	-	mlr5326 (M. loti) Bcep2139 SMa1024 (pSymA) SMc00921 Magn4014 Chlo2357	45%/114 (1e-13) 41%/108 (8e-11) 45%/91 (8e-10) 47%/80 (2e-9) 38%/88 (1e-8) 31%/105 (6e-5)	NP_106012.1 ZP_00029342 NP_435800.1 NP_384890.1 ZP_00051328 ZP_00019356
SMR0141 SilE	c125535-125966	143	-	-	SilE (S. typhimurium, pMG101) SilE (R476b) 11739772 SilE (pMIP233) SilE (pMIP235) SilE (pWR23) PcoE (pRJ1004, cop bind., 8594334) STY0609 (copper binding)	100%/143 (2e-54) 100%/127 (2e-48) 100%/127 (2e-48) 100%/127 (2e-48) 92%/127 (1e-44) 47%/142 (1e-16) 35%/82 (0.014)	AAD11743.1 AAC68931.1 AAC68941.1 AAC68944.1 AAC68950.1 CAA58532.1 NP_455145
SMR0142 SilS	c126211-127692	493	-	-	SilS (pMG101) 9930866 YbcZ (E. coli) CopS (E. coli K-12) SilS (R476b) SilS (pMIP233) SilS (pWR23) SilS (pMIP235) Reut5289/ CopS PcoS (pRJ1004)	93%/497 (0) 56%/480 (e-144) 56%/480 (e-144) 100%/198; 100%/133 99%/198; 99%/133 99%/198; 98%/130 98%/198; 99%/133 32%/475 (2e-65) 28%/465 (9e-52)	AAD11744 NP_415102 BAA35204.1 AAL68929;30 AAL68939;40 AAL68951;52 AAl68945;46 ZP_00026274 CAA58530
SMR0143 SilR	c127685-128365	226	-	-	SilR (pMG101) CopR (E. coli K12) YclA (E. coli) Reut5288/CopR PcoR (pRJ1004) Reut3251/CzcR CopR (S. typhi CT18)	92%/228 (1e-108) 87%/225 (e-105) 87%/225 (e-105) 67%/225 (4e-76) 58%/222 (5e-65) 55%/222 (2e-64) 54%/222 (2e-59)	AAD11745 BAA35205 NP_415103 ZP_00026273 CAA58529 CAA67086 NP_455593
SMR0144 SilC	128555-129940	461	-	-	SilC (pMG101) YlcB (E.coli) OprK (E.coli K12) CusC (E. coli, 11004187)	97%461 (0) 71%/450 (e-173) 71%/450 (e-173) 69%/341 (e-123)	AAD11746 NP_415104 BAA35206 AAF70174
SMR0145	129968-130321	117	-	-	ORF96 (pMG101)	96%/90 (e-38)	AAD11747.1

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

					YclC (E. coli) CusX (E. coli)	47%/117 (e-18) 46%/117 (8e-18)	NP_415105 NP_752590
SMR0146 SilB	130435-131727	430	-	-	SilB (pMG101) YclD (E. coli) CusB (E. coli) CzcB (P. putida) Reut5269	96%/430 (0) 67%/409 (e-158) 60%/409 (e-157) 39%/383 (2e-72) 37%375 (e-62)	AAD11748.1 NP_415106 NP_752591 NP_747487 ZP_00026255
SMR0147 SilA	131738-134884	1048	-	-	SilA (pMG101) YbdE (E. coli) CusA (E. coli) CzcA (P. putida) Reut5268	98%/1048 (0) 87%/1044 (0) 87%/1044 (0) 66%/1047 (0) 66%/1035 (0)	AAD11749.1 NP_415107 NP_752592 NP_747488 ZP_00026254
SMR0148	134971-135411	146	-	-	ORF105 (pMG101) SMa1008 (pSymA) SMb20881 (pSymB) PP5377 (P. putida) Reut5281/CopG	96%/105 (3-50) 43%/141 (6e-24) 41%/130 (5e-23) 40%/138 (2e-19) 37%/119 (6e-14)	AAD11752.1 NP_435790 NP_437715 NP_747478 ZP_00026267
SMR0149 SilP	135538-137985	815	-	-	SilP (pMG101) SMa1087 (pSymA) Reut5280 ActP (Rhizobium leguminosaurum) SMb21578	95%/816 (0) 66%/733 (0) 53%/808 (0) 54%/808 (0) 49%/679 (1e-164)	AAD11750 NP_435835 ZP_00026266 AAD26860 NP_437588
SMR0150	138026-138223	65	-	-	PP5390 (P. putida) SMa1089 (pSymA)	57%/38 (2e-4) 62%/53 (0.13)	NP_747491 NP_435836
SMR0151	c138257-138994	245	-	-	ORF191 (pMG101) YebA (S. typhi) YebA (E. coli)	94%/172 (2e-81) 40%/210 (6e-41) 40%/210 (7e-41)	AAD11751 NP_456457 H64947
SMR0152 CopE2	c139283-139732	149	-	-	SilE (pMG101) pcoE (pRJ1004)	34%/118 (2e-9) 35%/142 (6e-6)	AAD11743 CAA58532
SMR0153 CopA	139967-141784	605	-	-	pcoA (pRJ1004) McoA (K. pneumoniae) Psyr4025 Reut5287 CopA (R. metallidurans)	99%/605 (0) 99%/605 (0) 73%/610 (0) 54%/613 (0) 54%/613 (0)	CAA58525 AAN52531 ZP_127705 ZP_00026272 CAC07979
SMR154 CopB	141784-142680	298	-	-	pcoB (pRJ1004) Psyr4024 Reut5286 CopB (R. metallidurans)	100%/296 (e-168) 60%/264 (e-91) 44%/224 (6e-47) 44%/225 (8e-27)	CAA58526 ZP_127704 ZP_00026271 CAC07890
SMR0155 CopC	142720-143100	126	-	-	pcoC (pRJ1004) Reut5285 CopC (R. metallidurans) Psyr4023	100%/126 (2e-56) 64%/114 (e-30) 64%/114 (5e-28) 63%/112 (8e-27)	CAA58527 ZP_00026270 CAC07981 ZP_127703
SMR0156 CopD	143105-144034	309	-	-	pcoD (pRJ1004) Psyr4022 Reut5284 CopD (R. metallidurans) SMa1198 (pSymA)	100%/309 (e-139) 38%/305 (5e-46) 41%/306 (6e-44) 39%/306 (6e-40) 28%/255 (1.5)	CAA58528 ZP_127702 ZP_00026269 CAC07982 NP_435900
SMR0157 CopR	144089-144769	226	-	-	pcoR (pRJ1004) Psyr4021 YlcA (E. coli) Reut5288 czrR (R. metallidurans)	100%/226 (e-118) 60%/226 (3e-67) 60%/222 (7e-67) 63%/225 (3e-66) 57%/220 (2e-63)	CAA58529 ZP_127701 NP_415103 ZP_00026273 CAC67086

