The Phages and Phage-like Elements of *Rhodobacter capsulatus*

by

© Alexander P. Hynes

A thesis submitted to the

School of Graduate Studies

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Biology

Memorial University of Newfoundland

May 2014

St. John's Newfoundland and Labrador

ABSTRACT

Rhodobacter capsulatus is a purple non-sulfur bacterium that produces the gene transfer agent RcGTA. This phage-like particle is capable of transferring ~4-kb of host DNA to other *R. capsulatus* cells in a process analogous to transduction. The genes known to encode this particle are located in a ~15-kb region called the RcGTA structural gene cluster. As this cluster is larger than the packaging capacity of the particles, RcGTA is non-replicative. In addition to lacking key functions required by a phage, the RcGTA structural gene cluster lacks genes likely to encode functions necessary for its gene transfer activity, such as a cellular release mechanism. The costs and benefits of RcGTA production for *R. capsulatus* are unclear, making it difficult to explain its persistence in the genome. I investigated *R. capsulatus* SB1003 prophages to search for other genes involved in RcGTA production and gene transfer activity. I identified two functional prophages, RcapNL and RcapMu, which I characterized and found to be linked to RcGTA production. I also compared the genome-wide transcriptional profiles in a variety of *R. capsulatus* strains and growth conditions affecting RcGTA production to identify genes consistently co-regulated with the RcGTA structural gene cluster. I found nine such genes at six separate loci. In characterizing several of these genes, I identified a gene required for release of RcGTA by cell lysis, a pair of putative tail fibre-encoding genes, and an additional gene whose product is required for binding to the recipient cell. Finally, I characterized features of RcGTA production by quantitatively determining the packaging frequency of all loci on the *R. capsulatus* chromosome. Remarkably, while any gene can be transferred, the RcGTA structural gene cluster region was under-represented

in the particles relative to the genome average. The search for a mechanistic explanation of this anomaly led to the discovery that RcGTA gene expression is elevated in ~3% of the cells in a population, and that these cells undergo lysis to release RcGTA particles. This provided us with the first quantification of the cost of RcGTA production.

ACKNOWLEDGEMENTS

I would like to acknowledge, first and foremost, my supervisor Dr. Andrew S. Lang for his patience, guidance and mentorship. His interest and accessibility have made graduate school engaging, his rigour has driven me to improve, and his continued involvement and support have made this research possible. I also would like to thank my thesis committee for their attention and efforts, despite setbacks and an unnerving propensity for storms, break-ins and other disasters on days of scheduled committee meetings. Of my numerous co-authors, all of whom are acknowledged in more detail in the co-authorship statement, I'd like to make special mention of Dr. Ryan G Mercer, a colleague and friend who made grad school more enjoyable. His feedback was invaluable and, as trailblazer for Lang lab doctoral candidates, his trials, tribulations and thesis served as guideposts for my own. Finally, I must thank my father, Dr. Michael F Hynes. Having a parent in your field is a boon I would wish on every graduate student. His wealth of knowledge, both in our field and of academia in general, has benefited me tremendously.

My doctoral program and the research I have performed has been supported by a graduate student fellowship from the Memorial University School of Graduate Studies, a fellowship from Natural Sciences and Engineering Research Council of Canada (NSERC), and support from the Broad Institute's Marine Virome Sequencing Project with matching contributions from Newfoundland and Labrador Research and Development Corporation (RDC). The research in Dr. Lang's laboratory was supported by grants from NSERC, RDC and the Canadian Foundation for Innovation (CFI).

Table of Contents

List of Tables

List of Figures

List of Symbols, Nomenclature or Abbreviations

ANOVA – Analysis of variance

BLAST – Basic local alignment search tool

bp – Base pair

BSA – Bovine serum albumin

CAS – CRISPR associated

CHP – Conserved hypothetical protein

CRISPR – Clustered regularly-interspaced short palindromic repeats

CTAB - Cetyl trimethylammonium bromide

dsDNA – Double-stranded deoxyribonucleic acid

DNA – Deoxyribonucleic acid

DNase – Deoxyribonuclease

dNTPs – Deoxynucleotide triphosphates

DUF – Domain of unknown function

EDTA – Ethylenediaminetetraacetic acid

EMSA – Electrophoretic mobility shift assay

EPS – Extracellular polysaccharide OR exopolysaccharide

EM – Electron micrograph

FDG – Fluorescein di-β-d-galactopyranoside

GEO – Gene expression omnibus

GI = Genbank identifier

GTA – Gene transfer agent

HGT – Horizontal gene transfer

HP – Hypothetical protein

HSD – Honestly significant difference

HSL – Homoserine lactone

IPTG – Isopropyl-β-D-thiogalactopyranoside

kb – Kilobase pair

LB – Luria-Bertani

Mb – Megabase pair

M-MuLV – Moloney murine leukemia virus

NCBI – National center for biotechnology information

Ni-NTA - Nickel nitriloacetic acid

Nr – Non-redundant

OD – Optical density

ORF – Open reading frame

PAM – Protospacer associated motif

PAGE – Polyacrylamide gel electrophoresis

PCR – Polymerase chain reaction

PEG – Polyethylene glycol

PFGE – Pulsed field gel electrophoresis

PVDF – Polyvinylidene difluoride

RcGTA – *Rhodobacter capsulatus* gene transfer agent

RCV – *Rhodobacter capsulatus* medium V

RMA – Robust multi-array average

RMS – Restriction/modification system

RNA – Ribonucleic acid

RNase – Ribonuclease

RPM – Revolutions per minute

rRNA – ribosomal ribonucleic acid

SaPI – *Staphylococcus aureus* pathogenicity island

SDS – Sodium dodecyl sulfate

SDS-PAGE –Sodium dodecyl sulfate polyacrylamide gel electrophoresis

TBE – Tris-Borate ethyldiaminetetraacetic acid

TE – Tris/ethyldiaminetetraacetic acid

YPS – Yeast Extract/peptone/salts medium

List of Appendices

1 Introduction and Overview

1.1 Bacteriophages

Bacteriophages are viruses that infect bacteria. They consist of a genome, coding for and packaged within a "capsid" structure, and capable of infiltrating a bacterial cell. Once inside, the genome will eventually replicate and generate progeny capable of escaping the cell to begin the infection cycle anew. Phage genomes can be RNA or DNA, single-stranded or double-stranded, linear or circular, segmented or contiguous. The physical capsids can vary from polygons to rods, have elaborate protrusions or none, and possess tail structures ranging from the long and flexible to the short and rigid (phage morphotypes are reviewed in Ackermann, 2007). Their release from the host cell can be a rapid lysis event, a budding of mature particles (reviewed in Casjens & King, 1975) or even an extrusion process (reviewed in Marvin & Hohn, 1969), occurring minutes or generations after infection. In short, bacteriophages are incredibly diverse and, as they lack any universal common genetic elements, their classification has proven particularly difficult (Ackermann, 2011). Due to this variety and their small size, they were discovered, studied, and continue to be characterized primarily by their interaction with their hosts, which I will discuss in more detail.

1.2 Bacteriophages: discovery & impact

The term *bacteriophage* was coined by Félix d'Hérelle in 1917, upon discovering distinct zones of clearing in dysentery bacilli growth caused by an agent that could pass

through a fine porcelain filter. Even heavily diluted agent yielded equivalent zones of clearing. He deduced from this that the agent responsible for the lysis was replicative (d'Hérelle, 1917) and was, in his words, "an invisible microbe, a filterable virus, but a virus parasitic on bacteria" (Duckworth, 1976 cites d'Hérelle, 1949). Frederick Twort is co-credited with the discovery of bacteriophages in 1915. He believed that this infectious agent was probably an “enzyme with the power of growth” (Duckworth, 1976 cites Twort, 1915). Both the nature of bacteriophages and the credit for their discovery caused considerable controversy over the next 20 years (Van Helvoort, 1992; Duckworth, 1976).

d’Herelle went on to champion bacteriophage research, conducting many experiments investigating their therapeutic potential as antimicrobial agents (reviewed in Summers, 2001). This “phage therapy” met with some success and popularity, but was severely impeded by poor understanding of bacteriophages. As an example, phenol, which denatures protein, was added to phage preparations as a preservative (Housby & Mann, 2009). It was not until 1940, with the advent of the electron microscope, that phage particles were observed (Ruska, 1940). Shortly after this breakthrough penicillin entered mass production (Moyer, 1948) and research into bacteriophages for phage therapy fell by the wayside in the west. The emergence of bacteria resistant to even the most potent antibiotics and a shortage of new antimicrobials have resulted in a revival of this field in the 1990s.

In the intervening years, bacteriophage research rose to the forefront of genetics. These early years of fundamental phage research were centered around the “American Phage Group”, a loose association of scientists shaped largely by a summer course started by Max Delbrück at Cold Spring Harbor in 1945 (Stent, 1992). Many of the iconic

discoveries in molecular biology arose from the efforts of this group. The idea that mutations arise in the absence of selection was established by monitoring bacterial resistance to phage infection (Luria & Delbrück, 1943). The existence of recombination was established using mutants of *Escherichia coli* phage T2 (Hershey, 1946), which led to Meselson and Weigle's famous experiments demonstrating the involvement of double-strand DNA breaks in recombination (Meselson & Weigle, 1961). Work on phage T2 was also key to demonstrating that DNA was the genetic material (Hershey & Chase, 1952), and provided the first clues to the existence of mRNA from the increase in cellular RNA following phage infection (Hershey *et al.*, 1953; Volkin & Astrachan, 1956). The name "messenger" RNA arose from a characterization of gene regulation in phage λ (Jacob & Monod, 1961). The triplet nature of the DNA code was established (Crick *et al.*, 1961) by experiments taking advantage of the original techniques for mapping genetic mutations using *E. coli* phage T4 (Benzer, 1955).

The phage course was last offered in 1961 (Stent, 1992), marking the end of an era for the American phage group. Despite this, phage research remained a crucial component of fundamental molecular biology. As examples, phage experiments showed the possibility of integration (and excision) of genes to/from genomes (Campbell, 1961), proved that DNA and protein are collinear (Sarabhai *et al.*, 1964), demonstrated DNA recognition and binding by repressors (Ptashne, 1967), established the discontinuous nature of lagging-strand DNA synthesis (Okazaki *et al.*, 1968) and revealed the existence of chaperones involved in protein folding (Georgopoulos *et al.*, 1973). Phage genomes were the very first genomes sequenced: first that of *E. coli* RNA phage MS2 (Fiers *et al.*,

1976) and then that of *E. coli* DNA phage φX174 (Sanger *et al.*, 1978). The first synthetic genome was that of *E. coli* phage φX174 (Smith *et al.*, 2003).

In addition to an undeniable relevance as models and tools for genetic research, the biological impact of bacteriophages is tremendous. With estimates of global phage counts of ~10^{30} (Chibani-Chennoufi *et al.*, 2004), they are the most abundant biological entities on the planet. Bacteriophages therefore represent a colossal nutrient and genetic reservoir (Wilhelm & Suttle, 1999). The genetic diversity of phages is so high we have sampled only an estimated 0.0002% (Rohwer, 2003).

1.3 Bacteriophage interactions

Bacteriophages and their hosts share a long evolutionary history and are competing in what is probably the longest running and fastest moving evolutionary "arms race" on earth. Globally, bacteria become infected by phages at a rate of 10^{25} per second (Lima-Mendez *et al.*, 2007) and it is estimated that 15% of the marine bacterial population is lysed daily (Suttle, 1994; Chibani-Chennoufi *et al.*, 2004). It should come as no surprise, then, that the interactions between phage and host have resulted in many remarkable adaptations for the survival of both.

1.3.1 Phage-host interactions

For replication, a phage must adsorb to the host cell, inject its nucleic acids, replicate its genome, package its genome and exit the cell. Host cells have found ways to prevent all of the above (Labrie *et al.*, 2010) by creating physical barriers to adsorption,

by modifying or losing receptors, by destroying invading DNA, by modifying or losing intra-cellular proteins required by the phage for replication or assembly, and even by committing "altruistic suicide" to reduce phage spread through a population (Forde & Fitzgerald, 1999; Fineran *et al.*, 2009).

Perhaps the best-known means through which bacteria protect themselves from phages are the restriction/modification systems (RMS). Their initial discovery arose from the observation that phages replicating in some hosts could gain the ability to successfully infect otherwise "restrictive" hosts, while those grown in certain permissive hosts lose the ability to infect these same "restrictive" strains (Bertani & Weigle, 1953; Luria & Human, 1952). Werner Arber determined this had to be a means to restrict the integration or replication of foreign genetic material without interfering with "the potentially beneficial genetic exchange between cells of the same strain" (Arber, 1965). He established that the system had to have two parts: "r" or restriction (resistance to foreign DNA) and "m" or modification (the ability to confer resistance to r). We now know that the former are restriction enzymes, nucleases capable of sequence-specific cleavage of DNA, and the latter are modification systems capable of modifying bases to mask the sites recognized by the restriction enzymes. Phages can evade these systems by modifying their own DNA to avoid recognition by restriction enzymes, and in some cases encode nucleases to degrade host DNA (e.g. phage T4, Kutter & Wiberg, 1968) without endangering their own (modified) genome.

In 2007 an even more specific "restriction" system was characterized. Several bacterial genomes possess large arrays of "clustered, regularly-interspaced, short palindromic repeats" (CRISPRs) separated by spacers of ~20 to 30 nucleotides of DNA

(Jansen *et al.*, 2002). These spacers were observed to be homologous to sequences from bacteriophages or plasmids (Bolotin *et al.*, 2005; Mojica *et al.*, 2005; Pourcel *et al.*, 2005). These CRISPR arrays were shown to have a function in bacterial immunity (Barrangou *et al.*, 2007). The acquisition of new spacer sequences in the array, homologous to phage sequences, confers resistance to a phage bearing that same sequence. If the phage carries a single point mutation in the targeted sequence, the conferred immunity is bypassed (Barrangou *et al.*, 2007). This phage resistance system also proved effective against plasmids, impeding conjugation and transformation of plasmids bearing the targeted sequences (Marraffini & Sontheimer, 2008). The field of CRISPR study has since exploded, with researchers around the world investigating acquisition (Swarts *et al.*, 2012; Datsenko *et al.*, 2012; Fineran & Charpentier, 2012), chromosomal-targeting (Vercoe *et al.*, 2013), CRISPR classification (Makarova *et al.*, 2011), alongside many other aspects of CRISPR biology. Perhaps unsurprisingly, this system has been co-opted by phages; a recent study (Seed *et al.*, 2013) identified a phage infecting *Vibrio cholerae* that carries a functional CRISPR array targeting host genes involved in defense against phages.

An important example of the interactions between phage and host is the existence of lysogeny. Many phages are capable of lying effectively dormant in the host, replicating only as part of the genome of the host. While the concept of lysogeny dates back to 1921 (reviewed in Lwoff, 1953), it was not until 1950 that the term probacteriophage (or prophage) came to be, defined as "the form in which lysogenic bacteria perpetuate the power to produce bacteriophage" (Lwoff, 1953). The system by which *E. coli* phage λ

regulates its integration, dormancy and subsequent activation and excision is among the best-studied models of gene regulation (Ptashne, 2004).

The importance of the existence of prophages rose to greater prominence with the advent of genome sequencing. A large proportion of bacteria have recognizable prophages (>60% of full genome sequences, Casjens, 2003), which can compose as much as of 20% of a bacterium's genome. This can have important consequences for the host through phage-phage interactions (see 1.3.2), by offering a competitive advantage to the host (see 1.3.3) or by leading to acquisition and modification of genes (see 1.4.1).

1.3.2 Phage-phage interactions

Phages outnumber their hosts 10:1 (Brüssow & Hendrix, 2002) and are regularly competing for hosts and interacting with each other. The earliest characterized interaction between phages is the observation that lysogenic bacteria do not experience lysis when infected with homologous bacteriophages (Lwoff, 1953 cites Wollman, 1938). This immunity, now known as superinfection immunity, is a consequence of the lysogenic phage producing a repressor protein which serves both to maintain its dormancy and prevent an incoming phage from turning on the genes required for replication. The best-studied model for this mechanism is bacteriophage λ, which, when integrated, produces only the repressor protein CI. CI acts on the P_L and P_R promoters of phage λ, preventing entry into the lytic cycle as well as preventing of any other λ phage DNA that enters the host from integrating or entering a lytic cycle of its own (Ptashne, 2004). While this mechanism of exclusion is thought to provide a benefit to the host (Brüssow & Kutter,

2005), it is unclear how effective a means of protection this is, given its specificity and the diversity of bacteriophages infecting each host (Moineau & Lévesque, 2005).

Another similar system of immunity is that of superinfection exclusion, whereby an infecting phage causes some change in the host cell that actually prevents phage adsorption or the translocation of the phage genome into the cell. One example of a phage capable of this exclusion is *Salmonella* phage P22, which encodes two separate exclusion systems (Walsh & Meynell, 1967; Susskind *et al.*, 1971) that prevent co-infection but have no effect on phage DNA introduced by conjugation (Nagaraja Rao, 1968). The P22 exclusion system is not specific to DNA sequences at all; it will prevent transduction of any DNA carried by P22 particles (Ebel-Tsipis & Botstein, 1971). These systems appear to have a broader "immunity" than superinfection immunity, as P22 can exclude several *Salmonella* phages including L, MG40, and MG178 (Susskind *et al.*, 1974). λ lysogens, which only confer superinfection immunity to other λ phages, can exclude phages P22 and L (Susskind & Botstein, 1980).

The interaction between phages can be far more intricate than simply competition for hosts. The *E. coli* phage P4 was discovered in 1963 (Six & Klug, 1973 cites Six, 1963), and found to be a "satellite phage" that requires a P2-related helper-phage to produce phage progeny. The two phages share almost no genomic similarities (Lindqvist *et al.*, 1993). While P4 can maintain itself in the host as an integrated lysogen (Calendar *et al.*, 1981) or autonomous replicon (phagemid) (Goldstein *et al.*, 1982), it requires the structural genes encoded by P2 to form a virion and escape the cell. It encodes a protein that is able to modify the P2 capsid to create smaller head structures suited to packaging

its smaller genome (Shore *et al.*, 1978). As the P4 genome does not encode the majority of its structural components, its existence blurs the line between phage and plasmid.

Further complicating the definition of "bacteriophage" are the *Staphylococcus aureus* pathogenicity islands (SaPIs), regions of the *S. aureus* genome that are able to take advantage of phage infection to be packaged by the phage-encoded particle. This mobility was first observed in 1998 (Lindsay *et al.*, 1998), where infection by the generalized transducing phage 80α resulted in a high efficiency transfer of SaPIs into *recA*$^-$ recipients. In addition to encoding toxins, from which they derive their name, SaPIs encode the proteins to excise from the host's genome, replicate, hijack the helper phage's capsid structures and preferentially package themselves into the heads (Christie & Dokland, 2012). The excision and replication processes can only occur in the presence of the helper-phage (Úbeda *et al.*, 2008). This hijacking process significantly reduces burst sizes and so may serve as a host defense mechanism (Ram *et al.*, 2012).

1.3.3 Benefits to the host

The interactions between host and phage are not always purely antagonistic. Bacteriophages can provide several benefits to the host, in addition to the aforementioned immunity to other phage infections and being co-opted to move genes. Bacteriophages act not only as a tremendous selective pressure on bacteria, driving evolution, but also as a means of introducing new genetic diversity. Integration into bacterial genomes can disrupt genes and provide a framework for changes to genome architectures (Brüssow *et al.*, 2004). Infection by phages can also offer opportunities to introduce new genes through transduction (see 1.4.1).

The benefits can be more direct. A remarkable finding in 1975 was that λ lysogens, under certain growth conditions in which the phage was not induced, could greatly out-compete non-lysogens (Edlin *et al.*, 1975). The mechanism for this is unclear, but similar findings have suggested a role for prophages in regulating metabolic functions (Chen *et al.*, 2005). This has led to the hypothesis that prophages can directly benefit host survival by suppression of unneeded metabolic activities (Paul, 2008).

There are several instances of bacteria making use of functions encoded by prophage elements to improve their fitness. This can include the phage encoding a variety of genes beneficial to the host (e.g. Wang *et al.*, 2010), the longest-studied of which is the production of phage-like bacteriocins to kill competing microorganisms. Bacteriocins are polypeptides with antimicrobial action, typically restricted to affecting only organisms closely related to the producing species (reviewed in Daw & Falkiner, 1996). Some bacteriocins appear to be derived from phages and can range from tail-like structures (Strauch *et al.*, 2001; Thaler *et al.*, 1995; Jabrane *et al.*, 2002) to complete phage-like virions (Seaman *et al.*, 1964; Schwinghamer *et al.*, 1973). The best-studied of these is *Bacillus subtilis* PBSX, a phage-like particle whose production can be induced and which acts as a bacteriocin, killing sensitive *B. subtilis* cells (Okamoto *et al.*, 1968). The PBSX particles appear to package host DNA with no preference for phage sequences and are incapable of injecting the packaged DNA into other cells (Okamoto *et al.*, 1968). In addition to these deficiencies relative to a phage, they are encoded by an 18-kb region (Wood *et al.*, 1990) but are capable of packaging only 13 kb of DNA (Anderson & Bott, 1985). They are clearly non-replicative, and thought of as derived from (or being) a "defective phage" (Okamoto *et al.*, 1968). Production of PBSX appears to be a closely

regulated event that is lethal to the producing cell (McDonnell *et al.*, 1994), involving phage lysins encoded by PBSX (Longchamp *et al.*, 1994).

Another class of phage-like particles are gene transfer agents (GTAs) (Lang *et al.*, 2012) (see 1.5.3). In contrast with PBSX, these particles are capable of transferring host DNA to related cells in a process analogous to transduction (see 1.4.1) and do not appear to act as bacteriocins. The benefits to the host of carrying GTAs are not as clear, but appear to be linked to horizontal gene transfer.

1.4 Genetic exchange in bacteria

There is no doubt that bacterial genomes have been shaped by lateral or horizontal gene transfer (HGT). The techniques for detecting HGT events in genomes vary widely, resulting in reported estimates that 0 to 16% of the genomes of bacteria (Ochman *et al.*, 2000) or 0.5% to 25% of their open reading frames (ORFs) (Nakamura *et al.*, 2004) were "recently" acquired through HGT, that 50% of extended gene families in cyanobacteria have a history of HGT (Zhaxybayeva *et al.*, 2006) or that, cumulatively, ~81% of genes were involved in lateral gene transfer at some point (Dagan *et al.*, 2008). As the evolutionary histories of any given gene do not necessarily match those of the organism possessing them, there is an increasing movement away from representing organism phylogenies as a "Tree of Life", which is inherently constrained by assumptions of vertical transmission, to representing them as webs or networks (Doolittle & Bapteste, 2007; Olendzenski & Gogarten, 2009). This has met with considerable resistance

(Kurland *et al.*, 2003; Ge *et al.*, 2005; Riley & Lizotte-Waniewski, 2009) and engendered debates about the actual impact of HGT (Puigbò *et al.*, 2010).

1.4.1 Transduction

One of the means by which DNA can be transferred between bacteria is transduction. This is the transfer by phage particles of DNA other than that of the phage. Zinder and Lederberg detected transfer of genetic markers between cultures which shared a broth but in which the cells could not come into direct contact with one another (Zinder & Lederberg, 1952). This transfer could occur even after nuclease digestion of the shared medium. They dubbed this exchange of genetic material through a filterable agent "transduction". The mixture, it turns out, was serendipitous: the "donor" strain was lysogenic and produced a phage capable of generalized transduction.

Generalized transduction is the term for phage-mediated transfer of any genetic marker in the host chromosome, constrained only by the size of the marker relative to the packaging capacity of the phage. It is the result of headful packaging mechanisms where the phage mistakenly packages host DNA instead of its own (Ebel-Tsipis *et al.*, 1972; Ikeda & Tomizawa, 1965). This is in contrast with specialized transduction, where only specific genes, determined by properties of the phage, are packaged. For example, in *E. coli* phage λ, imprecise excision can result in the packaging of the *gal* (Morse *et al.*, 1956) or *bio* (Del Campillo-Campbell *et al.*, 1967) genes adjacent to the λ integration site. Deletion of the natural integration site can force λ integration into other sites, resulting in phages able to transduce other genes (Shimada *et al.*, 1972). Transducing λ particles carry

phage DNA as well as the adjacent host DNA. As the host DNA replaces phage DNA, these particles are typically defective and will not yield viable progeny.

In both specialized and generalized transduction, I have described a gene transfer event that happens when a phage particle "mistakenly" packages host DNA, resulting in phage particles capable of transduction but (usually) incapable of infection. Transduction is not always a mistake; some phages, such as *E. coli* phage Mu, always include host DNA in functional virions (reviewed in Mizuuchi & Craigie, 1986). When Mu packages the regions flanking its integrated genome, it can package and transduce almost any marker because it can transpose into any site in the host genome as part of its replication (Howe, 1973).

1.4.2 Other means of horizontal gene transfer

Transduction is not the only mechanism by which bacteria can acquire new genetic material. Classical examples include conjugation: the transfer of plasmid DNA mediated by cell-to-cell contact (Tatum & Lederberg, 1947) and transformation: the uptake of naked DNA from the environment through a specialized competence apparatus (Chen & Dubnau, 2004; Claverys *et al.*, 2009; Johnsborg *et al.*, 2007). In addition to these, more recent discoveries include transfer of genetic material through nanotubes (Dubey & Ben-Yehuda, 2011), genetic exchange mediated through vesicles (Mashburn-Warren & Whiteley, 2006; Chiura *et al.*, 2011), and temporary cell fusion (Rosenshine *et al.*, 1989).

1.4.3 The benefits of horizontal gene transfer

Despite the evidence of a significant effect of horizontal gene transfer on bacterial genomes (see 1.4.0), it is still unclear what the benefits, if any, of HGT are. Evidence of past transfer events is insufficient to argue that HGT is beneficial and would be selected for. In other words, "even if harmful exchange events were 100-fold more common than beneficial ones, we would only see the latter in genomes today" (Redfield, 2001). Many of the models used to explain the benefits of sex in eukaryotes require considerable adjustment to explain genetic exchange in bacteria (Vos, 2009). The best evidence for the benefits of HGT is the maintenance of the mechanisms for genetic exchange. Conjugation and transduction, however, are carried out by selfish and self-transmissible clusters of genes, and benefit the transferred genes over the host (Redfield, 2001). Competence, which is host-encoded, shares many regulatory cues with nutrient starvation and could potentially be a system to pick up environmental DNA as a nutrient or DNA repair mechanism rather than a means of genetic exchange (Finkel & Kolter, 2001). While they are poorly characterized, the existence and persistence of host-encoded gene transfer agents (GTAs) as apparent facilitators of HGT may prove to be a strong argument in favor of the benefits of HGT; they are neither selfish nor plausibly involved in nutrition.

1.5 *Rhodobacter capsulatus* and GTAs

Rhodobacter capsulatus was the first organism in which a GTA was identified (Marrs, 1974), and is the source of the archetypal GTA, "RcGTA".

1.5.1 *Rhodobacter capsulatus*

R. capsulatus, formerly *Rhodopseudomonas capsulata* (Imhoff *et al.*, 1984), is a gram-negative, rod-shaped bacterium of the class α-proteobacteria and the order *Rhodobacterales*. The *R. capsulatus* strains initially studied were isolated in pasture and agricultural soil, sewage settling ponds, stagnant cisterns, lakes, and brackish waters of the Baltic Sea (Weaver *et al.*, 1975). It is termed a purple nonsulfur bacterium as it is capable of anoxygenic photosynthesis where the electron donor is not sulfur. It was deemed useful for laboratory study as it was hardy, capable of fast growth, amenable to long-term storage and its genome is easily manipulated (Weaver *et al.*, 1975). It has served as a model for study of anoxygenic photosynthesis (Pemberton *et al.*, 1998), of nitrogen fixation (Madigan, 1995) and because of its distinctive lipids (Imhoff & Bias-lmhoff, 2004).

A project to obtain the complete genome sequence for *R. capsulatus* SB1003 was commenced in 1992 (Fonstein *et al.*, 1992; Haselkorn *et al.*, 2001) and finally completed in 2010 (Strnad *et al.*, 2010). The genome consists of a 3.7-Mb circular chromosome and a 134-kb circular plasmid. Of particular interest are the presence of six CRISPR loci, one of which possesses 40 spacers (Grissa *et al.*, 2007), and the presence of 237 phage-related genes (Strnad *et al.*, 2010). I have identified clusters of phage genes large enough to be intact prophages (Figure 1-1), including the cluster now known to be responsible for the production of RcGTA (Lang & Beatty, 2000).

1.5.2 Phages, prophages and bacteriocins in *R. capsulatus*

R. capsulatus is a species originally classified, in part, by susceptibility to a group of purified phages (Weaver *et al.*, 1975). These 16 phages, one of which had been identified previously (Schmidt *et al.*, 1974), are the only literature references to phages infecting *R. capsulatus*. The majority of strains tested showed no evidence of lysogeny, although strain B10 was lysogenic for two phages with detectable lytic activity (Weaver *et al.*, 1975). The genome-sequenced strain SB1003 was derived from B100, a B10 derivative "cured" of these two phages (Solioz, 1975) by unreported means.

An early study of *R. capsulatus* identified bacteriocin production in all 33 strains tested (Wall *et al.*, 1975a), suggesting that antimicrobial activity is common in this species. These bacteriocins were distinguishable from phages and not associated with RcGTA production.

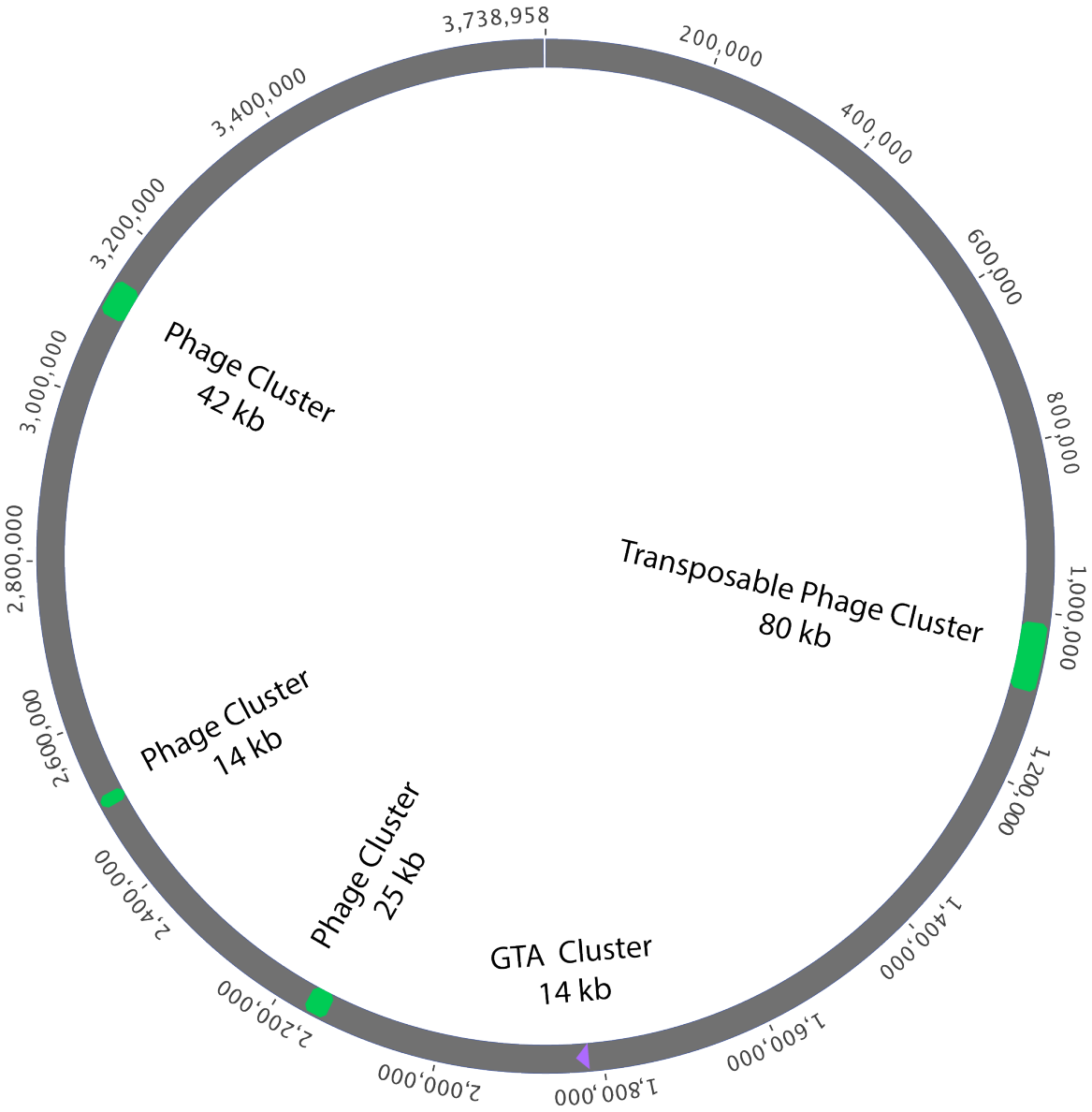

Figure 1-1: Prophage elements of *R. capsulatus* SB1003. The circle is a representation of the SB1003 chromosome, with base numbers from base 1 (white line) labeled. Each green region represents a cluster of phage-derived genes likely to be a prophage or prophage remnant, with the size of the cluster identified. The large (80 kb) cluster contains a region resembling a transposable prophage. The structural cluster encoding the phage-like gene transfer agent (GTA) is coloured in purple.

1.5.3 RcGTA: The early years

The initial discovery of a means of transferring various genetic markers between *R. capsulatus* strains established that the transfer shared many properties with transduction (Marrs, 1974). The gene transfer activity did not require cell-to-cell contact, it was retained in cell-free filtrates, and it was impervious to treatment with DNase. It was distinguished from transduction partly because of the lack of detectable lysis and the inability to induce particle production with mitomycin C, but also because the agent's sedimentation constant revealed that it was considerably smaller than any known phage (Marrs, 1974). The ability to produce and receive genes from RcGTA varied across *R. capsulatus* strains (Wall *et al.*, 1975a), but the ability to produce RcGTA could not be transferred by RcGTA, at least not in a single transfer event (Yen *et al.*, 1979 cites Jasper & Marrs, unpublished). Markers could be transferred at rates as high as 4×10^{-4} per recipient cell (Solioz *et al.*, 1975). The DNA content was double-stranded, but clearly much smaller than that of known phages (approx. $3.6 \times l0^{6}$ Da) (Solioz & Marrs, 1977), and the kinetics of release of RcGTA particles from cells did not resemble those of induction of lysogenic phages (Solioz *et al.*, 1975). The high transduction rates and ability to co-transduce markers ~5 genes apart made this GTA a useful tool for genetic mapping and manipulation (Yen & Marrs, 1976).

The creation of an RcGTA overproducer strain made the further characterization of RcGTA possible (Yen *et al.*, 1979). The particles were visualized by electron microscopy (Figure 1-2) and resembled small tailed phages. The DNA within the particles possessed a complexity comparable to that of the entire *R. capsulatus* genome

(as determined by hybridization kinetics and restriction digests), suggesting the particles could package and transfer host genes at random. The transfer of DNA was found to be mediated by recombination of the packaged DNA using mechanisms present in the recipient cell (Genthner & Wall, 1984). Although these GTAs were used to generate linkage maps (Wall & Braddock, 1984) and mutagenize *R. capsulatus* (e.g. Wong *et al.*, 1996), their properties were not investigated again in detail for some time.