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

SMR0158 CopS	144766-146166	466	-	-	pcoS (pRJ1004) Reut5289 CopS (R. metallidurans) YbcZ (E. coli) Psyr4020 SilS (pMG101) czcS (R. metallidurans)	99%/466 (0) 33%/460 (4e-64) 32%/460 (2e-63) 29%/474 (3e-55) 31%/451 (e-51) 28%/470 (4e-49) 28%/451 (e-42)	CAA58530 ZP_00026274 CAC07977 NP_415102 ZP_127700 AAD11744 CAA67087
SMR0159 CopE1	144766-146166	144	-	-	pcoE (pRJ1004) SilE (pMG101)	92%/144 (2e-33) 45%/142 (e-15)	CAA58532 AAD11743
SMR0160	147196-148014	272	HCM1.181 R0160 HCM1.182 R0159	60%/151 (e-47) 60%/151 (e-47) 67%/115 (4e-42) 67%/115 (4e-42)	ORF26 (pCTX-M3, C. freundii) ORF20 (pWWO, P. putida) ydiA (pCollb-P9) Rsph4197 ORF62 (Rts1) HCM2.0104c	22%/281 (4e-12) 23%/259 (2e-11) 25%/207 (4e-8) 24%/224 (5e-8) 24%/208 (0.14) 18%/227	NP_774997 NP_542813 NP_052499 ZP_00008253 NP_640023 NP_569576
SMR0161	148011-149216	401	tlpA (R0158) HCM1.183	40%/406 (5e-57) 40%/406 (5e-57)	<20% identity to numerous proteins (bacterial and eukaryotic)		
SMR0162	149280-149483	67	HCM1.184c R0157	59%/67 (2e-13) 59%/67 (2e-12)	-	-	-
SMR0163	149496-150815	439	HCM1.185c R0156	81%/431 (0) 81%/431 (0)	ORF6 (pWWO) ORF544 (pNL1) SMa0901 (pSymA) mlr9256 (pMLa) yp034 (p42d, R. etli) Reut5628	36%/434 (2e-19) 25%/348 (3e-15) 23%/382 (3e-15) 25%/334 (e-13) 22%/381 (4e-12) 23%/361 (9e-7)	NP_542801 NP_049116 NP_435741 NP_085787 NP_659881 ZP_00026603
SMR0164	150838-151005	55	HCM1.186c R0155	61%/55 (3e-11) 61%/55 (4e-11)	-	-	-
SMR0165 Dcm	151066-152493	475	HCM1.187c R0154	72%/460 (0) 71%/460 (0)	EcoRII cytosine methylase Dcm (E. coli)	67%/430 (e-175) 64%/438 (e-162)	CA28725 NP_416470
SMR0166	152708-153223	171	R0153 (parB) HCM1.188	74%/170 (e-64) 72%/170 (6e-64)	ParB (RP4) Nuc (pUCD5409, S. flexneri)	37%/166 (6e-25) 36%/160 (1e-21)	AAA26415 AAA75246
SMR0167	153226-154122	298	HCM1.189 R0152	51%/297 (2e-87) 51%/297 (2e-87)	ORF151 (Rts1)	22%/254 (8e-5)	NP_640112
SMR0168	154148-154378	76	-	-	-	-	-
SMR0170	154874-155107	77	HCM1.191	56%/60 (3e-16)	-	-	-
SMR0171	155445-155732	95	HCM1.192	33%/86 (e-6)	-	-	-
SMR0172	155769-155999	76	-	-	-	-	-
SMR0173	156140-156415	91	-	-	ORF280 (Rts1)	43%/89 (2e-11)	NP_640240
SMR0174 SfpA	156907-158385	492	-	-	ORF3 (Y. enterocolitca plasmid frag.) ORF89 (Rts1) ORF51 (R391) ORFA (Tn1404) 10543801 Reut2691 Rsph4233 SMb20070 (pSymB)	99%/492 (0) 78%/492 (0) 74%/495 (0) 74%/494 (0) 74%/494 (0) 65%/491 (e-166) 65%/483 (e-158)	CAA73749 NP_640050 AAM08007 AAD47994 ZP_00023729 ZP_00008287 NP_436610
SMR0175	158404-159231	275	-	-	ORF2 (Y. enterocolitca plasmid frag.) ORF88 (Rts1) ORFB (Tn1404) ORF52 (R391) Reut2690 Rsph4234	97%/275 (e-140) 43%/278 (3e-52) 43%/282 (8e-49) 36%/283 (e-44) 36%/284 (2e-37) 31%/188 (e-16)	CAA73748 NP_640049 AAD47991 AAM08025 ZP_00023728 ZP_00008288

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

SMR0176 ArsC	c159291-159716	141	-	-	ArsC (R773) ArsC (E. coli) ArsC (R46)	90%/140 (3e-66) 88%/140 (2e-65) 87%/140 (4e-64)	AAA21096 NP_417960 NP_511240
SMR0177 ArsB	c159729-161018	429	-	-	ArsB (R46) ArsB (E. coli) ArsB (R773) Psyr4015	86%/429 (e-167) 86%/429 (e-166) 85%/429 (e-164) 73%/413 (e-129)	NP_511239 NP_417959 AAA21095 ZP_00127695
SMR0178 ArsR	c161064-161384	106	-	-	ArsR (E. coli) ArsR (R733) ArsR (R46) Psyr4014 Rsph4238	63%/105 (e-33) 59%/105 (e-30) 56%/102 (4e-29) 43%/64 (2e-11) 42%/71 (2e-9)	P_47958 BVECAR NP_511236 ZP_00127694 ZP_00008292
SMR0179 ArsH	161471-162175	234	-	-	ArsH (P. putida) Psyr4017	74%/204 (5e-83) 69%/231 (2e-87)	ZP_00127697 NP_744859
SMR0180	c162208-163611	467	HCM1.201 R0148	73%/460 (0) 96%/465 (0)	Y4bF (pNGR234) TnIS1202 ORF167 (Rts1)	47%/425 (e-101) 28%/328 (6e-28) 81%/49 (e-14)	AAB91621 Q54513 NP_640127
SMR0181	163803-164120	105	-	-	-	-	-
SMR0182	164143-164448	101	-	-	-	-	-
SMR0183	164493-165164	223	-	-	Weakly similar to serine proteases		
SMR0184	165165-165422	85	-	-	-	-	-
SMR0185	165622-166029	135	HCM1.197c R0143	55%/127 (4e-38) 55%/127 (4e-38)	-	-	-
SMR0186	c166080-166607	175	-	-	ORF161 (Rts1)	36%/172 (2e-31)	NP_640122
SMR0187	c166774-167154	126	-	-	-	-	-
SMR0188	c167227-170562	1111	HCM1.195 R0140	46%/1112 (0) 45%/942 (0)	ORF164 (Rts1) ORF173 (Rts1)	47%/449 (e-114) 41%/668 (2e-94)	NP_640125 NP_640133
SMR0189	c170785-171138	117	HCM1.250c R0139	28%/103 (e-4) 29%/114 (2e-8)	-	-	-
SMR0190	c171292-171900	202	HCM1.251c R0138	50%/202 (3e-50) 52%/130 (8e-30)	-	-	-
SMR0191 RetA	c172114-173601	495	-	-	RetA (R471a, S. marcescens) MatRa (pNL1) SMb21167 (pSymB) Many more intron-assoc. maturases	100%/495 (0) 425%/211 (e-36) 32%/377 (e-34)	AAC82519 NP_049088 CAC49233
SMR0192 MucA	174003-174437	144	HCM1.252 R0137	81%/144 (4e-63) 81%/144 (2e-62)	MucA (pCTX-M3) MucA (R446b) MucA (R471a) ORF7 (Rts1) MucA (R394) 11106794 MucA (R46) MucA (pKM101) RumA (R391) ORF79 (Rts1)	56%/119 (2e-31) 56%/119 (e-31) 56%/119 (e-31) 55%/124 (5e-28) 55%/124 (6e-28) 52%/121 (5e-26) 52%/121 (5e-26) 48%118 (e-20) 41%/122 (6e-20)	NP_774966 AAC82517 AAC82521 NP_639968 AAF14052 NP_511216 AAB36886 AAM08055 NP_640040
SMR0193 MucB	174427-175680	417	HCM1.253 R0136	83%/417 (0) 83%/402 (0)	MucB (pCTX-M3) MucB (R446b) MucB (R471a) UmuC (E. coli) MucB (R46) MucB (pKM101) RumB (R391)	44%/411 (e-105) 45%/411 (e-105) 44%/411 (e-104) 48%/417 (e-102) 46%/408 (e-101) 46%/408 (e-101) 43%/416 (7e-93)	NP_774967 AAC82518 AAC82522 NP_415702 NP_511217 AAD17384 AAM08023

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

					ORF80 (Rts1) MucB (R394) ORF8 (Rts1)	38%/415 (8e-78) 45%325 (e-76) 45%325 (2e-76)	NP_640041 AAF14053 NP_639969
SMR0194	c175726-176535	269	HCM1.254c R0135	75%/268 (e-111) 75%/268 (e-111)	bcfH (S. typhimurium) STY0033 (S. typhi)	28%/275 (5e-24) 28%/275 (6e-24)	NP_459033 NP_454638
SMR0195	176642-177253	203	HCM1.255 R0134	69%/203 (2e-82) 69%/203 (2e-82)	-	-	-
SMR0196	c177313-177735	140	HCM1.256c R0133	46%/140 (3e-25) 46%/140 (3e-25)	-	-	-
SMR0197	c177769-178161	130	HCM1.257c R0132	58%/130 (5e-35) 58%/130 (2e-35)	-	-	-
SMR0198 DsbC	c178186-178935	249	HCM1.258c R0131	84%/249 (e-120) 84%/249 (e-120)	DsbC (Erwinia sp.)	35%/246 (6e-36)	CAA54108
SMR0199	179083-179622	179	HCM1.259 R0130 (bfpH)	89%/171 (2e-86) 89%/171 (2e-86)	bfpH (pB171) trbN (pB4) trbN (X. fastidiosa) trbN (R751) trbN (RP4) trbN (pCTX-M3) Reut5666 iagB (S. typhi, cell invasion protein)	35%/135 (8e-15) 35%/116 (5e-13) 31%/158 (8e-13) 33%/119 (1e-12) 32%/119 (4e-12) 34%/120 (2e-9) 33%/114 (2e-9) 30%/117 (2e-5)	NP_053074 NP_590155 NP_779548 NP_044252 AAA26440 NP_775063 ZP_00026640 NP_457271
SMR0200	179900-180262	120	HCM1.261 R0129	68%/120 (3e-43) 70%/86 (5e-31)	-	-	-
SMR0201	180386-181447	353	-	-	YP0258 (Y. pestis) yhhZ (E. coli)	50%/216 (2e-53) 46%/234 (6e-51)	NP_404857 NP_417899
SMR0202	181457-181963	168	-	-	YrhA (E. coli K-12) y3542 (Y. pestis)	26%/136 (4e-7) 23%/105 (0.009)	NP_417900 NP_670839
SMR0203 TrhG	c182086-186036	1316	HCM1.262 R0128 (TrhG)	75%/1331 (0) 75%/1331 (0)	S011 (S. typhimurium) TraG (pED208, S. typhimurium) TrhG (pCAR1, Pseudomonas resinovorans) ORF242 (Rts1) TraG (SXT) TraG (R391) 20649 (Vibrio vulnificus)	22%/836 (7e-28) 23%/564 (2e-4) 31%/134 (0.37) 20%/106 (1e-25) 19%/1050 (6e-16) 18%/1045 (4e-15) 20.8%/1090	AAK02037 AAM90723 NP_758692 NP_640202 AAL59677 AAM07996 NP_762602
SMR0204 TrhH	c186045-187460	471	HCM1.263c R0127 (TrhH)	86.6%/471 (0) 86.6%/471 (0)	TraH (R391) TraH (SXT) TraH (R100) TrhH (pCAR1) 12547188 TraH (pED208) S012 (S. typhimurium) ORF241 (Rts1) TraH (F) TraH (pSLT) Rsph_p_4274 (R. sphaeroides) TraH (Neisseria gonorrhoeae) TraH (pNL1) 20648 (Vibrio vulnificus)	28%/297 (2e-25) 27%/297 (3e-24) 26%/425 (7e-22) 26%/418 (3e-21) 26%/362 (6e-17) 26%/341 (4e-20) 30%/258 (3e-19) 26%/354 (1e-17) 26%/355 (2e-16) 22%/363 (4e-12) 22%/452 (2e-11) 21%/455 (1e-10) 25%/383 (2e-17)	AAM08008 AAL59676 AAD28727 NP_758691 AAM90722 AAK02038 NP_640201 AAA98080 NP_490584 ZP_00008325 AAK77138 AAD03948 NP_762601
SMR0205 TrhF	c187450-188496	348	HCM1.264c R0126 (TrhF)	84%/348 (1e-164) 84%/348 (1e-164)	TraF (R391) TraF (SXT) ORF240 (Rts1)	29%/278 (4e-33) 29%/281 (8e-33) 26%/321 (7e-25)	AAM08018 AAL59678 NP_640200