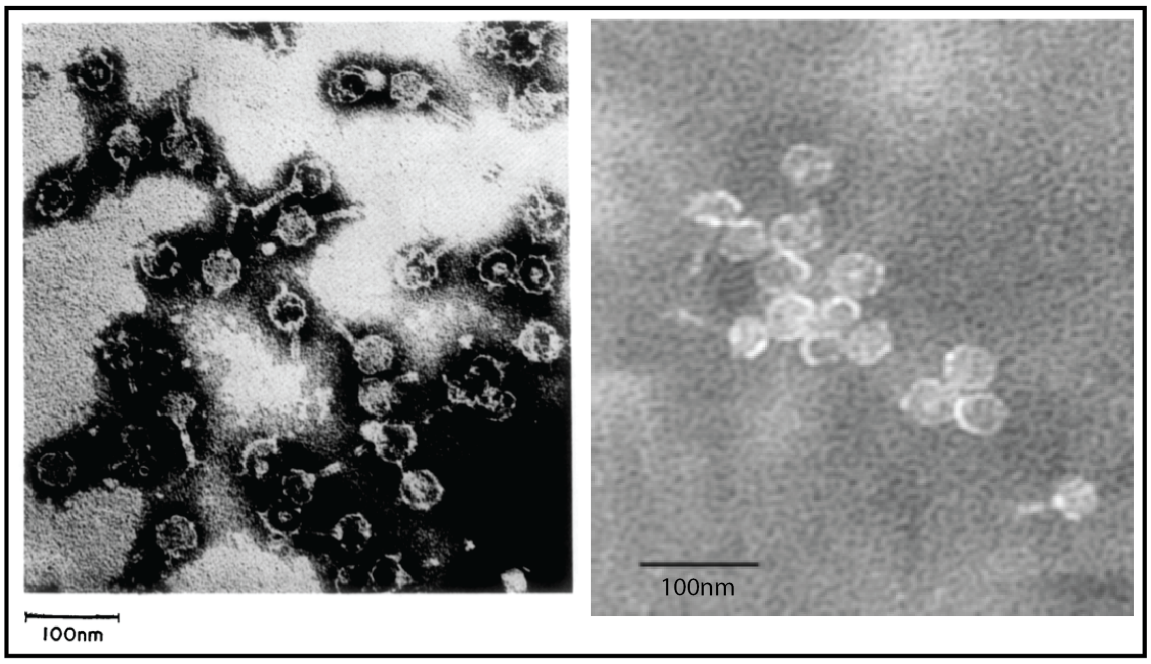

Figure 1-2: Electron micrographs of RcGTA particles. The left panel, from (Yen *et al.*, 1979), and the right panel, from (Chen *et al.*, 2008), both represent particles negatively stained with phosphotungstic acid. The heads are approximately 30 nm in diameter, and the tails 50 nm in length. Permission to reprint these micrographs was obtained from the copyright holders.

1.5.4 RcGTA: In recent years

In 2000, Lang & Beatty published the first characterization of the genes required to produce RcGTA. By transposon mutagenesis of an RcGTA overproducer strain, Y262, they were able to find a cluster of genes ~15-kb in size which appeared to encode the RcGTA particle (Figure 1-3). Organized in a head-to-tail fashion similar to that of the genomes of most tailed phages (Casjens *et al.*, 1992), it lacked genes coding for identifiable replication and integration functions. The absence of these genes is consistent with RcGTA's ability to package host DNA that recombines into genomes using host mechanisms (Wall & Braddock, 1984). The cluster also lacks readily identifiable tail fibre genes, lysis genes, and overall is less than half the size of related phages (e.g. ~40 kb for *E. coli* phage HK97, Juhala *et al.*, 2000). As RcGTA is encoded by a ~15-kb cluster and packages only ~ 4 kb, it is non-replicative and, by definition, not a phage. A recent study purified RcGTA particles and identified the associated proteins (Chen *et al.*, 2008). In addition to nine proteins encoded by the structural cluster, four other proteins encoded elsewhere in the genome co-purified with the particles.

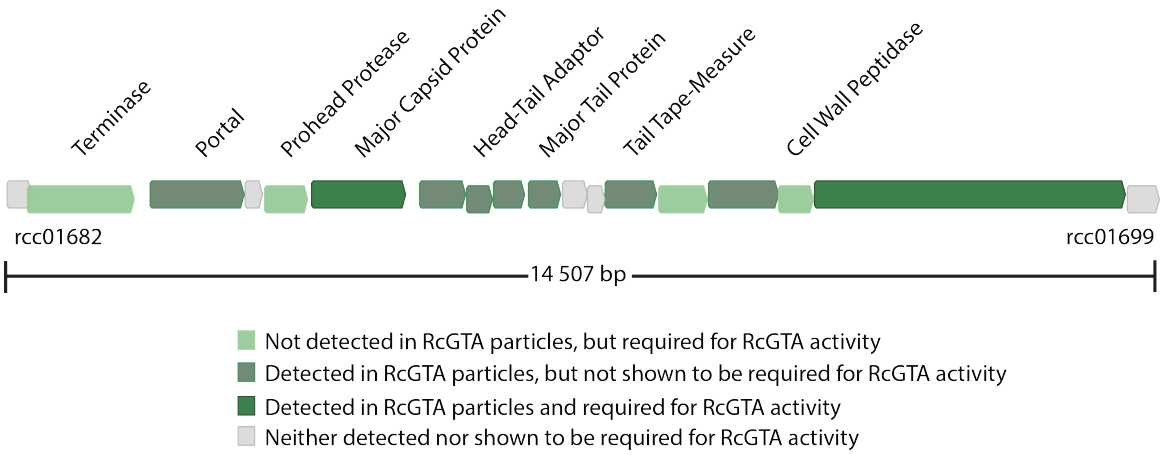

Figure 1-3: The RcGTA structural gene cluster. All genes/ORFs are transcribed from left to right. Vertical offset in neighbouring genes indicates different reading frames. The cluster spans from gene *rcc01682* to *rcc01699*, a total of 14 507-bp. Genes with homology to a phage gene are annotated by predicted function. Proteins detected in RcGTA particles are as reported by (Chen *et al.*, 2008), while those shown to be required for RcGTA activity are as reported by (Lang & Beatty, 2000; Fogg *et al.*, 2012).

In the same Lang & Beatty study, a transposon mutant with drastically reduced gene transfer activity was found to be a disruption of the *R. capsulatus* homologue of the *Caulobacter crescentus* cell-cycle regulator *ctrA* (Lang & Beatty, 2000). By analyses of northern blots and reporter gene constructs, it was determined that CtrA was responsible, in part, for the regulation of RcGTA production through the regulation of expression of the RcGTA structural gene cluster (Lang & Beatty, 2000). The creation of site-directed *ctrA* mutants revealed that both the phosphorylated and un-phosphorylated forms of CtrA are involved in RcGTA regulation (Mercer *et al.*, 2012). A genome wide transcriptomic and proteomic analysis of a *ctrA* null mutant revealed 227 dysregulated genes (Mercer *et al.*, 2010), several of which were investigated in detail and also found to be involved in RcGTA production (Mercer *et al.*, 2012).

A second regulatory pathway, involving quorum sensing, was found to be involved in regulating RcGTA production (Schaefer *et al.*, 2002). Quorum sensing is a mechanism thought to serve as a means to sense bacterial density (Fuqua *et al.*, 1994), or, perhaps, measure local diffusion rates (Redfield, 2002). It is frequently involved in coordinating behaviours in large populations of cells, as exemplified by the fluorescence of *Vibrio* (now *Aliivibrio*) *fischeri* (Nealson & Hastings, 1979) or competence of *Streptococcus pneumoniae* (Morrison & Lee, 2000) upon reaching a requisite cell density. A knockout of the acyl-homoserine-lactone (HSL) synthesis gene *gtaI* drastically reduced RcGTA production, which could be restored by addition of HSL to the cultures. The *gtaI* gene was found to be paired with a quorum sensing receptor, *gtaR*, which negatively regulates *gtaI* (Leung *et al.*, 2012). This same quorum-sensing pair appears to regulate extra-cellular polysaccharide synthesis and capsule production, the latter of which was found to be required for efficient adsorption of RcGTA particles to recipient cells (Brimacombe *et al.*, 2013).

Overall, these aspects of regulation show a remarkable integration of RcGTA production (and recipient ability) with host regulatory mechanisms. This regulation is complex and involves multiple pathways. To date, no protein interacting directly with the RcGTA promoter has been identified.

1.5.5 RcGTA phylogenetics

An investigation into the conservation of RcGTA genes in other organisms (Lang *et al.*, 2002) followed shortly after the discovery of the structural cluster in *R. capsulatus*. The authors identified conserved GTA structural gene clusters in several

Rhodobacterales, Rhodospirillales, *Caulobacterales* as well as *Rhizobiales,* and posited a "predominantly vertical descent from a GTA-like entity in an α-proteobacterial progenitor". This hypothesis was supported by a further analysis that suggested that RcGTA is likely the remnant of an ancestral prophage that existed prior to the divergence of the major phylogenetic divisions of α-proteobacteria (Lang & Beatty, 2007). GTA phylogenies revealed that the gene encoding the major capsid protein could yield phylogenies comparable to those produced by 16S rDNA. The authors suggested that since RcGTA seems to have existed since the origin of the α-proteobacteria and is still functional in *R. capsulatus,* it probably confers a selective advantage. Lending further support to this hypothesis, another member of the *Rhodobacterales (Silicibacter* (now *Ruegeria*) *pomeroyi* DSS-3) has since been shown to posses GTA-mediated gene transfer activity (Biers *et al.*, 2008).

1.5.6 Other GTAs

Gene transfer activity by particles too small to be functional phages is not unique to the *Rhodobacterales.* To date, four other gene transfer agents have been identified: VSH-1 of the spirochaete *Brachyspira hyodysenteriae* (Humphrey *et al.*, 1997), Dd1 of the δ-proteobacterium *Desulfovibrio desulfuricans* (Rapp & Wall, 1987), VTA of the archaeon *Methanococcus voltae* (Bertani, 1999) and BaGTA of *Bartonella grahamii* (Berglund *et al.*, 2009). None of these bear any known genetic relationship to one another or to RcGTA, but all are small tailed-phage-like particles capable of transferring small (4 to 14 kb) fragments of host DNA (reviewed in Lang *et al.*, 2012). In every case, they appear to be encoded primarily by clusters larger than the DNA molecule they can

package in a single particle; VSH-1 is actually encoded over multiple loci (Stanton *et al.*, 2009).

Given that GTAs are present in multiple phyla of prokaryotes, it is possible that a large proportion of the phage particles observed in environmental samples may in fact be GTAs. This proposal may explain the high proportion of cellular genes found in viral metagenomes (Kristensen *et al.*, 2010).

1.6 Research goals

GTAs are poorly understood. Even for RcGTA, one of the best-studied GTAs, the "life cycle" is still an enigma (Figure 1-4), its regulation is incompletely characterized (1.5.4), its release occurs through unknown mechanisms (1.5.3), and its adsorption and translocation occur at unknown receptors. The structural cluster appears to be missing key elements for its function (1.5.4), including a lysin, holin, and tail fibres. With so little information about its cost and benefit to the host, it is difficult to establish the evolutionary purpose served by the RcGTA particles.

In this thesis, I searched for additional genes encoded outside the RcGTA structural gene cluster required for RcGTA gene transfer activity. Because of the propensity for "satellite" or otherwise defective phages (1.3.3) to take advantage of a "helper" phage, I began my search in the other prophage clusters of *R. capsulatus* SB1003 (Chapter 2). *R. capsulatus* SB1003 has a history of association with bacteriophages (1.5.2), including at least four potentially intact prophage clusters. These clusters could be contributing genes necessary for RcGTA production.

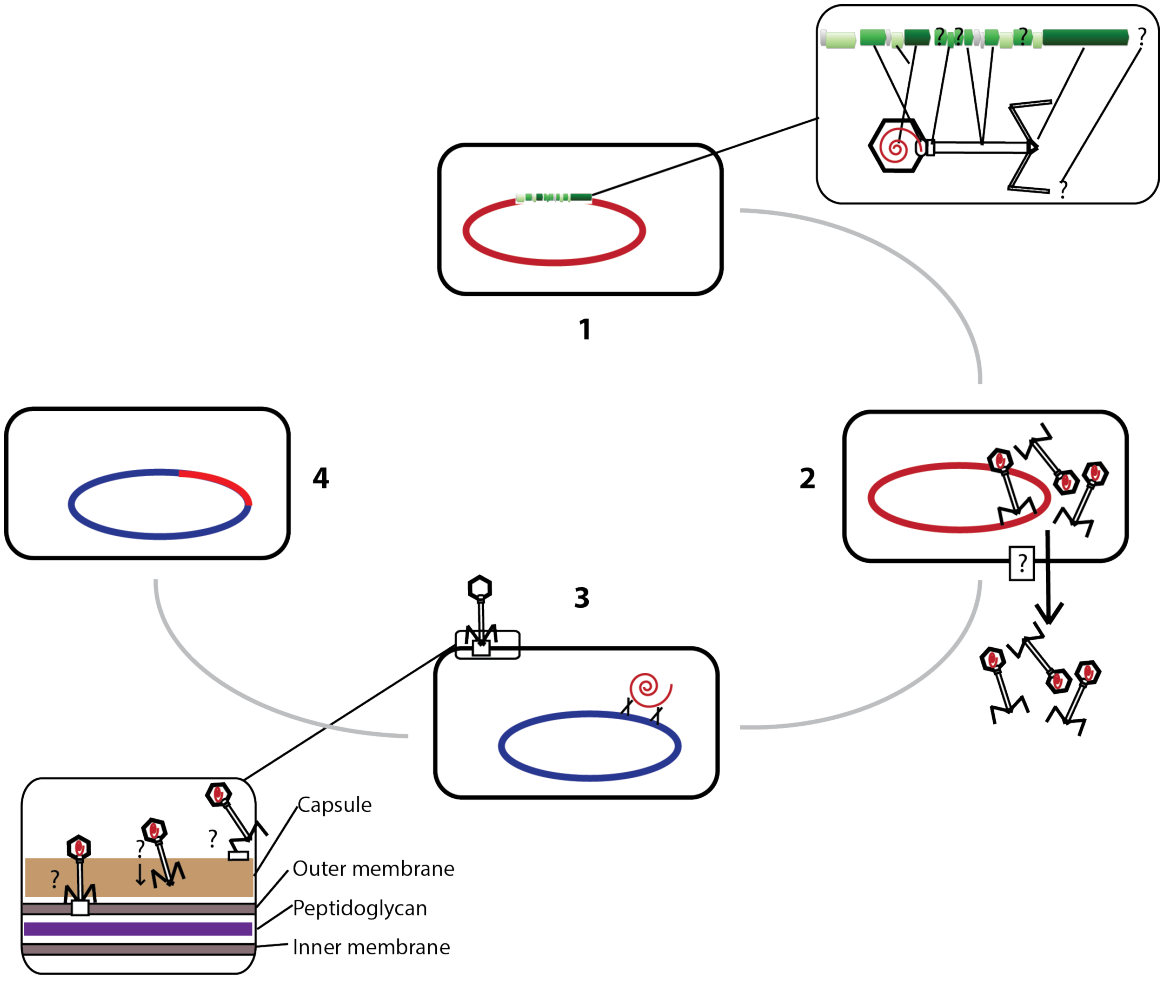

Figure 1-4: The "life cycle" of RcGTA. **1.** The production of RcGTA particles inside the *R. capsulatus* cells is primarily encoded by the RcGTA structural cluster, although some necessary genes appear to be absent from the cluster (e.g. tail fibres). The DNA packaged within the particles is host-derived. **2.** The release of particles from the cells is by an unknown mechanism. **3.** The attachment to the recipient cells. The adsorption is to an unknown site; the penetration of the capsule and the specific attachment to a receptor on the outer-membrane are suspected, but not known. The subsequent recombination event depends upon host factors. **4.** The resulting gene transfer recipient is a recombinant carrying DNA from the RcGTA producing cell.

As an alternate approach to discovering the "missing RcGTA functions", I investigated genes consistently co-regulated with the RcGTA structural gene cluster (Chapter 3). If any genes elsewhere in the genome are required for functional RcGTA production they are likely to be regulated in a manner similar to that of the structural cluster.

Using the information gleaned from the two approaches, I hoped to piece together an expanded "genome" of genes required for RcGTA production. The additional functions encoded by these genes would provide valuable insight into the evolutionary history (Chapter 3), as well as to the costs and benefits (Chapter 4) of RcGTA production.

Co-Authorship and Contributions

Chapter 2 consists of unpublished work, which will be modified into a manuscript for submission after this thesis is submitted for examination. I was responsible for the experimental design, in consultation with A.S. Lang. Contributions to the research in this chapter include: A.S. Lang's original findings of a link between phage and RcGTA production, those of students who made mutant strains I later tested for phage activity, and that of the Broad Institute where sequencing and preliminary annotation of our phage DNA sample took place. I carried out the isolation, characterization, annotation and subsequent deletion of the phages, as well as all the bioinformatic analyses included in this chapter. The work was carried out with support from the Broad Institute's Marine Virome Sequencing project, RDC Newfoundland, NSERC and the Memorial University of Newfoundland's School of Graduate Studies. This chapter makes multiple references to a published manuscript based on another portion of this work, published in collaboration with P. Fogg, E. Digby, and J.T. Beatty from the University of British Columbia (Fogg, P. C. M., A. P. Hynes, E. Digby, A. S. Lang & J. T. Beatty (2011) Characterization of a newly discovered Mu-like bacteriophage, RcapMu, in *Rhodobacter capsulatus* strain SB1003. *Virology* **421**: 211-221.), included as Appendix 2.

Chapter 3 represents a manuscript in preparation for submission. I was responsible for the experimental design, in consultation with A.S. Lang and R.G. Mercer. Contributions to the research in this chapter include: the creation of a mutant of *rcc00171* with help from Katherine Grebe, the creation of mutants of *rcc01079* and *rcc01080* with

help from Heidi Matchem and Marc Gruell, the creation of a mutant of *rcc01699* by Amanda Peach, and microarray data analysis and assistance with protein purification by Ryan G. Mercer. I performed the data analysis of the microarrays; performed all westerns, bio-assays, and inhibition assays used in this chapter; created the protein expression construct, purified the protein and performed all subsequent assays using it; and I performed all bioinformatic analyses included in the chapter. I authored the manuscript with considerable editorial input by A.S. Lang. The work was carried out with support from NSERC and the Memorial University of Newfoundland's School of Graduate Studies.

Chapter 4 represents a manuscript published in 2012 (Hynes, A. P., R. G. Mercer, D. E. Watton, C. B. Buckley & A. S. Lang, (2012) DNA packaging bias and differential expression of gene transfer agent genes within a population during production and release of the *Rhodobacter capsulatus* gene transfer agent, RcGTA. *Molecular Microbiology* **85**: 314-325). I was responsible for the experimental design, in consultation with A.S. Lang. Contributions to the research in this chapter include: the generation of the original gene expression microarray data, mutation of and preliminary findings related to *rcc00555* by Ryan G. Mercer; the DNA-packaging microarray experiment by David E. Watton; the generation of the *rcc02539* mutant by Colleen B. Buckley; and the ligation assays by A.S. Lang. I performed the array analyses, discovered the underpackaging phenotype, performed all the bio-assays, constructed the lacZ fusions and assayed them by flow cytometry, and performed all the assays characterizing SB555. I wrote the manuscript, with considerable editorial input by A.S. Lang. The final submission benefited greatly

from the comments and suggestions of the (anonymous) reviewers. The work was carried out with support from NSERC and the Memorial University of Newfoundland's School of Graduate Studies.