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

					TraF (pCAR1) TraF (pED208) TraF (pNL1) Rsph_p_4273 (R. sphaeroides) TraF (F) TraF (pSLT) 20646 (Vibrio vulnificus) 20647 (Vibrio vulnificus)	31%/292 (2e-24) 25%/259 (2e-12) 24%/201 (4e-12) 19%/233 (3e-6) 20%/219 (4e-6) 21%/228 (2e-4) 20%/224 (8e-5) 24%/104	NP_758690 AAM90720 AAD03949 ZP_00008324 AAB61943 NP_490581 NP_762599 NP_762600
SMR0206	c189088-189600	170	HCM1.266c R0124 (TrhY)	85.9%/170 (3e-57) 85.9%/170 (3e-57)	-	-	-
SMR0207	c189602-190402	266	HCM1.268c R0123 (TrhR)	77.8%/266 (8.3e-85) 77.8%/266 (8.3e-85)	-	-	-
SMR0208	190950-191111	53	-	-	DR0518 (D. radiodurans) (86 aa)	57%/26 (5.5)	NP_294241
SMR0209	191162-191647	161	HCM1.269 R0122 (TraH)	76.4%/161 (8.5e-49) 76.4%/161 (8.5e-49)	-	-	-
SMR0210	191644-192381	245	HCM1.270 R0121	68.2%/245 (5.9e-57) 68.2%/245 (5.9e-57)	-	-	-
SMR0211 TraI	192385-195537	1050	HCM1.271 R0120 (TraI)	63%/1060 (0.0) 63%/1060 (0.0)	TraI (SXT) TraI (R391) TraI (pCAR1) ORF201 (Rts1) ORFSG44 (P. aeruginosa) Hsom1101 (H. somnus) XAC2196 (Xanthomonas axonopodis pv. citri) Pflu_p_5158 (Pseudomonas fluorescens) Reut_p_2852 (R. metallidurans) Bcep_p_3583 (Burkholderia fungorum) PHG365 (pHG1, Ralstonia eutropha) Reut5678 (R. metallidurans)	31%/261 (4e-18) 31%/244 (5e-18) 31%/239 (1e-18) 30%/203 (7e-14) 40%/86 (8e-7) 33%/127 (1e-7) 35%/92 (5e-7) 39%/86 (9e-7) 36%/105 (6e-6) 34%/105 (4e-4) 16%/238 16%/233	AAL59675 AAM08003 NP_758664 NP_640161 AAN62266 ZP_00123099 NP_642513 ZP_00087872 ZP_00023889 ZP_00030759 AAP86114 ZP_00026652
SMR0212 TraG	195537-197621	694	HCM1.272 R0119 (TraG)	83%/694 (0.0) 83%/694 (0.0)	TraD (SXT) TraD (R391) Reut_p_5679 (R. metallidurans) TraG (pHG1) ORF202 (Rts1) TraG (pCAR1) STY4562 ORFSG68 (P.aeruginosa) Hsom1138 (H. somnus) Bcep_p_3602 (B. fungorum) ORF106 (pWWO, P. putida) Reut_p_4237 (R. metallidurans) Pflu_p_5177 (P. fluorescens) XAC2259 (X. axonopodis pv. citri) Psyr4044 (P. syringae) TrwB (R388) IcmO (Leigionella pneumophilia)	31%/542 (9e-60) 31%/542 (1e-59) 26%/664 (3e-48) 26%/655 (2e-46) 26%/551 (2e-30) 27%/540 (e-42) 24%/628 (7e-30) 24%/652 (1e-28) 25%/553 (3e-28) 24%/632 (4e-28) 23%/585 (2e-27) 24%/626 (1e-27) 23%/655 (9e-27) 25%/449 (3e-26) 23%/635 (1e-21) 44%/38 (1.3) 22%/244 (1.5)	AAL59680 AAM08004 ZP_00026653 AAP86115 NP_640162 NP_758665 NP_458648 AAN62290 ZP_00123135 ZP_00030778 NP_542873 ZP_00025244 ZP-00087890 NP_642576 ZP_00127721 CAA44852 CAA75326
SMR0213	197618-198730	370	HCM1.273 R0118	65%/370 (e-144) 65%/370 (e-144)	-	-	-
SMR0214	198717-199379	220	HCM1.275	81.4%/220 (2.8e-70)	s043 (SXT)	31%/96 (4.9)	AAL59721

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

TraJ			R0117 (TraJ)	81.4%/220 (2.8e-70)	ORF36 (R391) ORF124 (pCAR1) ORF204 (Rts1) ORF105 (pWWO) STY4563 (S. typhi) SG67 (P. aeruginosa) Hsom1136 (H. somnus) Pflu5176 (P. fluoroscens) PHG369 (pHG1) Reut4238 (R. metallidurans) Bcep3601 (B. fungorum) Reut5682 (R. metallidurans) Psyr4045 (P. syringae) XAC2260 (X. axonopodis pv. citri)	30%/96 (7.0) 28%/46 21%/184 18%/158 16%/178 17%/160 18%/154 16%/154 14%/158 14%/159 14%/159 13%/169 18%/127 20%/89	AAM08039 NP_758667 NP_640164 NP_542872 NP_458649 AAN62289 ZP_00123133 ZP_00087889 AAP86118 ZP_00025245 ZP_00030777 ZP_00026656 ZP_00127722 NP_642577
SMR0215	199390-200571	393	HCM1.276 R0116	72%/393 (e-162) 72%/393 (e-162)	sppA (P. syringae) * and numerous other proteases	44%/125 (e-19) and 29%/131 (0.003)	NP_793611
SMR0216	200573-201304	243	HCM1.277 R0115	62%/242 (2e-88) 62%/242 (2e-88)	BXA0003 (pXO1-01) ORF143 (pCAR1)	32%/115 (2e-11) 29%/144 (e-8)	NP_652769 NP_758686
SMR0217	c201591-201944	117	HCM1.278c R0114	73%/116 (7e-41) 73%/116 (7e-41)	-	-	-
SMR0218	c202010-202294	94	HCM1.279c R0113	55%/60 (e-8) 55%/60 (e-8)	-	-	-
SMR0219 Tn10R	202920-204128	402	HCM1.194 R0085	99%/402 (0) 99%/401 (0)	>97% identity to numerous eukaryotic and bacterial transposases		
SMR0220 TetD	c204138-204554	138	R0084 TetD	100%/138 (3e-61)	TetD (S. flexneri) TetD (E. coli) Rob (E.coli) 10802742, 1D5Y ykgA (E. coli) SoxS (S. typhi) MarA (E. coli) 9724717	100%/138 (3e-61) 98%/138 (e-60) 60%/104 (3e-31) 40%/122 (4e-20) 50%/102 (5e-20) 45%/102 (2e-17)	AAD50249 TETD_ECOLI NP_418813 NP_752353 NP_418486 NP_416048
SMR0221 TetC	204642-205235	197	R0083 TetC	100%/197 (e-101)	TetC (S. flexneri) * numerous others ~30 identity	100%/197 (e-101)	AAD50248
SMR0222 TetA	c205348-206553	401	HCM1.241c R0082 TetA	99%/401 (0) 99%/401 (0)	TetA (S. flexneri) * highly conserved antiporter	100%/401 (0)	AAD50247
SMR0223 TetR	206635-207258	207	HCM1.243 R0081 TetR	100%/207 (e-112) 100%/207 (e-112)	TetR (S. flexneri) * highly conserved repressor	100%/207 (e-112)	AAD50246
SMR0224	c207236-207922	228	HCM1.244c R0080	99%/228 (e-122) 100%/228 (e-122)	JemC (S. flexneri) * ~30% identity to ArsR-family transcriptional repressors	100%/228 (e-122)	AAD50245
SMR0225	c07930-208316	128	HCM1.245c R0079	100%/128 (e-62) 100%/128 (e-63)	ydjB (R100) BMEI0153 (Brucella melitensis)	100%/128 (5e-63) 50%/127 (6e-28)	NP_052926 NP_539071
SMR0226	c208309-208629	106	HCM1.246c R0078	100%/106 (e-49) 100%/106 (e-49)	ydjA (R100) JemB (S. flexneri) BMEI0152	100%/106 (e-49) 100%/106 (e-49) 53%/100 (5e-23)	NP_052927 AAD50244 NP_539070
SMR0227 GtlS	209073-210278	401	R0077 Glts HCM1.247 (psuedogene)	99%/77 (0)	ydhA (R100) JemA (S. flexneri) * highly conserved transporter	99%/77 (0) 99%/77 (0)	NP_052926 AAD50243
SMR0228 Tn10L	c210644-211852	402	HCM1.249c R0076	100%/402 (0) 99%/402	>97% identity to numerous eukaryotic and bacterial transposases		
SMR0230	c212539-212907	122	HCM1.196 R0142	45%/132 (e-20) 45%/132 (e-20)	-	-	-