2 The Phages of *Rhodobacter capsulatus*

2.1 Introduction:

Rhodobacter capsulatus is an α-proteobacterium of particular interest as the source of the archetypal gene transfer agent, RcGTA. This small, host-encoded phage-like particle is coded for primarily by a ~14-kb structural gene cluster (Lang & Beatty, 2000), and is capable of transferring 4-kb stretches of host DNA to other *R. capsulatus* cells in a process analogous to transduction. It is non-replicative, and has been termed by several a "defective phage" (Solioz *et al.*, 1975; Yen *et al.*, 1979; Lang & Beatty, 2000; Redfield, 2001). The structural gene cluster is organized in a head-to-tail fashion (Lang & Beatty, 2000), akin to the organization of most tailed phages (Casjens *et al.*, 1992). Notably absent from this cluster are genes identifiable as being involved in replication, integration or excision. These are functions that would not be required by a non-replicative entity, so their omission is consistent with the observed transduction activity of RcGTA. Conspicuous in their absence are any genes predicted to encode a cellular release mechanism (e.g. a lysin) or tail fibres. An early electron micrograph of RcGTA particles showed the presence of tail fibres (Yen *et al.*, 1979), and it would be unprecedented for a tailed phage particle to be released from cells without lysis.

There are several known phages and phage-like elements that do not encode all their own functions but that can still produce functional phage particles. Among these are the "satellite phages", exemplified by phage P4 (Six & Klug, 1973). Phage P4 is capable of lysogenizing its host but is only capable of producing virions in the presence of a

phage of the P2 family. P4 hijacks several of the functions encoded by P2, including the capsid itself, to produce phage particles capable of infecting other cells (Shore *et al.*, 1978). Another similar example is that of *Staphylococcus aureus* pathogenicity islands (SAPIs) (Christie & Dokland, 2012), which, upon infection of the SaPI-containing cell with certain transducing phages, can replicate (Úbeda *et al.*, 2008) and be preferentially packaged by the phage (Christie & Dokland, 2012).

Two previous studies have investigated *R. capsulatus* phages: one investigated the host ranges of 16 phages in 33 *R. capsulatus* strains (Wall *et al.*, 1975a), and the other explored the energetics of infection by one of those phages (Schmidt *et al.*, 1974). The genome-sequenced strain of *R. capsulatus*, SB1003, is derived from strain B100, a strain supposedly "cured" of two prophages by unreported means (Solioz, 1975). However, the SB1003 genome sequence includes 237 phage-related genes (Strnad *et al.*, 2010) in at least four clusters large or complete enough to potentially be intact prophages (Figure1-1). The presumed "missing" RcGTA functions, the existence of satellite phages in other orders of Bacteria, and the presence of several identifiable prophage regions in the *R. capsulatus* genome prompted me to further investigate the phages of *R. capsulatus* and to explore potential RcGTA-phage interactions. I have attempted to isolate any functional prophages and cure them from *R. capsulatus* in order to identify phages or phage genes involved in RcGTA production. Any characterization of these phages would also help remedy a dearth of knowledge about α-proteobacterial phages. While many phages have been well characterized, the α-proteobacterial phages remain highly underrepresented. When I embarked upon this study in 2010, 880 phage genomes were available through PhAnToMe (*http://www.phantome.org/PhageSeed/Phage.cgi*) and 660 through EBI

(*http://www.ebi.ac.uk/genomes/phage.html*). Combined, the two databases contained only 9 unique α -proteobacterial phages. The same databases at time of writing (July 2013) contain 1184 and 1329 phage sequences, respectively, of which only 36 are from α-proteobacterial phages.

2.2 Experimental procedures

2.2.1 Bacteriophage production

I attempted phage induction in *R. capsulatus* strain SB1003 by damaging its DNA with DNA-damaging agents such as ultraviolet light and carbadox (0.1 to 5 mg ml^{-1}), by high (37°C) and low (20°C) temperatures, by growth under photosynthetic conditions, as well as growth in a minimal medium, RCV (Beatty & Gest, 1981). In the end, optimal yields were achieved using growth conditions found to favor aerobic RcGTA production. *R. capsulatus* cells were grown in 50 ml of complex YPS medium (Wall *et al.*, 1975a) at 30°C, placed in a 250 ml flask shaking at 225 RPM. The cultures were grown for 48 h, and phage particles were harvested (see 2.2.2). Strains used for this procedure are listed in Table 2-2.

2.2.2 PEG precipitation of phage particles

Cells were pelleted by centrifugation at 6800 *g* and the supernatant decanted. This process was repeated until the supernatant was clear. The final supernatant was treated with 1 mg ml^{-1} DNase I (Sigma-Aldrich, Oakville, Canada) for 1 h at 37°C according to the manufacturer's recommendations. NaCl was added to a concentration of 1 M, and

then centrifuged again. The supernatant was brought to 10% (w v^{-1}) PEG8000, incubated at 28°C for 2 h and the precipitate was collected by centrifugation at 6800 *g* for 20 min. The pellets were resuspended in G buffer (Solioz & Marrs, 1977) by gentle shaking (60 RPM) overnight at 4°C.

The above protocol was refined for later purifications by increasing the first spin to 8000 *g* and by the adding a filtration step in which the supernatant was passed through a 0.45-µm filter. This removed the need for additional spins and increased yields.

2.2.3 DNA extraction from phage particles

PEG precipitate suspensions were treated with 2 U DNase I and 1.2 U RNase A at 37°C for 30 min, followed by addition of 0.5 M EDTA (pH 8.0) to a final concentration of 0.005 M and heat inactivation at 75°C for 20 minutes. DNA was extracted using phenol:chloroform:isoamyl alcohol (25:24:1), precipitated with ethanol, and re-hydrated in pH 8.0 TE buffer or water. This protocol was refined as follows to increase the yields. After heat inactivation of the nucleases, SDS, EDTA and proteinase K (New England Biolabs, Pickering, Canada) were added to a final concentration of 0.5% w v^{-1}, 0.02 M and 0.05 mg ml^{-1} respectively, and incubated at 55°C for 1 h. This was followed by DNA purification as described above.

2.2.4 Separation of phage and RcGTA particles

PEG precipitate suspensions were loaded on top of a sucrose gradient created by the successive addition of equal portions 30%, 25%, 20%, 15% and 10% w v^{-1} sucrose in G buffer that had been left to stand overnight to linearize. The gradients were then

centrifuged at 50 000 *g* in a swinging-bucket rotor, and fractions were collected by drainage from a hole pierced in the bottom of the tube. The collected fractions were then centrifuged at 300 000 *g* for 1 h, and the resulting pellets resuspended in G buffer. Each resuspension was then subjected to DNA extractions (as described above) and western blotting with probing for the major capsid protein of RcGTA, as previously described (Mercer *et al.*, 2012).

2.2.5 Pulsed field gel electrophoresis (PFGE)

Visualization and sizing of DNA bands was performed by pulsed field gel electrophoresis (PFGE) on a Bio-Rad CHEF Mapper (Bio-Rad, Mississauga, ON) using their proprietary auto-algorithm for appropriate separation of bands. Gels were run using 1% pulsed field-certified agarose in 0.5x Tris-Borate-EDTA (TBE) buffer at 14°C.

2.2.6 DNA sequencing

Initially, purified phage DNA was digested with BamHI and the resulting restriction fragments were cloned and sequenced by traditional Sanger sequencing to confirm the presence of phage-like sequences. Undigested phage DNA purifications were then run on an agarose gel and phage DNA was extracted by electro-elution and concentrated with a centrifugal filter unit with a nominal molecular weight limit of 30,000 (Millipore, MA). The DNA concentration was determined by spectrophotometry and 280 ng was submitted to the Broad Institute (Cambridge, MA) for high-throughput (Roche 454) pyrosequencing (Margulies *et al.*, 2005). Gaps in the sequence were closed by PCR amplification across the gaps ("Gap" primers, Table 2-1), and Sanger sequencing

performed on the products by The Centre for Applied Genomics (TCAG) (Toronto, Canada). Sequenced data was obtained for two phages, which we named RcapMu and RcapNL.

Table 2-1: Primers and gene disruption sites used in this study

Primer name	Use	Primer sequence	Disruption site
2TNF	K.O. of RcapMu	ATCGTCGATCTCGGCGCGCA	StuI/BmgBI
2TNR		AGGGCTATGCGATTTCGCAGGGG	
TerF	RcapNL Screening & K.O. of RcapNL	GGACACCTCGCGGTTGGTGG	NruI
TerR		GCAGCATGATCGGCGACGGA	
CapsidR	RcapNL Screening	ACCGGCTTCTTGCGCAGCAT	
CapsidF		ATGGCGACAGCGGCCTGTTC	
GapRR	Closing RcapNL sequence	ATGCCGTGTCGAAGACCCCG	
GapRF		TCGTCAAGGGGAAACTTCCCGC	
GapLR	Closing RcapNL sequence	GGGAACGACAGCCGCTCGAT	
GapLF		GCATCACCAAAACGCACCCGC	
InvLR	Finding RcapNL ends	CGGTCTGTGGCGGCGAAGAT	
InvLF		CGCCTCAAATTCCCGCCCGT	
InvRR	Finding RcapNL ends	CGATACAGGACGGCGGCAGG	
InvRF		ATGACGACGAAGCGGGGTGC	

2.2.7 Bioinformatics

Sequence assembly was performed independently using Geneious Pro 5.3.4 (Drummond *et al.*, 2011) with settings for a high sensitivity (gap free) assembly. Annotations from the Broad institute and from RAST (Aziz *et al.*, 2008) were compared to automated Glimmer 3.02 (Delcher *et al.*, 2007) and Genemark (Besemer & Borodovsky, 2005) annotations, and to a manual annotation based primarily on assumed start codon preference, putative *R. capsulatus* ribosome binding site locations and ORF

length. Any predicted ORF was then compared to existing CDS databases by blastp and PSI-BLAST (Altschul *et al.*, 1990; Altschul *et al.*, 1997; Wheeler *et al.*, 2007), and the predicted start positions optimized by alignment with close homologues, where relevant. The RcapNL sequence is available in GenBank (GI: 461474957), as is the sequence for RcapMu (GI: 356870838). Annotations for hypothetical proteins were revised to "proteins of unknown function" if proteomic evidence of their expression was obtained (see below).

2.2.8 Proteomics

An existing *R. capsulatus* whole-cell proteomic dataset (Mercer *et al.*, 2010) was searched for peptides matching predicted RcapNL proteins. The peptide information can be accessed through *http://omics.pnl.gov/.*

2.2.9 Plaque assays

R. capsulatus cells of various strains (Table 2-4) were grown overnight in RCV minimal medium, pelleted and re-suspended in either G buffer or fresh medium. 100 µl of PEG precipitate resuspensions from 48 h aerobic cultures or filtrates from 48 h phototrophic cultures were incubated with 300 µl recipient cell suspension and incubated for 1 h at 35°C. The mixture was then combined with molten 0.5% YPS agar, and poured as an overlay above 1.5% YPS agar plates. Plates were grown for 48 h at 30°C and examined for plaques.

2.2.10 Phage & prophage DNA end characterization

I attempted PCR on restriction digests of chromosomal *R. capsulatus* SB1003 DNA with a variety of restriction enzymes, using primers reading in opposite directions from several sites in the RcapNL genome ("inv" primers, Table 2-1).

To characterize the DNA ends, RcapNL DNA purified from SBKMu cultures was treated with either T4 DNA polymerase (New England Biolabs) followed by T4 DNA ligase (New England Biolabs), or by a treatment with T4 DNA ligase only – each according to the manufacturer's recommendations. The ligations were then used as template for PCR using primers reading outwards from either end ("Inv" primers, Table 2-1) of the linear genome. The PCR products were sent to The Centre for Applied Genomics (TCAG), and sequenced by Sanger sequencing.

2.2.11 Phage "knockouts"

Phage knockouts were generated by disrupting genes thought to be essential for phage replication. The knockout strains, then, would be deficient for production of phage particles but not cured of the phages. The plasmid for making knockouts of RcapMu was constructed by amplifying the region of RcapMu containing the two transposase genes with the 2TN primers (Table 2-1). The PCR product was cloned into pUC19 (New England Biolabs) with the SphI site deleted by cleavage with SphI and subsequent blunting and re-ligation. This plasmid was digested with SphI and BmgBI, which released a 1.4-kb fragment of the RcapMu sequence which was then replaced with the 1.4-kb SmaI fragment of the kanamycin resistance-encoding KIXX cartridge (Barany, 1985). The resulting strain was called SBKMu. The plasmid for making knockouts of RcapNL was

constructed by amplifying the terminase region of RcapNL with the Ter primers (Table 2-1). The PCR product was cloned into pGem-T-Easy (Promega, Madisson, WI), and the 381-bp NruI fragment was replaced with either the KIXX cartridge SmaI fragment or the SmaI fragment of the spectinomycin resistance-encoding Ω cartridge (Prentki & Krisch, 1984). The resulting strains were called SBKNL (KIXX) or SB ΩNL

The plasmids carrying the disrupted genes were conjugated into an RcGTA overproducer, *R. capsulatus* DE442, using *Escherichia coli* C600 (pDT51) (Taylor *et al.*, 1983), and the disrupted genes were transferred by RcGTA into SB1003 with recombinants selected on the basis of acquired antibiotic resistance. I also created a strain, SBKMuΩNL incapable of producing both RcapMu and RcapNL. To confirm the phenotypes of all the knockouts, I analyzed purified phage DNA by PFGE (Figure 2-5), and visualized them using a gel documentation system. The images were adjusted for brightness and contrast, with all regions of each image manipulated equivalently, using Adobe Photoshop C.S. 6.0 (Adobe Systems Inc, San Jose, CA).

2.2.12 GTA activity assays

Gene transfer bio-assays were performed as described in "*Rhodobacter capsulatus* Gene Transfer Agent (RcGTA) Activity Bioassays" (Bio-protocols.org, available in Appendix II) (Hynes *et al.*, 2012). Normalized culture aliquots were used to inoculate anaerobic photoheterotrophic cultures in YPS medium that were grown for 48 h. The gene transfer activity was measured using the recipient strain DW5 (*puhA*$^-$) (Wong *et al.*, 1996), monitoring the gene transfer activity of cell-free filtrates by counting the number of DW5 colonies able to grow photoheterotrophically.

2.3 Results

2.3.1 Phage detection

PEG precipitations and DNA purifications using the initial technique (i.e. without Proteinase K addition) yielded an additional >10-kb band co-purifying with RcGTA-packaged DNA (4 kb) (Figure 2-1, Lane 2). Unlike RcGTA DNA, this larger fragment could be digested into fixed banding patterns (data not shown), leading me to hypothesize that it was phage DNA. Attempts at induction with ultraviolet light, carbadox treatment, or by modifying growth conditions yielded no appreciable increase in the intensity of the band. Furthermore, I observed that mutants deficient in RcGTA production failed to produce this additional DNA band (Figure 2-1, lane 4) in detectable amounts. Testing of numerous RcGTA production mutants (Table 2-2), including knockouts of genes regulating RcGTA production and knockouts of genes encoding the structure itself, consistently failed to yield any of this larger DNA band using the standard extraction protocol (Extraction 1, Table 2-2). Strains with gene knockouts that have no effect on RcGTA activity did produce the additional band (Table 2-2).

In an effort to increase phage and RcGTA DNA yields, purification protocols were modified to include a proteinase K digestion prior to phenol:chloroform:isoamyl alcohol extraction. This resulted in an increased intensity of the phage band (Figure 2-1, Lane 2 vs 1), and the detection of the phage band in RcGTA-deficient mutants (Figure 2-1, Lane 3; Table 2-2).

Sucrose gradient separation was used to discriminate between RcGTA and particles containing this additional DNA. Fractions containing RcGTA or phage DNA

were probed for the presence of RcGTA major capsid protein by western blot. None of the larger DNA band-containing fractions contained detectable RcGTA capsid whereas the fractions with RcGTA DNA did (data not shown).

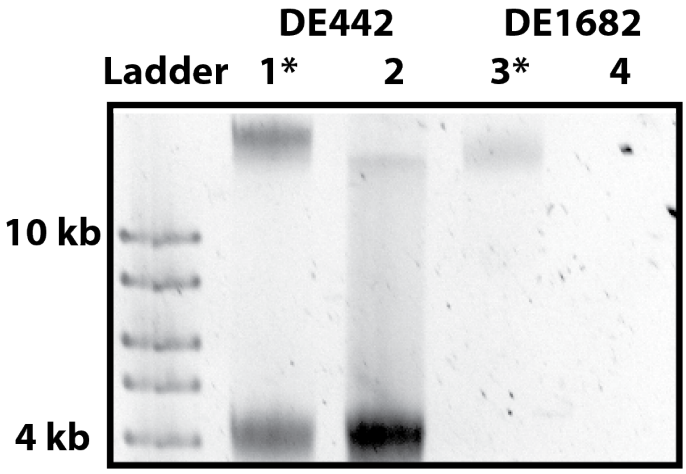

Figure 2-1: Detection of phage DNA co-purifying with RcGTA-packaged DNA. Visualization of DNA Extractions from PEG precipitates, with proteinase K treatments denoted by (*). DNA extractions from particles produced by DE442 (Lanes 1 & 2) result in a visible ~4-kb RcGTA DNA band accompanied by a larger, >10-kb band whose intensity increases with a proteinase K treatment (Lane 1). DNA extractions from particles produced by DE1682 (Lanes 3 & 4), a mutant deficient in RcGTA production, result in no RcGTA DNA and a large >10-kb band that is only detectable with a proteinase K treatment (Lane 3).

Table 2-2: Strains tested for presence of a detectable phage band.