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

SMR0231	c213133-213669	178	HCM1.280c R0100	63%/177 (3e-58) 63%/177 (3e-58)	-	-	-
SMR0232	c213737-214153	138	HCM1.281c R0109	54%/136 (2e-34) 54%/136 (2e-34)	-	-	-
SMR0233	c214219-214515	98	-	-	-	-	-
SMR0234	c214650-215351	233	R0107	55%/61 (3e-11)	-	-	-
SMR0235	c215666-215947	93	HCM1.284c R0105	61%/93 (6e-22) 61%/93 (6e-22)	-	-	-
SMR0236	c216013-217197	394	HCM1.286c R0104 HCM1.58c R0042 HCM1.63c R0037	65%/403 (e-131) 65%/403 (e-131) 35%/224 (2e-21) 35%/224 (2e-21) 28%/190 (2e-5) 28%/190 (2e-5)	ParB (Caulobacter vibrioides)	41%/60 (0.020)	AAB51268
SMR0237	c217614-217844	76	-	-	-	-	-
SMR0238	218301-219245	314	-	-	-	-	-
SMR0239	219344-219943	199	-	-	-	-	-
SMR0240	220003-220353	116	-	-	-	-	-
SMR0241	c220489-220812	107	-	-	-	-	-
SMR0242	220885-221205	105	-	-	-	-	-
SMR0243 InsL	221857-222969	370	-	-	IS186 Tn (E. coli) IS231 Tn (B. thuringiensis) ORF39 (pX01)	99%/370 (0) 23%/385 (6e-9) 22%/287 (2e-8)	NP_416895 CAD30116 NP_052735
SMR0243a Hok	c223163-223321	52	HCM1.290c R0101 Stm	86%/52 (2e-16) 86%/52 (2e-17)	Hok (R100) FlmA (F plasmid)	69%/52 (6e-12) 69%/52 (9e-12)	NP_052939 AAA99216
SMR0245 InsK	c223420-224271	283	-	-	IS150 Tn (E. coli) ISPsy9 Tn OrfB ISSmu1 Tn IS861 Tn (S. agalactiae) IS3-Spn1 Tn IS1661 (Y. pestis) IS1353 (R100)	100%/283 (e-156) 58%/272 (2e-89) 51%/274 (2e-67) 50%/272 (3e-66) 47%/269 (5e-61) 48%/258 (8e-60) 39%/279 (2e-54)	NP_418015 NP_794751 NP_720882 NP_688077 AAF73806 NP_403776 NP_052891
SMR0246 InsJ	c224268-224789	173	-	-	IS150 (E. coli) YPo0119 ISPsy9 OrfA	100%/173 (6e-87) 27%/163 (2e-9) 27%/171 (e-7)	NP_418014 NP_403777 NP_794752
SMR0247	c225122-225727	201	R0100 HCM1.01c	42%/198 (9e-39) 44%/175 (7e-34)	-	-	-
SMR0248	c225944-226225	93	HCM1.02c R0099	76%/86 (2e-29) 76%/86 (7e-29)	-	-	-
SMR0249	c226601-226912	103	HCM1.03c R0098	60%/97 (5e-26) 60%/97 (5e-26)	ORF291 (Rts1)	30%/105 (0.002)	NP_640251
SMR0250	c227135-227335	66	HCM1.10c R0093 HCM1.13c R0092	58%/55 (e-9) 58%/55 (e-9) 29%/41 (2.1) 29%/41 (2.1)	-	-	-
SMR0251	c227375-227599	74	HCM1.13c R0092 HCM1.30c R0072	79%/74 (4e-26) 79%/74 (4e-26) 50%/63 (e-9) 50%/63 (e-9)	-	-	-
SMR0251a	c227654-227857	67	HCM1.15c	28%/67 (0.0064)	-	-	-
SMR0252	c228410-228901	163	HCM1.20c	51%/162 (9e-45)	STY4595	41%/58 (0.005)	NP_4586789