Strains	Description	Source	GTA Activity	Phage Band Detected (Extraction 1)	Phage Band Detected (Extraction 2)
SB1003	rif resistant pseudo-wt	(Yen & Marrs, 1976; Strnad *et al.*, 2010)	++	+	+
DE442	GTA overproducer	?	++++	+	+
B10	Wild type cured of phage	(Weaver *et al.*, 1975)	++	+	+
A1	GTA capsid mutant	Gift from J.T. Beatty	0	-	+
DE1682	GTA *orfg1* mutant	Lang lab, unpublished	0	-	+
SBMF1	Polar GTA *orfg2* mutant	Lang lab, unpublished	0	-	+
YK8	*orfg2* mutant	(Lang & Beatty, 2000)	0	-	NT
ALS1	Disruption of *gtaI*	(Schaefer *et al.*, 2002)	+	-	+
ΔcckA	Deletion of *cckA*	(Mercer *et al.*, 2012)	+	-	+
SBRM1	Disruption of *ctrA*	(Mercer *et al.*, 2010)	0	-	+
SB555	Disruption of *rcc00555*	(Hynes *et al.*, 2012)	+	+	+
SB171	Disruption of *rcc00171*	Chapter 3	0	-	+
SB1685	Disruption of *rcc01685*	(Hynes *et al.*, 2012)	++	+	NT
SB1699	Disruption of *rcc01699*	Chapter 3	++	+	NT
SBchpT	Disruption of *chpT*	(Mercer *et al.*, 2012)	+	-	+

NT = Not tested

? = Strain of uncertain provenance. (Fogg *et al.,* 2011) cites (Yen *et al.*, 1979)

2.3.2 Phage discovery

Preliminary sequencing indicated that the larger DNA band contained sequences that mapped to an ~80-kb transposable phage cluster in the published genome (Strnad *et al.*, 2010) (see Figure 1-1). However, many of the detected sequences did not appear in the published *R. capsulatus* genome sequence. Complete sequencing by 454 pyrosequencing revealed that the phage band was in fact composed of two separate phages: one, a transposable prophage present in the *R. capsulatus* chromosome sequence, and the second, a new phage absent from the existing *R. capsulatus* genome sequence. The first we named RcapMu, and investigated in detail in collaboration with J.T. Beatty (Fogg *et al.*, 2011; See Appendix 2). The second phage, with three times the number (7,126) of mapping reads, RcapNL, is discussed here.

2.3.3 Genomics & Proteomics of RcapNL

The complete genome of RcapNL includes 64 ORFs, the majority of which (41/64) share no recognizable homology or domains with proteins of annotated function and were therefore annotated as "hypothetical proteins". By using my genome annotation to probe existing *R. capsulatus* SB1003 proteomic databases (Mercer *et al.*, 2010), I was able to detect peptides for 35 of the 64 ORFs, including 21 of the "hypothetical proteins". This allowed me to revise the annotation for these to "proteins of unknown function".

Most of the 64 ORFs shared no recognizable homology to known phage proteins; those few that did were similar to sequences from widely differing phages (e.g. transposable and lambda-like) from several bacterial host classes (e.g. α- and γ-

proteobacteria), and often with weak homology (E-value >e-15) (Table 2-3). There is no evidence of any previously described closely related phage. There is, however, evidence of related prophages in the genomes of *Rhodobacter sphaeroides* ATCC 17025 (Accession NC_009428), *Paracoccus denitrificans* PD1222 (Accession NC_008688) and even elsewhere in *R. capsulatus* SB1003 (Accession NC_014034) (Figure 2-2). RcapNL (see later, section 2.3.6) and these three prophages are integrated at the site of a tRNA, and organized similarly relative to that tRNA.

Within *R. capsulatus*, RcapNL has regions of homology to the aforementioned prophage cluster (Figure2-2), as well as possessing two genes that are similar to a pair of genes in the 14-kb phage cluster (*rcc02331*, *rcc02327*), another pair of genes homologous to two genes in the RcGTA structural cluster (*rcc01684*, *rcc01691*), and one gene homologous to a gene of RcapMu (*rcc00995*). The tail-fibre gene *RcapNL_29* was homologous to *rcc01079*.

Table 2-3: RcapNL predicted ORFs and their best phage BLAST hits, where available

Name	AA length	Predicted product	Best phage hit (E-value)	Id[a], Sim[b], Cov[c]
RcapNL1	252	Protein of unknown function		
RcapNL2	121	HNH endonuclease	*Xanthomonas* phage Xop411 (2e-11)	48% 61% 61%
RcapNL3	162	Protein of unknown function		
RcapNL4	565	large terminase	*Xanthomonas* phage phiL7 (2e-92)	37% 53% 96%
RcapNL5	426	HK97 family portal protein	Enterobacteria phage SfI (2e-54)	35% 53% 84%
RcapNL6	315	peptidase S49	*Vibrio parahaemolyticus* phage VP16T (3e-14)	33% 51% 54%
RcapNL7	434	HK97 family major capsid protein	*Burkholderia* phage phi644-2 (1e-76)	42% 59% 99%
RcapNL8	153	hypothetical protein		
RcapNL9	73	hypothetical protein		
RcapNL10	192	Protein of unknown function		
RcapNL11	109	phage head-tail adapter protein	*Rhizobium* phage 16-3 (7e-06)	34% 47% 98%
RcapNL12	103	Protein of unknown function		
RcapNL13	119	phage HK97 gp10-like protein		
RcapNL14	136	GTA orfg8-like protein		
RcapNL15	143	GTA orfg9-like major tail protein		
RcapNL16	190	Protein of unknown function		
RcapNL17	86	Protein of unknown function		
RcapNL18	519	phage tail-related protein		
RcapNL19	228	phage tape measure protein	*Xanthomonas oryzae* phage OP1(8e-08)	46% 59% 42%
RcapNL20	216	hypothetical protein		
RcapNL21	195	hypothetical protein		
RcapNL22	138	hypothetical protein		
RcapNL23	142	hypothetical protein		

RcapNL24	587	phage tail component protein	*Rhizobium* phage 16-3 (1e-17)	25% 39% 96%
RcapNL25	651	hypothetical protein		
RcapNL26	106	hypothetical protein		
RcapNL27	207	peptidoglycan-binding protein	*Serratia* phage phiMAM1 (1e-22)	41% 56% 66%
RcapNL28	67	hypothetical protein		
RcapNL29	85	tail fibre protein		
RcapNL30	144	hypothetical protein		
RcapNL31	102	Protein of unknown function		
RcapNL32	73	XRE family transcriptional regulator		
RcapNL33	75	Protein of unknown function		
RcapNL34	366	phage integrase/recombinase		
RcapNL35	80	hypothetical protein		
RcapNL36	360	phage Gp37Gp68 family protein	*Mycobacterium* phage Che9c (2e-64)	44% 52% 98%
RcapNL37	169	Protein of unknown function		
RcapNL38	77	PRK12775-containing protein		
RcapNL39	627	C-5 cytosine-specific DNA methylase	*Pectobacterium* phage ZF40 (4e-109)	38% 50% 93%
RcapNL40	54	hypothetical protein		
RcapNL41	68	hypothetical protein		
RcapNL42	322	DUF2303-containing protein	*Burkholderia* phage BcepMigl (4e-11)	27% 39% 86%
RcapNL43	118	Protein of unknown function		
RcapNL44	78	Protein of unknown function		
RcapNL45	86	hypothetical protein		
RcapNL46	75	hypothetical protein		
RcapNL47	89	Protein of unknown function		
RcapNL48	102	Protein of unknown function		
RcapNL49	76	Protein of unknown function		
RcapNL50	212	phage repressor	*Pseudomonas* phage D3112(1e-15)	36% 58% 63%

RcapNL51	95	hypothetical protein		
RcapNL52	95	Protein of unknown function		
RcapNL53	143	Protein of unknown function		
RcapNL54	90	Protein of unknown function		
RcapNL55	438	Protein of unknown function	*R. capsulatus* phage RcapMu (4e-13)	46% 61% 21%
RcapNL56	114	hypothetical protein	*Pseudomonas* phage JBD67 (4e-10)	45% 56% 75%
RcapNL57	116	Protein of unknown function		
RcapNL58	63	Protein of unknown function		
RcapNL59	97	hypothetical protein		
RcapNL60	160	Protein of unknown function		
RcapNL61	388	phage P4 alpha zn-binding domain		
RcapNL62	77	hypothetical protein		
RcapNL63	101	hypothetical protein		
RcapNL64	622	P4 family phage/plasmid primase	*Lactobacillus* phage A2 (3e-57)	33% 47% 75%

Genes highlighted in grey produce proteins detected by proteomic analysis.
a = percentage identity
b = percentage similarity
c = percent coverage

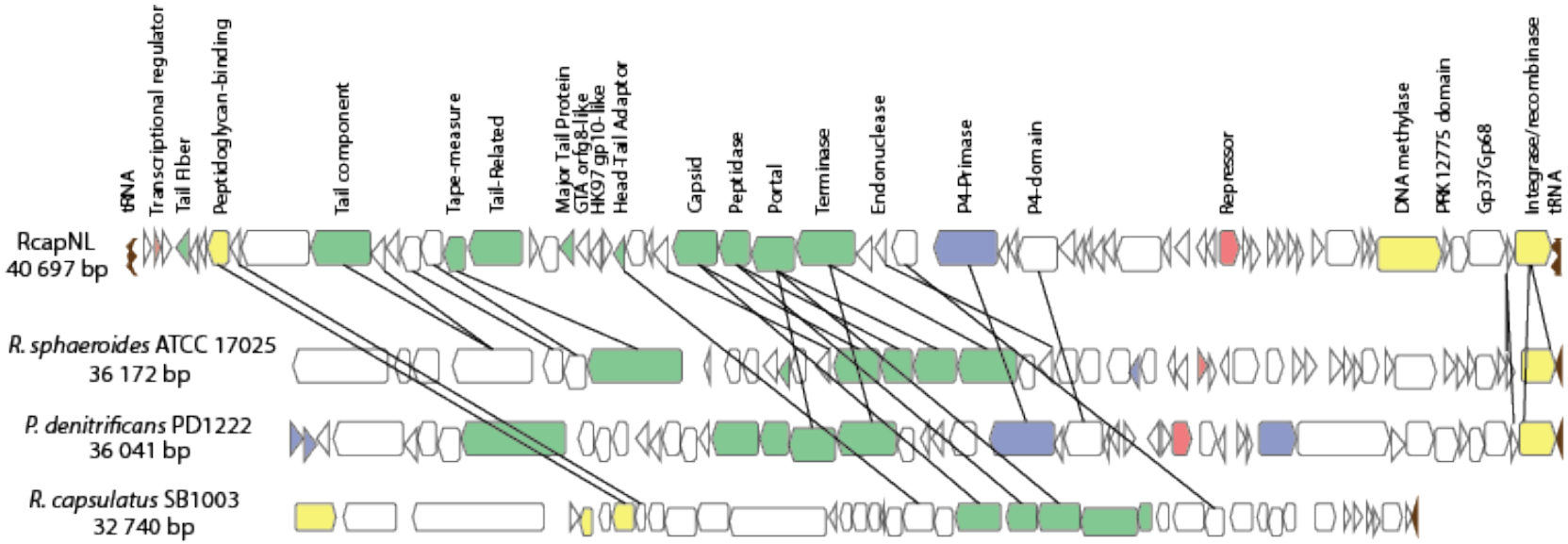

Figure 2-2: Comparison of the genome of RcapNL to three related prophage clusters. Vertical offset in neighbouring genes indicates different reading frames, and the arrows indicate the direction of transcription. The cluster in *R. sphaeroides* is from position 1 806 135 to 1 769 964 (Accession NC_009428). The cluster in *P. denitrificans* is from position 984 618 to 948 578 on chromosome 1 (Accession NC_008688). The cluster in *R. capsulatus* is from position 2 165 178 to 2 132 438 (Accession NC_014034). Black lines indicate recognizable homology at the amino acid level ($e < 10^{-15}$). Colouring indicates function of the encoded protein, according to existing annotation: green indicates proteins involved in virion assembly, blue indicates replication-related functions, red indicates regulatory functions, and yellow indicates host interaction. Brown highlights the tRNA sequences adjacent to the prophages.

2.3.4 RcapNL *cos* sites

To investigate the DNA ends of the RcapNL genome, I attempted to amplify outwards from the ends of purified RcapNL DNA using PCR. After self-ligation I obtained PCR products, confirming that the self-ligated DNA was circular. Sanger sequencing of the products revealed 26 bp of sequence not present in the 454 reads. As the preparatory protocol for 454 sequencing would remove any 3' overhangs, I attributed this new sequence to 3' cohesive overhangs at the ends of the RcapNL genome (Figure2-4).

The predicted terminase sequence shows the highest similarity to terminases known to produce 3' overhangs. The highest similarity to a characterized phage terminase was to that of *Xanthomonas oryzae* Phage 10, which produces 9-bp 3' cohesive ends (Boyd & Brüssow, 2002).

2.3.5 RcapNL prophage

The RcapNL genome sequence is not present in the existing *R. capsulatus* SB1003 genome sequence. However, the properties of its genome, and the fact that it has never been detected as a plasmid or extra-chromosomal replicon, suggested it must be integrated into the *R. capsulatus* genome. To determine if this was a recent event (acquired, perhaps, in this lab) I probed the genomes of a variety of *R. capsulatus* strains for the presence of the capsid and terminase genes by PCR ("Cap" and "Ter" primers, Table 2-1). Several other strains clearly possess a similar phage (Table 2-4).

I also investigated all of the *R. capsulatus* CRISPR arrays for evidence of spacers targeting RcapNL. In the largest, 40-spacer CRISPR, 7 spacers scattered throughout the array were >90% identical to sequences of RcapNL (Figure 2-3). None of the other spacers had significant homology ($e<10^{-5}$) to any known sequences. The targeted regions in the phage sequence, known as proto-spacers, were preceded by a DNA motif (Figure 2-3) consistent with the protospacer-associated-motifs (PAMs) found in active CRISPR systems (Mojica *et al.*, 2009; Makarova *et al.*, 2011). The predicted crisper-associated (CAS) proteins at this locus are consistent with a type I-C system (Makarova *et al.*, 2011), also known as "dvulg".

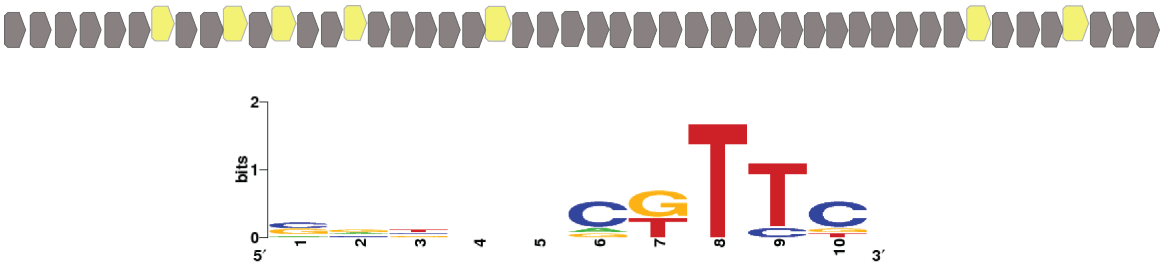

Figure 2-3: The large *R. capsulatus* SB1003 CRISPR array and its properties with respect to RcapNL. (Top) A depiction of the 40-spacer CRISPR type I-C array in SB1003. Grey regions represent the repeats. Yellow regions indicate spacers with homology to any known sequence. Spacers with no homology to any known sequence are present in between all other repeats, but not depicted. All the yellow regions appear to target sequences in RcapNL, with mismatches ranging from none to as many as 3-bp out of the 34-37-bp region of homology. (Bottom) A weblogo depiction of the conserved motifs on RcapNL adjacent to phage regions targeted by spacers (proto-spacers). The conserved "GTTC" motif appears to be a proto-spacer associated motif (PAM), consistent with active CRISPR systems.

2.3.6 RcapNL integration

To find the integration site of RcapNL, I attempted inverse PCR using a variety of primer pairs and restriction enzymes. None yielded any sequence information outside that of the phage itself. In an unrelated attempt to sequence phage-packaged DNA, I obtained

RcGTA-packaged chromosomal DNA that included seemingly chimeric fragments mapping to both RcapNL and SB1003. The host-component of these phage-host sequences mapped to an Arginine tRNA at position 712 145 in the chromosome (Figure 2-4), which suggested this was the phage integration site. This was confirmed by PCR amplification of the phage-host junction in genomic DNA isolates. There is a region of 15 bases in the phage with homology to the integration site in the host. I have named the phage region, located between a hypothetical protein *RcapNL33* and the integrase *RcapNL34*, *attP*, and the host region forming part of a tRNA, *attB* (Figure 2-4).

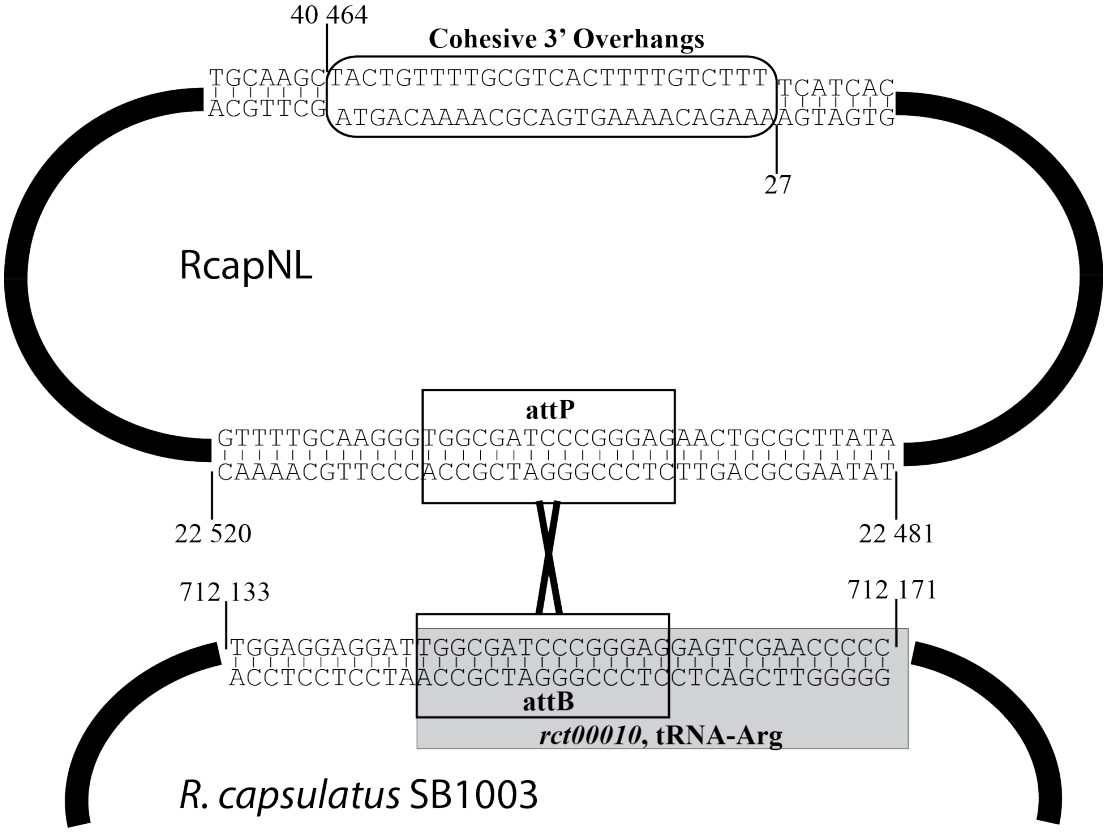

Figure 2-4: Proposed model for integration of RcapNL into the *R. capsulatus* genome. The long 26-bp *cos* sites (top) pair, circularizing the phage. The phage *attP* region, homologous to a 15 bp region at the end of a host tRNA-Arg gene (*rct00010*), recombines with this attachment (*attB*) site, presumably via the phage-encoded integrase. The resulting prophage begins and ends with intact *attB* sites, therefore leaving the tRNA intact.

2.3.7 Host range

The absence of amplified RcapNL terminase or capsid sequences in many of the tested stains (Table 2-4) suggested they might be good candidate hosts for RcapNL replication, as they would lack any superinfection immunity conferred by the prophage. I used SBKMu culture filtrates and PEG precipitates as a source of RcapNL, and incubated these with all strains listed in Table 2-4. I observed no detectable lysis or plaques (Table 2-4).

Table 2-4: Screening *R. capsulatus* natural isolates and phage-knockouts for the presence of RcapNL by PCR and the ability to serve as a host for RcapNL, purified from SBKMu.

Strain	Source	Capsid	Terminase	Plaque
SB1003	(Yen & Marrs, 1976)	+	+	-
DE442	(Yen *et al.*, 1979)	+	+	-
YW1	(Weaver *et al.*, 1975)	-	-	-
YW2	(Weaver *et al.*, 1975)	+	+	-
B6	(Weaver *et al.*, 1975)	-	-	-
B10	(Weaver *et al.*, 1975)	+	+	-
SP18	(Weaver *et al.*, 1975)	-	-	-
SP36	(Weaver *et al.*, 1975)	-	+*	-
H9	(Weaver *et al.*, 1975)	-	-	-
P12F1	(Weaver *et al.*, 1975)	-	+*	-
SBKNL	This Study	+	-	-
SBΩNL	This Study	+	-	-
SBKMuΩNL	This Study	+	-	-

*Detectable band of unexpected size

2.3.8 Confirming phage-knockout strains

In order to further investigate the link between RcGTA production and the phage, I assayed phage knockout strains SBKMu, SBKNL, SBΩNL and SBKMuΩNL (See

2.2.11) for gene transfer activity. To confirm the knockouts, DNA extracted from particles harvested from each strain was subjected to PFGE (Figure 2-5). The strain lacking RcapMu (SBKMu) and producing only RcapNL was indistinguishable from strains lacking RcapNL (SBKNL, SBΩNL) and producing only RcapMu. Restriction digests were used to confirm the appropriate phage was absent (data not shown). In all tested strains, including the strain lacking both RcapMu and RcapNL, a larger >48 kb DNA smear was detectable, but only when the DNA extraction included treatment by proteinase K.

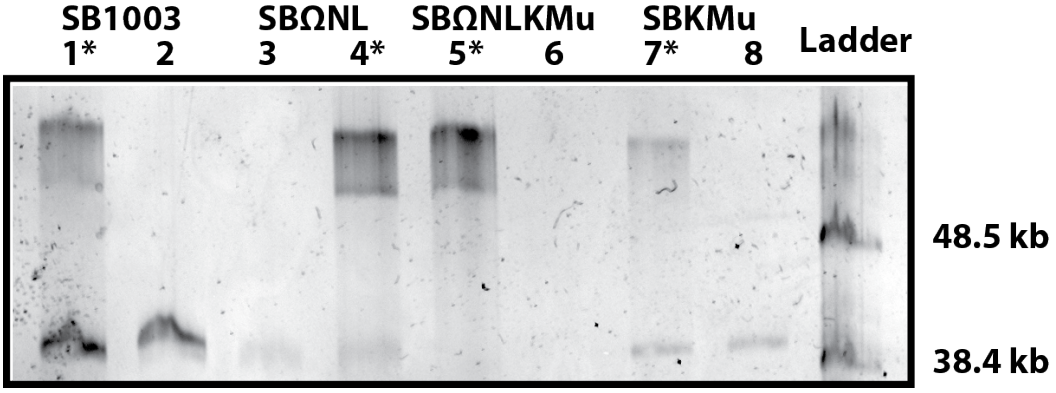

Figure 2-5: Characterization of phage DNA band sizes by pulsed-field gel electrophoresis. The sizes of the two distinct bands of the ladder are labeled (right). The addition of proteinase K (lanes marked *) resulted in the appearance of a larger >48-kb band in all tested strains, including a mutant producing neither RcapMu nor RcapNL (Lane 5). The ~40-kb bands correspond to RcapMu and RcapNL (lanes 1 and 2), RcapMu alone (Lanes 3 and 4), or RcapNL alone (Lanes 7 and 8).

2.3.9 RcapH

This DNA band that is only detectable after treatment with proteinase K, which I hypothesize to be a third phage, RcapH, was sized at 80.85 kb (Figure 2-6). Its production appears to be entirely RcGTA independent, as it can be purified in comparable amounts from RcGTA-deficient strains (e.g. Figure 2-1). Digestion of this 80.85-kb DNA with a

variety of restriction enzymes (BamHI, SalI, SacI, EcoRI, MseI, PacI) consistently failed. An attempt to sequence DNA extracted from the phage fraction of the SBΩNL/KMu strain on the high-throughput Ion Torrent instrument yielded very little data, and all sequences obtained mapped to the entirety of the *R. capsulatus* genome. This was interpreted to result from contamination by RcGTA-packaged genomic DNA.

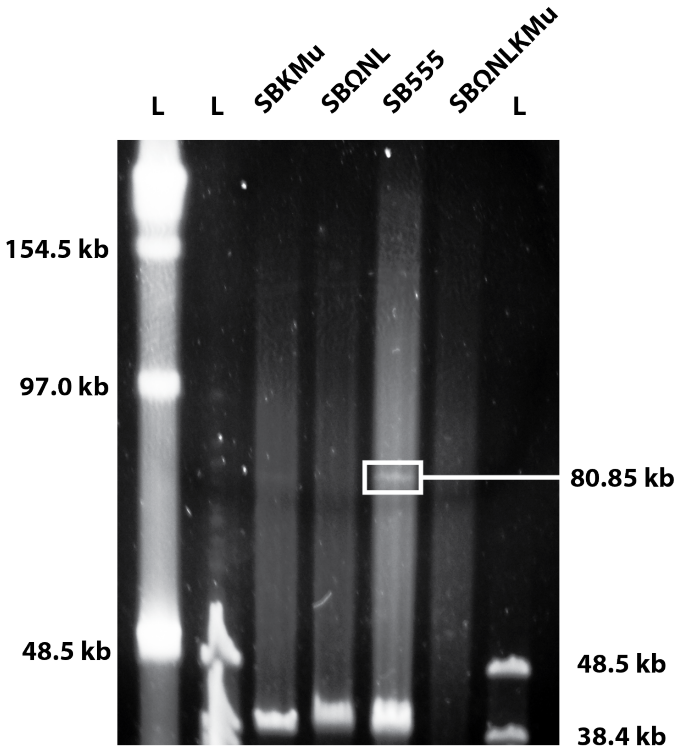

Figure 2-6: Sizing of the RcapH phage DNA band by pulsed-field gel electrophoresis. Distinct ladder bands of known size are labeled to the left and right of the image. All experimental samples were treated with proteinase K. The run conditions were optimized to examine the larger band (>48-kb) seen in Figure 2-2, which appears to be a distinct band at 80.85 kb that is only distinguishable in high-yield preparations (SBKMu & SB555).

2.3.10 Linking phage production and RcGTA activity

In order to further investigate the link between RcGTA production and the phage, I assayed phage-deficient knockout strains SBKMu, SBKNL, SBΩNL and SBKMuΩNL for gene transfer activity. All strains were indistinguishable from the parental strain for gene transfer activity.

2.4 Discussion

My efforts to purify phage from *R. capsulatus* SB1003 identified at least two (RcapNL, RcapMu), and possibly a third (RcapH), phages co-purifying with RcGTA particles.

The 40.5-kb genome of RcapNL (GI: 461474957) was absent from the host strain SB1003's published genome sequence. As the phage had not been detected as an independent replicon, I suspected it was present as an integrated prophage. I probed other *R. capsulatus* strains for evidence of this phage, and was able to amplify similar capsid and terminase genes in several other *R. capsulatus* strains (Table 2-4). In addition, I examined the CRISPR arrays of *R. capsulatus* SB1003 for evidence of previous association with RcapNL. CRISPR spacers are acquired by exposure to foreign DNA and enable species carrying functional CRISPR arrays to acquire sequence-specific immunity (Barrangou *et al.*, 2007). As the spacers are acquired by addition at the 5' end, the arrays serve as a "fossil-record" for exposure to foreign DNA; each spacer represents an older acquisition event than that upstream of it (Vale & Little, 2010). Unfortunately, as acquisition (at the 5' end of the array) and loss (from the 3' end) are poorly understood

events with irregular timing, the spacers do not serve as a good "clock" to estimate the timing of phage-host interactions.

The largest CRISPR array in the *R. capsulatus* SB1003 genome, possessing 40 spacers, is accompanied by CRISPR-associated (Cas) genes identifying the array as a type I-C or Devulg system (Makarova *et al.*, 2011). Of its 40 spacers, only 7 have significant ($e < e\text{-}5$) similarity to any known sequence; all 7 of these are similar to RcapNL (Figure2-3). These spacers are distributed throughout the array, suggesting multiple interactions separated by other phage-host interactions and, presumably, long stretches of time. This leads us to believe that the association between RcapNL and *R. capsulatus* SB1003 is a long one and that the presence of RcapNL in the lab strains is unlikely to be due to a recent event. The 33 remaining spacers bear no similarity to any known sequence.

RcapNL showed little genetic similarity to known phages. While the phage genome included identifiable structural components (Table 2-3), the majority of the ORFs (41/64) shared no homology to known proteins and I was unable to assign any function to them. RcapNL's genome is recognizably homologous to prophages in other *Rhodobacterales* (Figure2-2). The observed homologies and similarities in genome organization amongst these elements suggest that RcapNL is related to phages that have integrated in the genomes of several *Rhodobacterales*.

Although I was unable to find a host that RcapNL could infect, I was able to determine some important features of RcapNL replication. Re-sequencing of the SB1003 genome revealed chimeric phage-host reads, which were confirmed by PCR to be the integration site of the phage. I found that RcapNL had long, 26 bp 3' cohesive ends. The

long, 26 bp overhangs of RcapNL are, to my knowledge, unprecedented. I propose a model for integration whereby the 3' *cos* sites pair, circularizing the phage genome, and the *attP* and *attB* sites allow for phage-mediated recombination resulting in an integrated prophage (Figure 2-4). Interestingly, because the *attB* site is the last 15 bp of the tRNA, integration maintains an intact copy of the arginine tRNA.

RcapMu, a transposable phage present as part of the *R. capsulatus* SB1003 published genome, has a 39.3-kb genome (Genbank Identifier (GI): 356870838) and runs as a distinct band at ~40.5 kb that is indistinguishable from RcapNL (Figure 2-5). It is from this observation, along with restriction digests, that I was able to determine that it packages >1.2 kb of host DNA at the 3' end of it's genome, as Mu-like phages are known to do (Mizuuchi & Craigie, 1986). Its life cycle, integration sites, transposition activity and structural proteins were characterized in partnership with Paul Fogg (see Appendix 2), and despite being one of very few sequenced α-proteobacterial phages, its homology to known transposable phages allowed a relatively robust annotation and attribution of function for the predicted ORFs.

RcapMu and RcapNL appear to be directly linked to RcGTA production. In structural or regulatory mutants deficient for RcGTA activity, the co-purifying phage band was absent (Table 2-2, Figure 2-1). Given that strains carrying gene knockouts which do not affect RcGTA production (e.g. SB1685) did not affect presence of a detectable phage band (Table 2-2), this link does not appear to be an artifact of the mutant generation process. While modifications to the purification and extraction techniques resulted in increased phage yields and detectable DNA, even in RcGTA-deficient

mutants, the majority of this increased yield was attributed to the detection of a third phage, RcapH (Figure 2-5).

RcapH's genome, which appears as a smear of DNA centered around a band of 80.85 kb (Figure 2-6), is released in detectable amounts only after proteinase K treatments of PEG precipitates from *R. capsulatus* cultures (Figure 2-5). This could be explained by the presence of proteins covalently bound to the DNA, trapping the DNA in the phenol-chloroform interface in the absence of a proteinase K treatment. The DNA band is present even in mutants in which RcapMu or RcapNL have been knocked-out (Figure 2-5). The phage DNA appears intractable to Ion Torrent sequencing and to digestion by a variety of restriction enzymes, which suggests modified or otherwise protected DNA. As there is no prophage cluster of that size in the sequenced *R. capsulatus* SB1003 genome, this phage's genome remains an enigma.

The fact that knockouts of RcapNL, RcapMu or both failed to affect gene transfer activity establishes that any transduction activity by RcapMu or RcapNL is negligible relative to the overall levels observed in RcGTA producing cultures. It also establishes that the production and release of functional RcGTA particles is not dependent on the production and release of functional RcapNL or RcapMu. There is no evidence from my work, or any previous studies, that RcGTA is dependent on individual genes encoded by either of these two phages, but we cannot rule out this possibility..

As the association between RcGTA activity and phage production is apparent in both regulatory mutants and structural mutants (e.g. capsid), it is likely that this is a dependence on more than a shared regulatory cue initiating production and depends, in some way, on a the production of RcGTA particles. I investigated the possibility that the

phages were dependent on the cell lysis mechanism used by RcGTA for their release from the cell. Rcc0555 is involved in RcGTA release through cell lysis (Hynes *et al.*, 2012) but its disruption still resulted in a detectable 40.5 kb band (Figure 2-6), so RcapMu and RcapNL are not dependent on RcGTA release for their own escape from the cell. Attempts to induce these phages with carbadox or UV irradiation failed, but if RcGTA production results in damage to the genome by the action of the packaging mechanisms, this RcGTA-production-dependent damage could induce the prophages to begin replicating to "escape" a dying host cell. This would explain the apparent dependence of phage production on structural components of RcGTA production.

There does not appear to be any evidence of a link between RcGTA's structural gene cluster and the genomes of RcapMu and RcapNL that would explain the dependence of the phages upon RcGTA. Several other phages seem to share more in common with RcGTA's structural cluster; a transposable phage with an associated RcGTA-like structural cluster was reported (Paul, 2008), while another of the few characterized α-proteobacterial phages contains several GTA-like genes as well as homologues to *ctrA* and *rcc00171* (Huang *et al.*, 2011), both proteins known to be involved in RcGTA production (Chapter 3; Mercer *et al.*, 2010; Lang & Beatty, 2000).

3 The Non-Contiguous Genome of RcGTA, the *Rhodobacter capsulatus* Gene Transfer Agent

3.1 Introduction

Rhodobacter capsulatus is a purple non-sulfur bacterium that produces the gene transfer agent RcGTA. Gene transfer agents (GTAs) are host-encoded phage-like particles capable of transferring host DNA from the producing cell to recipient cells (Lang *et al.*, 2012). RcGTA packages essentially random (Hynes *et al.*, 2012) ~4-kb pieces of DNA (Solioz & Marrs, 1977; Yen *et al.*, 1979). Genes responsible for production of RcGTA were first identified by transposon mutagenesis and screening for mutants deficient in gene transfer activity (Lang & Beatty, 2000). This identified an ~15-kb region originally annotated as having 19 open reading frames (ORFs) (Lang & Beatty, 2001) called the RcGTA structural gene cluster. It is organized in a manner similar to the structural gene clusters in genomes of tailed phages (Casjens *et al.*, 1992). This cluster of genes is proposed to have a long evolutionary history within the class α-proteobacteria (Lang & Beatty, 2007), and is especially well conserved in the order *Rhodobacterales* (Lang *et al.*, 2002; Biers *et al.*, 2008; Lang *et al.*, 2012; Paul, 2008). This suggests that GTA production might be widespread in these bacteria and provide benefits to GTA-producing populations.

Encoded by a 15-kb cluster of genes but capable of packaging only 4 kb, RcGTA is completely dependent on chromosomal replication for its maintenance. While this is the case for most "defective" phages, RcGTA distinguishes itself from most defective

phages in several important aspects. First, it still encodes particles capable of packaging DNA that can be transferred to other cells. Secondly, the production of RcGTA is regulated by the host (Lang & Beatty, 2000; Mercer *et al.*, 2012; Leung *et al.*, 2012; Schaefer *et al.*, 2002) with production upregulated in the stationary phase of culture growth (Solioz *et al.*, 1975). The release of RcGTA particles has recently been shown to depend on lysis genes encoded elsewhere on the chromosome (Hynes *et al.*, 2012). This dependence on the host for replication, regulation and release demonstrates a remarkable integration of this virus-like particle with *R. capsulatus* biology.

A proteomic characterization of the RcGTA particles confirmed the presence of nine of the RcGTA structural cluster-encoded proteins, but also identified four additional proteins associated with the particles (Chen *et al.*, 2008). This suggests that additional loci might contribute components of the RcGTA particle. An unrelated GTA, VSH-1 produced by *Brachyspira hyodysenteriae*, is encoded by at least two separate loci in the host chromosome (Stanton *et al.*, 2009).

In this study, I have attempted to better define the genetic contributions to production of the RcGTA particle. I compared the transcriptional profiles of a variety of mutant strains and growth conditions known to affect RcGTA production to create a list of genes consistently co-regulated with the RcGTA structural gene cluster. I then tested the function of these genes with respect to RcGTA production, through a combination of genetic and biochemical approaches.

3.2 Experimental procedures

3.2.1 Strains and plasmids

The strains and plasmids used in this study are listed in Table 3-1. Insertional disruptions of *R. capsulatus* genes were performed by amplification of the gene of interest by PCR using the primers indicated in Table 3-2, ligation of the resulting product into the pGEM-T easy vector system (Promega, Madison WI), and subsequent insertion of the 1368 bp SmaI fragment of the KIXX cartridge (Barany, 1985) encoding kanamycin resistance at the restriction site listed in Table 3-2. The pGEM construct with the disrupted gene was then transferred by conjugation from *Escherichia coli* C600 (pDT51) (Taylor *et al.*, 1983) into *R. capsulatus* DE442 and the marked gene disruption transferred into *R. capsulatus* SB1003 by RcGTA transfer (Scolnik & Haselkorn, 1984). The disrupted gene replacements were confirmed by PCR with the original amplification primers and visualization of the size differences resulting from KIXX insertion.