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

			R0089	51%/162 (4e-44)	ORF23 (pCTX-M3)	24%/90	NP_774993
SMR0253	c228906-229217	103	HCM1.21c R0088	73%/100 (3e-28) 73%/100 (3e-28)	ORF61 (E. coli 536) * conserved in many E. coli strains	48%/49 (e-6)	CAD66209
SMR0254	c229734-230054	106	-	-	-	-	-
SMR0255	c230233-230463	76	HCM1.23c R0086	57%/68 (9e-14) 57%/68 (9e-14)	-	-	-
SMR0256 KlaC	c230635-231528	297	R0074	32%/95	KlaC (pCAR1)	23%/263 (0.22)	NP_758600
SMR0257 KlaB	c231518-232618	366	-	-	TelA (RK2 Te-r) KlaB (pCAR1) YaaN (B. subtilis) ORF182 (Rts1) * many tellurite resistance proteins	56%/359 (e-104) 57%/356 (e-101) 30%/340 (7e-37) 29%/368 (9e-33)	AAA98154 NP_758599 NP_387907 NP_640142
SMR0258 KlaA	c232627-233412	261	-	-	KilA (RK2 Te-r) KlaA (pCAR1) ORF183 (Rts1)	45%/256 (3e-61) 41%/249 (4e-50) 19%/202 (0.008)	AAA98153 NP_758598 NP_640143
SMR0259	Join c233643-233798 to c233802-233930	94	HCM1.25c	67%/95 (2e-24)	-	-	-
SMR0261	c234007-234615	202	HCM1.26c R0074 (1-42)	62%/201 (9e-67) 69%/142 (5e-9)	PA2593 (P. aeruginosa) Avin0923	34%/112 (9e-8) 38%/47 (2e-4)	NP_251283 ZP_00089256
SMR0262	c234729-235262	177	HCM1.27c	79%/174 (5e-174)	SF1348 (S. flexneri)	45%/135 (7e23)	NP_707245
SMR0263	c235317-235511	64	-	-	-	-	-
SMR0264	c235508-235819	103	HCM1.28c	70%/98 (2e-34)	-	-	-
SMR0265	c235882-236121	79	HCM1.30c R0072 HCM1.13c R0092	69%/66 (e-14) 69%/66 (e-14) 51%/70 (5e-14) 51%/70 (5e-14)	-	-	-
SMR0266	c236371-238755	794	HCM1.32c R0070	76%/792 (0) 75%/791 (0)	ebiG17 (Anopheles gambiae, contam?) alr4917 (Nostoc sp.) SAV2951 (Streptomyces avermitilis)	40%/635 (e-137) 28%/319 (5e-17) 30%/291 (3e-15)	EAA01880 NP_488957 NP_824107
SMR0267	c238921-239370	149	HCM1.33c R0069	61%/144 (e-43) 61%/144 (3e-43)	-	-	-
SMR0268	c239421-240212	263	HCM1.34c R0068	59%/264 (4e-81) 59%/264 (4e-81)	-	-	-
SMR0269	c240442-240699	85	HCM1.35c R0067	45%/55 (4e-5) 46%/45 (0.026)	-	-	-
SMR0270	c240765-241091	108	HCM1.36c R0066	62%/106 (e-33) 62%/106 (e-33)	-	-	-
SMR0271	c241334-241654	106	-	-	-	-	-
SMR0272	c241949-243199	416	-	-	SptAIM (S. paratyphi) PvuII (P. vulgarus)	31%/314 (e-33) 30%/393 (4e-31)	AAG42425 CAA32026
SMR0273	c243370-243978	202	-	-	CPE0665 (C. perfringens) SCC24.26c (S. coelicolor) RSp1337 (R. soanacearum)	30%/82 (6e-6) 32%/98 (4e-4) 27%/98 (0.017)	NP_561581 NP_626698 NP_522896
SMR0274	c244134-244406	90	-	-	-	-	-
SMR0275	c244456-244800	115	-	-	-	-	-
SMR0276	c244976-245353	125	-	-	-	-	-
SMR0277	c245759-246940	393	HCM1.38c R0065	62%/392 (e-143) 61%/316 (e-114)	-	-	-

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

SMR0278	c246949-247245	98	HCM1.39c	89%/98 (4e-39)	VCA0391 SMc02461 HigB (Rts1, ORF261)	43%/95 (e-11) 29%/98 (e-4) 30%/93 (0.01)	NP_232785 NP_385157 NP_640221
SMR0279	c247295-247765	156	HCM1.40c R0063	64%/155 (2e-47) 64%/155 (2e-47)	-	-	-
SMR0280	c248199-250589	796	HCM1.49c R0055 HCM1.56c R0045 R0047 HCM1.52 R0051 HCM1.50 R0054	38%/787 (e-139) 38%/787 (e-139) 29%/843 (8e-69) 26%/723 (e-35) 49%/109 (2e-23) 41%/130 (3e-20) 40%/125 (4e-18) 30%/160 (5e-13) 31%/138 (e-12)	ORF131 (pCAR1)	23%/805 (3e-33)	NP_758674
SMR0281	c250992-251243	83	-	-	-	-	-
SMR0282	c251451-251708	85			-	-	-
SMR0283	c251701-252132	143	-	-	ORF281 (Rts1)	38%/130 (2e-7)	NP_640241
SMR0284	c252236-252760	174	-	-	-	-	-
SMR0285	252776-252916	46	-	-	-	-	-
SMR0286	253083-253847	254	R0060 HCM1.43 HCM1.42 R0061	61%/255 (9e-85) 61%/255 (1e-84) 32%/264 (2e-29) 32%/264 (2e-29)	-	-	-
SMR0287	253865-254689	274	HCM1.57c R0043	29%/141 (9e-8) 29%/141 (e-7)	Rsph4197 Unnamed (pO157) ydcA (R100) ORF139 (Rts1) HCM2.0104c	29%/186 (5e-12) 29%/131 (2e-10) 29%/131 (8e-10) 27%/153 (2e-8) 30%/190	NP_82537 NP_052654 NP_052920 NP_640100 NP_569576
SMR0288	254783-255046	87	HCM1.45 R0058	66%/87 (5e-29) 66%/87 (e-28)	ORF287 (Rts1)	31%/67 (7e-4)	NP_640247
SMR0289	255090-255533	147	-	-	-	-	-
SMR0290	255767-258187	806	HCM1.56c R0045 R0055 HCM1.49c R0054 HCM1.50 HCM1.52 R0051 R0047 HCM1.51 R0052 R0053 HCM1.53	42%/814 (e-176) 39%/710 (e-136) 31%/812 (6e-89) 31%/812 (9e-89) 70%/202 (3e-75) 71%/200 (e-74) 85%/131 (2e-56) 86%/131 (5e-56) 60%/112 (2e-30) 74%81 (4e-28) 74%81 (4e-28) 85%/57 (2e-21) 86%/45 (3e-15)	ORF131 (pCAR1)	29%/799 (2e-86)	NP_758674
SMR0292 RepHI2	c260265-261353	362	R0036 RepHIA HCM1.64 RepA2 R0050 RepHIB HCM1.54 RepA	43%/285 (5e-59) 43%/285 (5e-59) 38%/276 (2e-46) 38%/276 (2e-46)	RepA (pHCM2) RepA (pMT1, Y. pestis) RepA (R386) RepA (pSLT) RepFIB (F plasmid) RepA (Vir plasmid, S. cholerasesuis) RepA (pO157) RepA (P1 phage)	51%/333 (e-83) 51%/346 (e-82) 52%/283 (5e-71) 50%/308 (5e-71) 52%/283 (9e-71) 49%/316 (9e-71) 52%/274 (2e-68) 41%/269 (2e-53)	NP_569603 NP_395385 AAA71884 NP_490513 AAA71886 NP_073241 NP_052630 BVBPRA