Table 3-1: Strains and experimental plasmids used in this study

	Description	Source
Strains		
R. capsulatus		
SB1003	Genome-sequenced strain	(Yen & Marrs, 1976); (Strnad *et al.*, 2010)
DE442	GTA overproducer, *crtD*$^-$ (provenance uncertain)	(Fogg *et al.*, 2011) cites (Yen *et al.*, 1979)
ALS1	*gtaI* knockout in SB1003	(Schaefer *et al*., 2002)
SBRM1	SB1003 with disruption in *ctrA*	(Mercer *et al*., 2010)
SB0171	SB1003 with disruption in *rcc00171*	This study
SB1079	SB1003 with disruption in *rcc01079*	This study
SB1080	SB1003 with disruption in *rcc01080*	This study
SB1699	SB1003 with disruption in *rcc01699*	This study
SB2623	SB1003 with disruption in *rcc02623*	This study
DW5	SB1003 *ΔpuhA*	(Wong *et al*., 1996)
Escherichia coli		
C600 (pDT51)	Plasmid mobilizing strain	(Taylor et al., 1983)
S17-1	Plasmid mobilizing strain	(Simon et al., 1983)
BL21(DE3)	competent cells for protein expression with T7 promoter	New England Biolabs (Pickering, Canada)
Plasmids		
pRK767	complementation vector	(Gill & Warren, 1988)
pRK2TF	pRK767 with *rcc01079* & *rcc01080* and 141 bp of 5' sequence	This study
pRK0171	pRK767 with *rcc00171* and 118 bp of 5' sequence	This study
pET28a(+)	protein expression vector	Merck Millipore (Billerica, MA)
pET28-171	*rcc00171* in pET28, C-terminal His-tag	This study

Table 3-2: Primers and associated restriction sites used in mutant construction, complementation and protein expression

ORF disrupted	Primer sequences	Disruption site
171	GCCCTTCGACACCTATCTGA GTATCGTCAGCACGAGAGCA	NruI
1079 & 1080	TTCATCTTGCAGCCCTTCTT CATAGGTCAGAACCGCCTGT	BlpI (*1079)* BsmI *(1080)*
1699	AGTTCGGAAAATTGGGAGGT AATGCAGCATCGAGACATTG	RsrII
2623	TGCCGGATTTCTTCTTTGTC TTCACGGCTAGGTCTGGTCT	StuI
Complementation	**Primer Sequences**	**Insertion into pRK767**
171	TAGGTACCCGCCCGGCGGCGTCT ATGGTACCACGCGCCCGCAGCCT	KpnI
1079/1080	TAGGTACCCGCGCCGCCTCTGC ATGGTACCCCGCGGCCAGCCG	KpnI
Fusion construct	**Primer Sequences**	**Insertion into pET28**
171	CGCGGAGATCACCCATGGCCGACCA TGGATCCACCCAGGGGCCGATGGC	NcoI/BamHI

Underlined sequences indicate the restriction site used

Constructs to complement mutants *in trans* were created by amplification of the wild type gene along with sufficient upstream region to include the probable promoter. The primers included 5' KpnI sites to facilitate cloning into the plasmid pRK767, with resulting plasmids transferred to *R. capsulatus* by conjugation from *E. coli* S17-1 (Simon *et al.*, 1983)

The pET28-171 expression plasmid was constructed by amplification of *rcc00171* using the primers indicated in Table 3-2, which included 5' restriction enzyme sites to directionally clone the insert into pET28a(+). The resulting construct was sequenced by The Centre for Applied Genomics (TCAG) (Toronto, Canada) using their own T7 and T7term primers to confirm the in-frame fusion to the C-terminal His-tag and conservation of the original *rcc00171* sequence.

R. capsulatus cells were grown under aerobic conditions at 30°C in RCV medium (Beatty & Gest, 1981) for general culturing or 35°C under anaerobic photoheterotrophic conditions in YPS medium (Wall *et al.*, 1975b) for high RcGTA production (transductions, bioassays, microarrays).

3.2.2 Microarray analyses

The data from the *R. capsulatus* SB1003 and SBRM1 (*ctrA*) (Mercer *et al.*, 2010), DE442 (Hynes *et al.*, 2012), and ALS1 (Brimacombe *et al.*, 2013) microarrays have been previously published and are available from the NCBI Gene Expression Omnibus (GEO) database under accession numbers GSE18149, GSE33176 and GSE41014.

In all cases, cells were grown photoheterotrophically and were harvested 4 h after reaching stationary phase, as determined by monitoring culture turbidity. For SB1003, an additional sample was taken in mid-logarithmic growth. RNA was extracted using the RNeasy Kit (Qiagen, Mississauga, Canada).

The *R. capsulatus* microarrays are Affymetrix custom whole-genome expression arrays (Affymetrix, Santa Clara, CA) described in Mercer *et al*. (Mercer *et al.*, 2010). cDNA synthesis, labeling and array hybridization were performed at the Michael Smith Genome Science Centre (Vancouver, Canada) as described in the Affymetrix Expression Analysis Technical Manual for Prokaryotic Samples. Raw data from the arrays were robust multi-array (RMA) normalized (Irizarry *et al.*, 2003), then normalized to the 50th percentile using GeneSpring 7.2 (Agilent Technologies, Santa Clara, CA). The signal intensities were used to identify genes with >2 fold change in intensity between the experimental and the reference dataset; the SB1003 early stationary (ES) sample.

3.2.3 RcGTA activity bioassays

Bioassays were performed as previously described (Hynes *et al.*, 2012). Normalized culture amounts were used to inoculate YPS anaerobic photoheterotrophic cultures that were grown for 48 h, with the RcGTA activity assayed by *puhA* transfer to DW5 cells. For protein complementation/inhibition assays, 3 to 30 μg of protein was added prior to the addition of RcGTA-containing filtrate. Controls received an equal volume of dialysate (see 3.2.7 "Protein purification") with 10% v v^{-1} glycerol, or a BSA solution in dialysate. All data were analyzed as the gene transfer activity relative to SB1003. The transfer rates were compared by one-way ANOVA and Tukey HSD tests.

3.2.4 Detection of RcGTA major capsid protein by western blotting

Cells and culture filtrates (0.45 μm) from the cultures used in RcGTA activity bioassays were assayed for the presence of RcGTA capsid protein by western blotting. Cultures were centrifuged at 17000 *g*, the supernatant was removed and the cells resuspended in an equal volume of pH 8.0 TE buffer. 5 μl of cell re-suspension or 10 μl of culture filtrates were subjected to SDS-PAGE, with blotting and detection of the capsid protein performed as previously described (Mercer *et al.*, 2012). Images were captured on an ImageQuant LAS 4000 (GE Life Sciences, Baie D'Urfe, Canada) and band intensities quantified, where relevant, with ImageQuantTL version 8.1. The blot images were manipulated for brightness and contrast, with all regions of each image manipulated equivalently, using Adobe Photoshop C.S. 6.0 (Adobe Systems Inc, San Jose, CA).

3.2.5 Assays of RcGTA binding to cells

To monitor the binding efficiency of RcGTA particles to *R. capsulatus* cells, RcGTA particles were added to G buffer (Solioz & Marrs, 1977) or G buffer containing recipient cells as done for gene transfer bioassays. After a 1 h incubation the cells were pelleted by centrifugation at 17000 *g* for 1 min and the supernatants were collected and analyzed for the remaining amount of capsid protein by western blotting as described above.

3.2.6 DNA extraction from RcGTA particles

Cultures were grown for 48 h at 35°C under anaerobic photoheterotrophic conditions in YPS medium. The DNA extractions were performed as previously described (Hynes *et al.*, 2012). Briefly, the cultures were filtered and the filtrates were treated with nucleases and proteinase K, and the DNA purified by phenol:chloroform:isoamyl alcohol (25:24:1) extraction and ethanol precipitation. The samples were then subjected to agarose gel electrophoresis.

3.2.7 Protein purification

Overnight cultures of *E. coli* BL1(DE3) containing pET28-171 were used to inoculate 200 ml of LB broth containing 25 μg ml^{-1} kanamycin. The cultures were grown at 37°C for 1 h, then induced by addition of isopropyl-β-D-thiogalactopyranoside (IPTG) to a final concentration of 1 mM, and allowed to grow for another 4 h. Cell pellets of induced cultures were resuspended in 4 ml lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, 0.1% (v v^{-1}) Benzonase® nuclease (Qiagen), 1 mg ml^{-1} lysozyme

(w/v); pH 8) and incubated on ice for 30 min. The lysates were centrifuged at 14000 *g* for 30 min at 4°C and supernatants were mixed 4:1 (v v^{-1}) with Ni-NTA agarose (Qiagen) and incubated at 4°C with slow shaking for 1 h. The samples were loaded into polypropylene columns, washed twice with wash buffer (50 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole; pH 8) and the fusion proteins eluted in elution buffer (50 mM NaH_2PO_4, 300 mM NaCl, 250 mM imidazole; pH 8). The purified proteins were dialyzed into a coupling buffer (20 mM sodium phosphate, 500 mM NaCl; pH 7.5) and quantified relative to known BSA standards using an ND-1000 Nanodrop spectrophotometer. The success of the purification was evaluated by SDS-PAGE and Coomassie Blue staining with samples from the pre-induced culture, induced culture, lysate, wash and eluate. The dialyzed purified protein was split into aliquots and stored at 4°C or at -80°C with the addition of glycerol to 10% (v v^{-1}).

3.2.8 Electrophoretic mobility shift assay

Purified Rcc00171 or BSA in dialysate were added to a final concentration of 0, 20, 80 or 160 μg ml^{-1} to 125 ng of a 500 bp PCR product or DNA from within RcGTA particles in a Tris Buffer (6 mM Tris HCl, 6 mM $MgCl_2$, 50 mM NaCl, 1 mM DTT; pH 7.5) and incubated for 20 min at room temperature. The samples were then run on a 1% agarose gel then stained with GelRed (Invitrogen, Ontario, Canada)

3.2.9 Polysaccharide lyase activity assays

To test for potential polysaccharide lyase activity of the purified protein, I exposed SB1003 resuspended in G buffer, as normally used in gene transfer bioassays, to

either 25 or 55 μg of BSA or purified Rcc00171. After 1 h, the samples were treated for a phenol-sulfuric acid carbohydrate quantification and a variation of Anthony's capsule stain (Brimacombe *et al.*, 2013). In also purified *R. capsulatus* SB1003 polysaccharide using cetyltrimethylammonium bromide (CTAB) as described for *E. coli* K5 (Clarke *et al.*, 2000). Capsular polysaccharides (12.5 μg) were then treated with 0.018-0.9 μg of Rcc00171 protein for 30 min in G buffer, and the resulting polysaccharides were run on a 15% polyacrylamide gel and visualized by a combined Alcian blue-silver staining method (Fau & Cowman, 1986).

3.3 Results

3.3.1 Identification of genes with transcript level changes according to changes in RcGTA production

All microarray data (ALS1 early stationary phase (ES), DE442 ES, SBRM1 ES, SB1003 logarithmic phase) were compared to SB1003 ES. These were chosen for comparison because they each differ significantly in RcGTA production with respect to SB1003 ES (Schaefer *et al.*, 2002; Mercer *et al.*, 2010; Solioz *et al.*, 1975; Yen *et al.*, 1979), and would allow me to identify genes that have modified transcript levels under conditions where RcGTA production is known to be affected. Genes with a signal intensity differential of >2-fold relative to SB1003 ES were assembled into gene lists and were compared to identify shared gene sets (Figure 3-1). This identified the genes whose regulation matched that expected for RcGTA genes. The four separate transcriptome comparisons to SB1003 ES are clearly distinct, each with their own sets of genes co-

regulated with RcGTA. A total of 26 genes were differentially regulated across all four transcriptome comparisons (Figure 3-1, Table 3-3). For actual fold-change values for each of these genes, see Table A-1. These 26 co-regulated genes were examined in detail and found to include all 17 annotated RcGTA structural gene cluster open reading frames (ORFs), in addition to 9 genes in six separate loci outside this cluster (Table 3-3). The only putative RcGTA structural gene cluster ORF absent was *rcc01699*, a small ORF at the end of the cluster annotated as *orfg16*. Two of the three-way comparisons contained 0 genes while 6 genes were shared in all but the DE442 comparison; and 13 genes were shared in all but the ALS1 comparison (Table A-2).

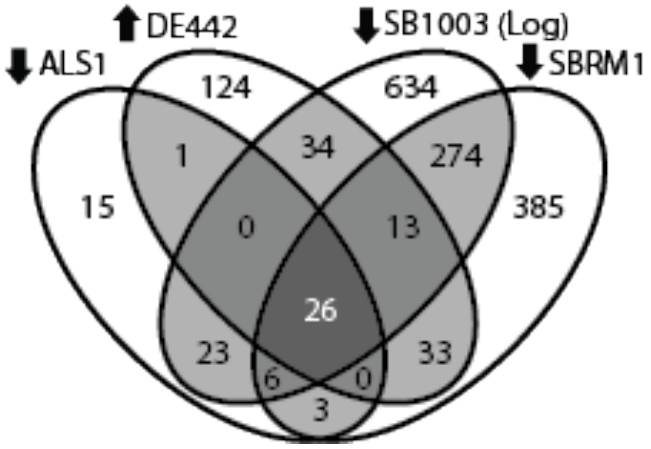

Figure 3-1: Identification of genes whose transcription varies with RcGTA production. Only genes with a >2 fold change in transcript level relative to the wild type in early stationary phase, as determined by comparison of microarray data, are counted for each comparison. The ovals contain genes downregulated in ALS1, SBRM1 or SB1003 (logarithmic phase), where production of RcGTA is known to be lower (down arrows), or up-regulated in the overproducer DE442, where production is higher (up arrow). The central (darkest) overlap of 26 genes consists of all the genes whose transcript level fluctuates with RcGTA production in all four tested comparisons.

Table 3-3: *R. capsulatus* genes co-regulated in transcriptome comparisons

Gene Name	Description (genome annotation)	BLAST	Notes	Involvement in RcGTA Activity
*rcc00171**	CHP	ribonuclease III	Attachment	Yes (This Study)
rcc0555	HP	Lysozyme family protein	Lysin	Yes (Hynes *et al*. 2012)
rcc0556	HP	DUF3154	Holin?	
*rcc01079**	CHP	tail fibre protein, putative	Tail Fibre	Yes (This Study)
*rcc01080**	CHP	tail fibre protein, putative	Tail Fibre	Yes (This Study)
rcc01682	CHP		*orfg1*	Yes (This study)
rcc01683	terminase-like		*orfg2*	Yes (Lang & Beatty, 2000)
*rcc01684**	portal, HK97 family		*orfg3*	
rcc01685	CHP		*orfg3.5*	No (Hynes *et al*. 2012)
rcc01686	prohead protease, HK97 family		*orfg4*	Yes (Lang & Beatty, 2000)
*rcc01687**	major capsid protein, HK97 family		*orfg5*	Yes (Florizone, 2006)
*rcc01688**	phage CHP	phiE125 gp8 family	*orfg6*	
*rcc01689**	phage CHP	phage head-tail adaptor, putative	*orfg7*	
*rcc01690**	CHP		*orfg8*	
*rcc01691**	major tail protein, TP901-1 family		*orfg9*	
rcc01692	CHP	DUF3356	*orfg 10*	
rcc01693	phage CHP	DUF2376	*orfg 10.1*	

*rcc01694**	phage CHP	phage tail minor protein	*orfg 11*	
rcc01695	phage CHP	DUF2460	*orfg 12*	Yes (Lang & Beatty, 2000)
*rcc01696**	phage CHP	DUF2163, Phage_BR0599	*orfg 13*	
rcc01697	cell wall peptidase, NlpC/P60 family	Cell wall-associated hydrolases	*orfg 14*	Yes (Fogg *et al.* 2012)
*rcc01698**	phage CHP	TIM-barrel-like domain, tail protein	*orfg 15*	Yes (Lang & Beatty, 2000)
rcc01865	HP	DNA replication initiation ATPase		
rcc01866	HP			
rcc02623	HP			No (This Study)
rcc02730	HP			

Rows highlighted in grey indicate genes located in the RcGTA structural gene cluster.

CHP = Conserved hypothetical protein.

HP = Hypothetical protein.

Genes marked with (*) indicate proteins found by (Chen *et al.*, 2008) in conjunction with purified RcGTA particles.

3.3.2 Identification of genes with roles in RcGTA activity

In a previous study, I investigated the lysis-related locus (*rcc00555-556*) and showed that *rcc00555* was required for RcGTA release from cells (Hynes *et al.*, 2012). I created mutant strains with disruptions of four of the remaining of the genes identified as possibly being co-regulated with the structural cluster, as well as of *rcc01699*; the only gene thought to be an RcGTA gene and absent from my lists. Of these, only the strains with disruptions of *rcc00171*, *rcc01079*, and *rcc01080* were affected for RcGTA gene transfer activity (Figure 3-2A). The amount of capsid protein found in cells and culture filtrates of these affected strains were then compared by western blotting (Figure 3-2B). The disruption of *rcc00171* completely abolished RcGTA activity without affecting capsid protein levels (Figure 3-2). The disruptions of *rcc01079* and *rcc01080* significantly reduced RcGTA activity without affecting the capsid protein accumulation inside or outside the cells (Figure 3-2). All mutations found to affect RcGTA activity were complemented *in trans*, although in the case of *rcc01079* and *rcc01080* it was necessary to introduce both genes to rescue RcGTA activity of either mutant. The *rcc01079* gene contains, in the latter portion of the gene, a predicted -1 frameshift signal identified by FSfinder (Moon *et al.*, 2004), which would result in a single *rcc01079/1080* gene product.

3.3.3 Roles of identified proteins in RcGTA function

The observation that gene transfer activity was affected in SB0171, SB1079 and SB1080 without an associated change in production and release of capsid protein led me

to suspect the genes were involved in binding of the RcGTA particles to cells. Additionally, the Rcc01080 protein sequence shows sequence similarity to known phage tail fibre proteins in the GenBank database. I performed binding assays and found that binding of particles from the *rcc00171* mutant strain was comparable to that of particles from the wild type (Figure 3-3). However, binding of particles from the *rcc01079* and *rcc01080* mutants to cells was significantly less than that of RcGTA from SB1003 (Figure 3-3). For these mutants, the decrease in free capsid protein was statistically undetectable.