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

					RepA (pSW1200, Pantoea stewertii) RepA (Rts1) RepA1 (pS268a, Serratia sp.) RepA2 (pS268a, Serratia sp.)	43%/257 (2e-52) 41%/271 (3e-52) 42%/140 (3e-24) 43%/128 (2e-22)	AAB66323 NP_639962 AAL84745 AAL84746
SMR0293	c262461-264917	818	HCM1.56c R0045 R0055 HCM1.49c R0047 R0051 HCM1.52 HCM1.50 R0054 R0052 HCM1.51 HCM1.53 R0053	81%/818 (0) 81%/705 (0) 35%/836 (3e-99) 35%/834 (6e-99) 83%/112 (4e-46) 56%/127 (3e-31) 52%/129 (9e-31) 40%/196 (3e-28) 39%/200 (9e-28) 45%/81 (e-10) 45%/81 (e-10) 54%/46 (7e-6) 38%/54 (0.001)	ORF131 (pCAR1)	30%/826 (e-100)	NP_758674
SMR0294	265290-265916	209	-	-	ORF4 (pCTX-M3) Atu0367 * and numerous other methionyl-tRNA formyl transferases (new domain at 98-156?)	44%206 (7e-47) 41%/53	NP_774970 NP_513073
SMR0295	c265974-266255	93	-	-	-	-	-
SMR0296	c266242-266685	147	-	-	-	-	-
SMR0297	c266746-266922	58	HCM1.56ac R0044	61%/57 61%/57	-	-	-
SMR0298	c266919-268610	563	HCM1.57c R0043	70%/558 (0) 70%/558 (0)	HCM2.104c ORF62 (Rts1) ORF139 (Rts1) ORF26 (pCTX-M3)	30%/574 (2e-67) 29%/522 (e-49) 28%/485 (2e-45) 28%/235 (9e-12)	NP_569576 NP_640023 NP_640100 NP_774997
SMR0299	c268858-270021	387	HCM1.58c R0042 HCM1.286c R0104 HCM1.63c R0037	72%/384 (e-159) 72%/384 (e-159) 30%/236 (8e-16) 29%/231 (2e-15) 29%/200 (2e-11) 29%/200 (2e-11)	ORF112 (pCAR1) ORF273 (Rts1) ParB (Chlamydophilia caviae)	28%/243 (2e-9) 24%/230 (2e-7) 24%/85 (0.019)	NP_758655 NP_640233 NP_828930
SMR0300 Dam	270370-271203	277	HCM1.59c R0041	70%/272 (e-111) 70%/257 (e-102)	STY1014 ORF53 (Rts1) ORF131 (Rts1) * >30% id to many DNA adenine methylases	39%/270 (2e-45) 35%/267 (9e-38) 34%/265 (7e-32)	NP_455493 NP_640014 NP_640092
SMR0301	c271212-272063	283	HCM1.60 R0040	63%/277 (2e-95) 63%/277 (2e-95)	-	-	-
SMR0302	c272301-273017	238	HCM1.61c R0039	53%/231 (7e-63) 53%/231 (7e-63)	-	-	-
SMR0303	c273014-273214	66	HCM1.62c R0038	53%/66 (3e-10) 51%/66 (5e-10)	-	-	-
SMR0304	c273234-274286	350	HCM1.63c R0037 HCM1.58c R0042 HCM1.286c	58%/355 (5e-90) 58%/355 (5e-90) 25%/298 (8e-16) 25%/298 (8e-16) 27%/195 (5e-7)	ORF112(pCAR1) ParB (Treponema pallidum) ORF273 (Rts1)	20%/189 (6e-4) 26%/165 (0.002) 23%/143 (0.33)	NP_758655 NP_218712 NP_640233

			R0104	27%/195 (5e-7)			

a. The co-ordinates of the R478 sequence begin with the A in the start codon of the *repHIA* gene. Genes which have co-ordinates beginning with a "c" are encoded on the anti-sense strand to the *repHIA* gene.

b. the percentage of amino acids aligned in the indicated range. E-value is as reported by BLAST, and if no value is reported, the alignment was created using PSI-BLAST. A solid dash indicates that no significant alignments were generated.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Curriculum Vitae: Matthew W. Gilmour

E-mail: mgilmour@ualberta.ca or matt.gilmour@ualberta.net

Professional Address: 1-28 Medical Science Building
Department of Medical Microbiology and Immunology
University of Alberta
Edmonton, AB, Canada
T6G 2R3
Tel: 780-492-4437; Fax: 780-492-7521

Home Address: 204-10530 84 Avenue
Edmonton, AB, Canada
T6E 2H4
Tel: 780-989-2998

EDUCATION:

- Ph.D. - Department of Medical Microbiology and Immunology (University of Alberta); September 1999 - March 2004, Supervisor: Dr. Diane Taylor
  - Thesis defense examination passed February 26th, 2004
  - Project: Determination of the essential genes, protein function, and protein-protein interactions required for the conjugative transfer (i.e. Type IV secretion) of R27, a 180 kbp resistance plasmid from *Salmonella enterica* serovar Typhi
  - Project: Comparative genomics of IncHI plasmids and related conjugative genetic elements. The IncHI2 plasmid R478 isolated from *Serratia marcescens* is 275 kbp, confers multiple drug resistance, and shows sequence identity to numerous other self-transmissible elements. My goals are to annotate the R478 sequence data to reveal the function and the evolutionary history of backbone and accessory genes, and to compare the IncHI1 plasmid backbone features to those encoded in the IncHI2 plasmid in order to determine the minimal set of determinants for plasmid maintenance and conjugative transfer. These aims are being achieved by using Sanger Institute software Artemis and ACT for the annotation and comparative analysis of genomes.
  - Project: Commissioned to write with Dr. Diane Taylor a chapter in the new online version of EcoSal (see www.ecosal.org. - formerly called '*Escherichia coli* and *Salmonella*: Cellular and Molecular Biology', edited by Neidhardt *et al.*); section 'Molecular Architecture of Bacterial Cells'; subsection 'Cytology'; tentative chapter title: 'Cytology of Bacterial Conjugation' (editor: Dr. B.B. Finlay).
  - Project: Collaboration with the research groups of Samuel Kaplan (University of Texas) and Frank Larimer (Oak Ridge National Laboratory) to complete the annotation of the *Rhodobacter sphaeroides* 2.4.1 genomic sequence. My particular role is to annotate the coding sequences of the 5 large plasmids present in the sequenced strain.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

- Experimental procedures routinely used include: bioinformatic analysis of DNA sequences, molecular DNA techniques, protein purification through affinity chromatography, fluorescence microscopy, bacterial two-hybrid analysis, cellular fractionation, functional analysis of purified proteins, and site-directed mutagenesis of whole genes and specific motifs

- B.Sc. with Specialization in Cell Biotechnology (with Distinction); June 1999, University of Alberta
  - This undergraduate program was a combination of both the Microbiology and Genetics programs offered by the Department of Biological Sciences, with a focus on obtaining formal training in laboratory techniques
  - Instructional laboratories include: Microbiology Laboratory Techniques (protein and DNA purification), Microbial Physiology Laboratory (metabolism of microorganisms), Introduction to Molecular Genetics Techniques (Molecular and Cellular biology), Bacterial Pathogenesis Laboratory (clinical analysis of bacterial isolates), Fermentation Biotechnology (large-scale fungal and bacterial cultures and product purification)
  - Thesis project: characterization of the membrane-associated protein TrhC required for bacterial conjugative transfer of plasmid R27 (Supervisor: Dr. Diane Taylor)
  - Received Dean's Honor Roll placement in third and final years of the program

PUBLICATIONS (Refereed):

1) **Gilmour MW** and Taylor DE. 2004. A subassembly of R27-encoded transfer proteins is dependent on TrhC nucleotide triphosphate-binding motifs for function but not formation. *Journal of Bacteriology*. 186(6):1606-13.

2) **Gilmour MW**, Gunton JE, Lawley TD, Taylor DE. 2003. Interaction between the IncHI1 plasmid R27 coupling protein and Type IV secretion system: TraG associates with the coiled-coil Mating pair formation protein TrhB. *Molecular Microbiology*. 49(1):105-16.

3) Lawley TD, **Gilmour MW**, Gunton JE, Tracz DM, Taylor DE. 2003. Functional and mutational analysis of conjugative transfer region 2 (Tra2) from the IncHI1 plasmid R27. *Journal of Bacteriology*. 185(2):581-91.

4) Lawley TD, **Gilmour MW**, Gunton JE, Standeven LJ, Taylor DE. 2002. Functional and mutational analysis of conjugative transfer region 1 (Tra1) from the IncHI1 plasmid R27. *Journal of Bacteriology*. 184(8):2173-80.

5) **Gilmour MW**, Lawley TD, Rooker MM, Newnham PJ, Taylor DE. 2001. Cellular location and temperature-dependent assembly of IncHI1 plasmid

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

R27-encoded TrhC-associated conjugative transfer protein complexes. *Molecular Microbiology*. 42(3):705-15.

6) Sherburne CK, Lawley TD, **Gilmour MW**, Blattner FR, Burland V, Grotbeck E, Rose DJ, Taylor DE. 2000. The complete DNA sequence and analysis of R27, a large IncHI plasmid from Salmonella typhi that is temperature sensitive for transfer. *Nucleic Acids Research*. 15;28(10):2177-86.

MANUSCRIPTS (Submitted for publication):

1) **Gilmour MW**, Thomson NR, Sanders M, Parkhill J, Taylor DE. The complete nucleotide sequence of R478: defining the backbone components of incompatibility group H conjugative plasmids through comparative genomics. Submitted to the *Journal of Molecular Biology* October 2003.

PUBLISHED ABSTRACTS

1) **Gilmour MW** and Taylor DE. Assembly requirements of TrhC-associated protein complexes which are involved in the conjugative transfer of R27. Poster presented at Plasmid Biology 2002. Pittsburgh, Pennsylvania, USA. June 22-28, 2002. Abstract published in *Plasmid*. 48(3):252.

2) Gunton JE, Lawley TD, **Gilmour MW**, Taylor DE. Characterization of the coupling protein of the IncHI1 plasmid R27. Poster presented at Plasmid Biology 2002. Pittsburgh, Pennsylvania, USA. June 22-28, 2002. Abstract published in *Plasmid.* 48(3): 256.

3) Tracz DM, Lawley TD, **Gilmour MW**, Gunton JE, and Taylor DE. Characterization of the Tra1 and Tra2 conjugative transfer regions of the IncHI1 plasmid R27: defining a type IV secretion system subfamily. Poster presented at Plasmid Biology 2002. Pittsburgh, Pennsylvania, USA. June 22-28, 2002. Abstract published in *Plasmid*. 48(3): 282.

4) **Gilmour MW**, Lawley TD, Gunton JE, Tracz DM, and Taylor DE. Characterization of the Tra1 and Tra2 conjugative transfer regions of the IncHI1 plasmid R27: defining a type IV secretion system subfamily. Poster presented at the UA-UC Conference on Infectious Diseases 2002. Banff, Canada. April 21-24, 2002. Abstract #1.

5) **Gilmour MW**, Lawley TD, Rooker M, Gunton JE, and Taylor DE. Fusion of green fluorescent protein to the conjugal transfer protein TrhC: localization and temperature-dependent expression. Poster presented at the EURESCO conference on the Biology of Type IV Secretion Processes: Euroconference

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

on the Medical and Ecological Implications. Lucca, Italy. September 7-12 2001. pg.74.

6) **Gilmour MW**, Lawley TD, Sherburne CK, Blattner FR, Burland V, Grotbeck E, Rose DJ, and Taylor DE. Highlights of the complete DNA sequence and analysis of R27, a large IncHI plasmid from *Salmonella typhi* that is temperature sensitive for transfer. Poster presented at the International Symposium on Molecular Biology of Bacterial Plasmids. Prague, Czech Republic. September 19-25, 2000. Abstract published in *Plasmid* 45(2):164-165.

INVITED PRESENTATIONS/VISITS:

1) University of British Columbia (Vancouver, Canada): March 16th, 2003; host: Dr. Brett B. Finlay

2) National Microbiology Laboratory (Winnipeg, Canada): June 4th, 2003; host: Dr. Lai-King Ng

3) The Sanger Institute (Wellcome Trust Genome Campus, Cambridge, UK): August 17-29, 2003; hosts: Dr. Julian Parkhill and Dr. Nick Thomson

4) University of Texas (Houston, USA); December 4th, 2003; host: Dr. Peter Christie

CONFERENCES/WORKSHOPS ATTENDED:

1) Canadian Bioinformatics Workshops: BIOINFORMATICS 2003, Vancouver, Canada. February 2003. Intensive two-week program with lectures/tutorials/laboratories focussing on databases, webpage design, sequence submission, sequence alignment, unix/linux platforms, BLAST, sequence analysis, phylogeny, PERL programming, RNA sequence analysis, and protein structural analysis. (Final grade: 91%, A+ standing)

2) Plasmid Biology 2002. Pittsburgh, USA. June 2002. Poster presented.

3) UA-UC Conference on Infectious Diseases 2002. Banff, Canada. Poster presented.

4) EURESCO conference on the Biology of Type IV Secretion Processes. Lucca, Italy. September 2001. Poster presented.

5) International Symposium on the Biology of Plasmids. Prague, Czech Republic. September 2000. Poster presented.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

GRANTS & AWARDS:

1) 2003-2004. Honorary Izaak Killam Memorial Scholarship. This is the highest honour awarded to Graduate Students by the University of Alberta, and the award component covers tuition, a $2600 funding increase, and a $2000 research allowance

2) 2003. Andrew Stewart Memorial Graduate Student Prize. $5000

3) 2002-2007. Canadian Institutes of Health Research operating grant (awarded to D.E. Taylor, principal investigator, and T.D. Lawley and M.W. Gilmour, co-applicants)

4) 2000-present. Alberta Heritage Foundation for Medical Research Studentship. $20000/year + $1500 research allowance/year

5) 2000. J Gordon Kaplan Graduate Student Award. $800

6) 1995. Alexander Rutherford Scholarship. $1500

INSTRUCTIONAL EXPERIENCE:

- Teaching Assistant for MMI 520 (Biology of Plasmids); 2000-present. Duties include lecturing students, grading presentations and grant proposals, and preparing and marking exams
- Teaching Assistant for MMI 351 (Bacterial Pathogenesis Laboratory); 2000-present. Duties were same as MMI 520, plus supervision of students
- Direct supervision and instruction of junior graduate students (James Gunton, Leah Standeven, and Dobryan Tracz) and undergraduate project students (Jamil Janmohamed, Eugene Lam, and Andrew Ting)

REFERENCES:

1) Dr. Diane Taylor (diane.taylor@ualberta.ca; 780-492-4777)

2) Dr. Laura Frost (laura.frost@ualberta.ca; 780-492-0672)

3) Dr. Glen Armstrong (glen.armstrong@ucalgary.ca; 403-220-6885)

4) Dr. Markus Stein (mstein@ualberta.ca; 780- 492-5495)

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.