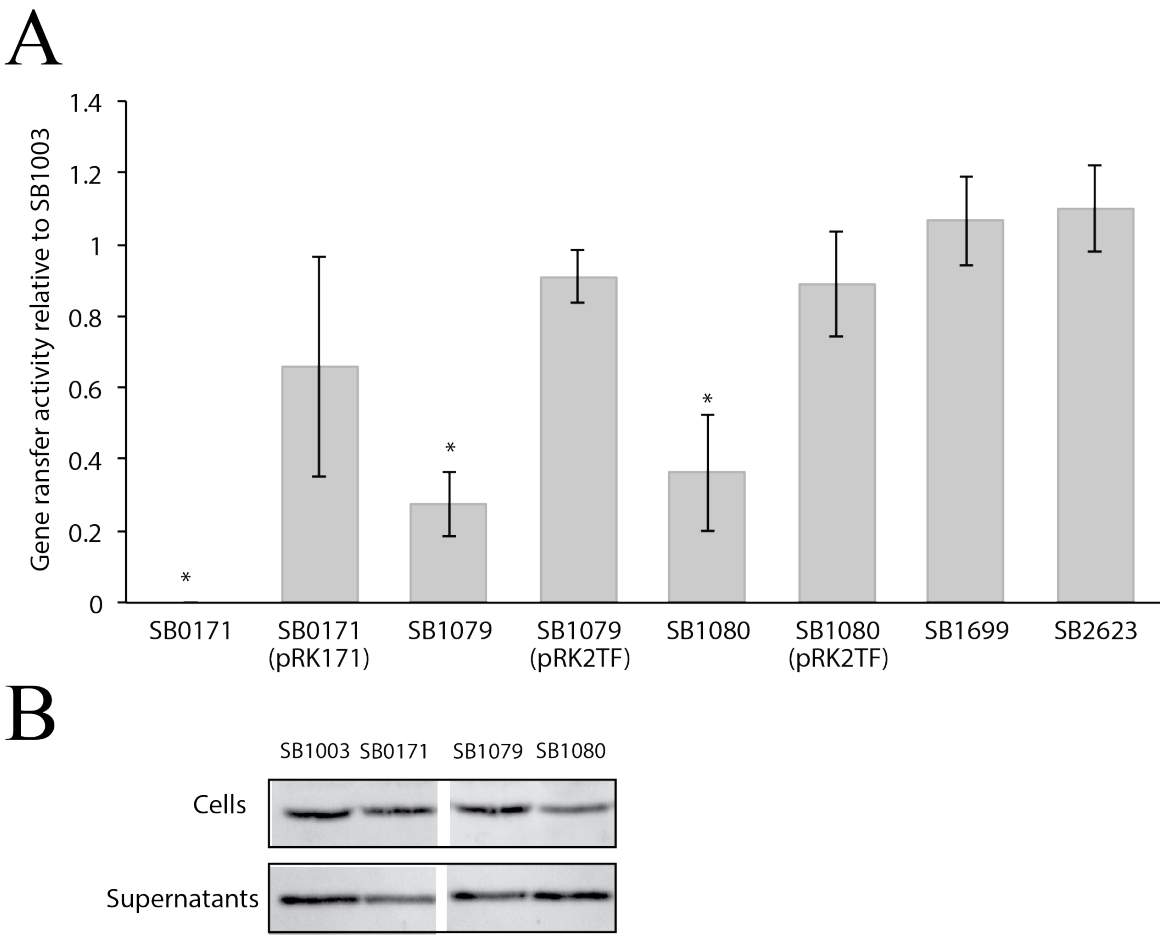

Figure 3-2: Effects of gene disruptions on RcGTA gene transfer activity and protein levels. **A.** Gene transfer activity of mutant strains and their associated complements. Complementation of SB1079 and SB1080 required re-introduction of both genes on pRK2TF The activity was determined as an average relative to SB1003 in three replicate bioassays, and the bars represent the standard deviation. An asterisk (*) denotes RcGTA gene transfer levels that differed significantly from the wild type ($P < 0.001$) determined by analysis of variance (ANOVA) and Tukey HSD. **B.** The relative abundance, as visualized by a representative western blot, of RcGTA capsid protein in the cells (top) and supernatants (bottom).

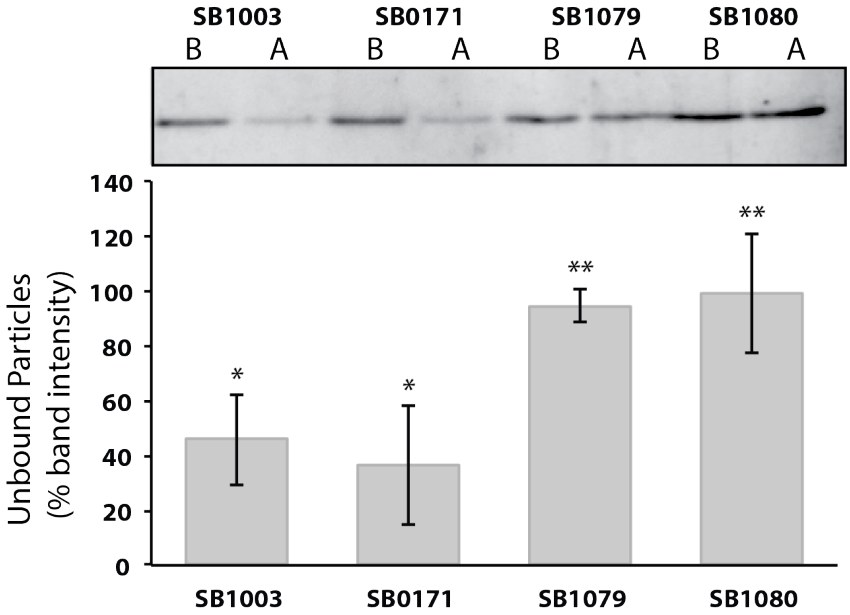

Figure 3-3: Binding efficiency of RcGTA particles produced by SB1003, SB0171, SB1079 and SB1080. The binding is visualized by (Top) a representative western blot comparing free capsid before (B) and after (A) exposure to recipient cells for 1 h followed by pelleting the cells, and measured by (bottom) quantified band intensity ratios of before/after over three replicate blots. (*) and (**) represent statistically distinct groups in which every member is significantly different from those of the other group ($p < 0.05$), but not from those within its group (ANOVA and Tukey HSD).

The finding that neither production nor binding of the particles was impaired in the *rcc00171* mutant prompted me to investigate the DNA within the particles produced by this mutant. DNA extractions confirmed that RcGTA particles produced by the *rcc00171* mutant contained DNA of the expected size (Figure 3-4). We could not find any justification for the annotation of homologous proteins as "ribonuclease III", so we began investigating other potential roles of the protein. I purified the Rcc00171 protein and performed an electrophoretic mobility shift assay (EMSA) to test for possible Rcc00171-DNA interactions. Incubation of Rcc00171 with either a 500-bp PCR product or purified RcGTA DNA showed no evidence of DNA binding by Rcc00171 (data not shown). As

the location of the *rcc00171* gene is adjacent to genes involved in polysaccharide export, I also used the Rcc00171 protein to test its potential as a capsular or extracellular polysaccharide (EPS) lyase. Incubation of cells with the purified protein produced no measurable decrease in cell-associated sugars as measured by phenol-sulfuric acid carbohydrate quantification and no visible effects upon the capsular structure as observed by microscopy of capsule-stained cells. Incubation of the purified protein with purified *R. capsulatus* EPS preparations yielded no evidence of breakdown of sugars into smaller units when visualized on an acrylamide gel (data not shown).

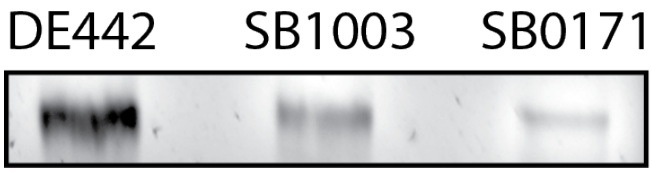

Figure 3-4: DNA extracted from RcGTA particles from DE442 (overproducer), SB1003 (wild type) and SB0171 strains, visualized by agarose gel electrophoresis. DNA yields obtains from preps is highly variable, so this non-quantitative analysis serves only to confirm the presence of DNA within the particles.

I used the purified protein to attempt rescue of the gene transfer activity for Rcc00171-deficient RcGTA particles. The purified protein was added to gene transfer bioassays with culture filtrates from the donor strains SB1003 and SB0171. Addition of the protein did not restore gene transfer activity to particles from strain SB0171. However, addition of the protein inhibited the gene transfer activity of SB1003 particles in a concentration dependent manner, with addition of 7.5, 15 and 30 μg of protein,

resulting in ~27%, ~17% and ~5% activity relative to the untreated control, respectively (Figure 3-5).

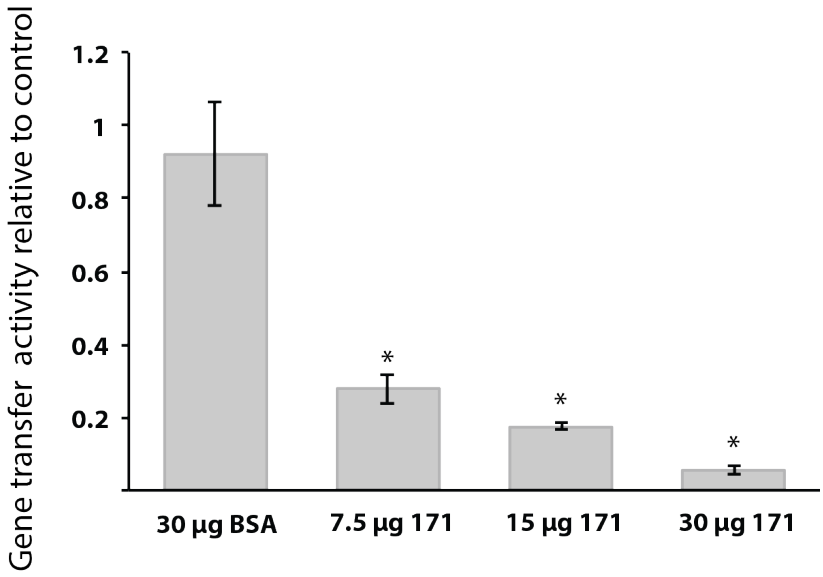

Figure 3-5: Inhibition of gene transfer activity in the presence of purified Rcc00171. Gene transfer activity was determined as an average relative to a no-treatment control in three replicate bioassays, and the bars represent the standard deviation. An asterisk (*) denotes RcGTA gene transfer levels that differed significantly from the wild type ($P < 0.05$) determined by analysis of variance (ANOVA) and Tukey HSD.

3.4 Discussion

Analysis of transcriptome data from four different pair-wise comparisons identified 26 genes as possibly co-regulated, their transcript levels varying with changes in RcGTA production (Figure 3-1). Of these, 17 are the ORFs within the RcGTA structural gene cluster. The other 9 genes are located in six separate loci. One of these loci, *rcc00555-rcc00556*, has already been characterized in some detail with *rcc00555* predicted to encode an endolysin and shown to be required for cell lysis for RcGTA release, and *rcc00556* suspected to encode an associated holin (Hynes *et al.*, 2012) (see

Chapter 4). I have generated strains with gene disruptions in five of the remaining genes to identify those involved in RcGTA production.

Among the 26 genes identified were 3 encoding proteins found to be associated with RcGTA particles by proteomics (Chen *et al.*, 2008), *rcc00171*, *rcc01079* and *rcc01080*. I have shown that these three genes are required for proper RcGTA function. My experiments implicate two of the proteins as tail fibres based on the properties of the RcGTA particles produced by strains lacking the genes (Figures 3-2 and 3-3). One of these proteins possesses sequence homology to known phage tail fibre proteins and both have homology to predicted tail fibres in the *R. capsulatus* RcapMu (Fogg *et al.*, 2011; accession number NC_016165) and RcapNL (NC_020489) phages. Insertional disruption of either *rcc01079* or of *rcc01080* resulted in a decrease of ~70% in gene transfer activity (Figure 3-2), which was accompanied by a decrease in binding ability (Figure 3-3) consistent with defective adsorption. This phenotype could only be rescued by *in trans* complementation of the individual mutants with both genes. With the presence of a conserved -1 frameshift signal (Moon *et al.*, 2004), it is possible the two ORFs are co-translated into a single Rcc01079/1080 protein. Interestingly, these two ORFs are located immediately upstream of genes required for production of the capsular polysaccharide RcGTA receptor (Brimacombe *et al.*, 2013).

The protein encoded by *rcc00171*, Rcc00171, is required for gene transfer activity (Figure 3-2). The Rcc00171 protein sequence contains a domain of unknown function (DUF 2739) at the N terminus (Marchler-Bauer & Bryant, 2004), and this gene was previously annotated as “ribonuclease III”. This annotation suggested its function might be associated with nucleic acid binding, but particles from the *rcc00171* mutant strain

contain DNA (Figure 3-4) and I found no evidence for DNA binding by the purified protein. Furthermore, I was unable to identify any justification for the ribonuclease III annotation. The impairment of gene transfer activity in the absence of Rcc00171 is not because of defects in capsid production, release, (Figures 3-2 and 3-4) or binding to cells (Figure 3-3). The ability of particles lacking this protein to bind to cells as efficiently as the wild type particles indicates it is required for a post-adsorption function. Like the tail fibres adjacent to capsule synthesis genes (discussed above), the two genes upstream of *rcc00171* are predicted to be involved in capsular polysaccharide export. This might imply *rcc00171* possesses a function related to the capsular polysaccharide. However, my attempts to demonstrate polysaccharide lyase activity with purified Rcc00171 were unsuccessful in three separate assays. In addition, the inability of the purified protein to rescue the gene transfer activity of Rcc00171-deficient particles does not support a role for the protein as a lyase. The inhibition of gene transfer of wild type RcGTA particles by the addition of Rcc00171 to RcGTA-cell mixtures (Figure 3-5) suggests another role entirely. I hypothesize that this protein is involved in attachment to a specific receptor following tail fibre mediated adsorption. The addition of extraneous Rcc00171 is presumably competing for and blocking attachment sites and therefore reducing the gene transfer rates in proportion to the concentration added.

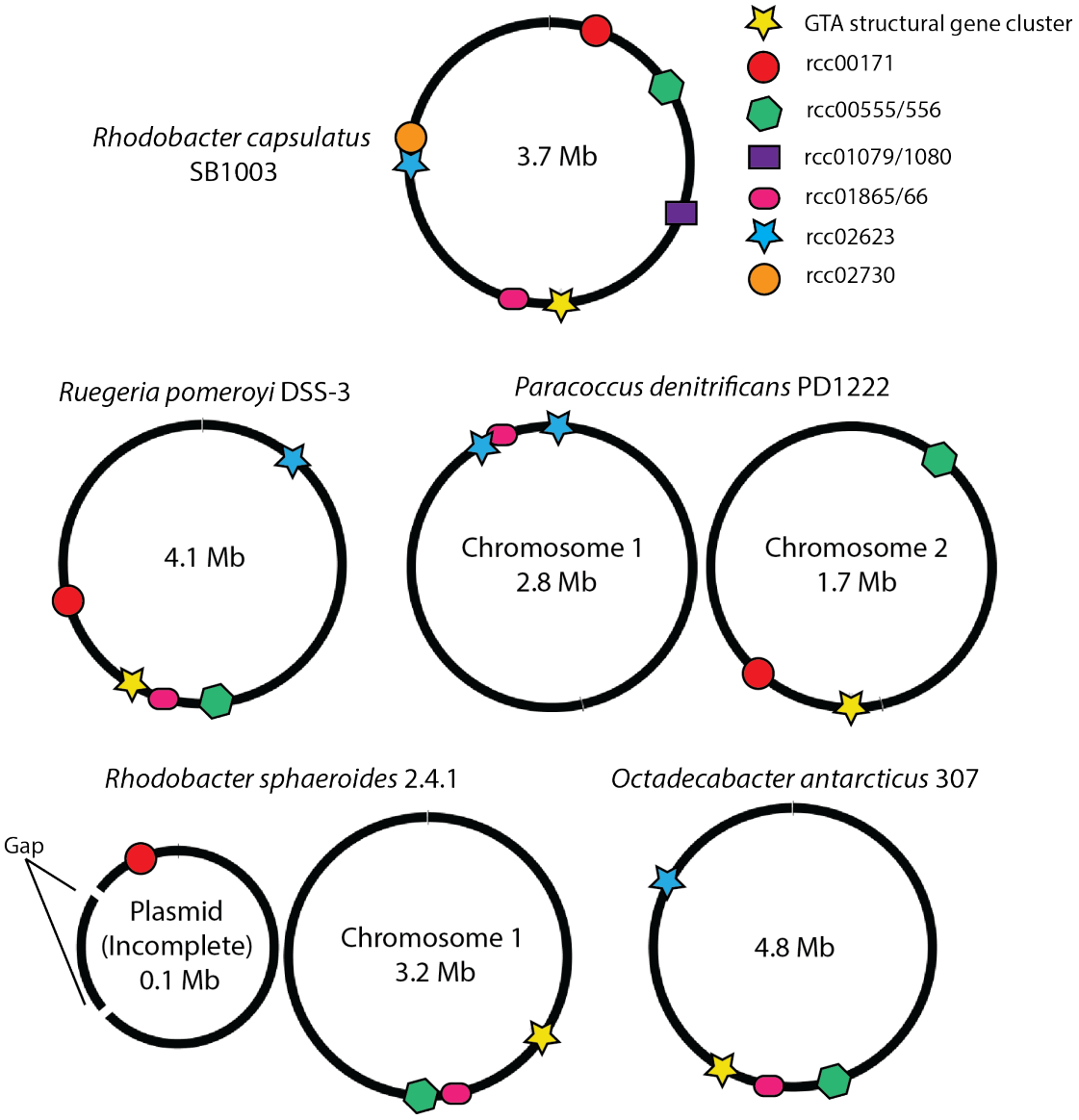

Figure 3-6: Comparative genomics of loci identified to be co-regulated with the RcGTA structural gene cluster. Each locus is identified by a coloured shape, and its relative position in the genome indicated. Any plasmids or chromosomes containing none of the loci of interest have been omitted. The genomes chosen for comparison, with accession numbers of the relevant replicons in brackets, are those of *Ruegeria pomeroyi* DSS-3 (NC_003911), *Paracoccus denitrificans* PD1222 (NC_008686 and NC_008687), *Rhodobacter sphaeroides* 2.4.1 (NC_007493 and NC_009007) and *Octadecabacter antarcticus* 307 (NC_020911).

One notable ORF not on my list of 26 genes is *rcc01699,* which is immediately downstream of *orfg15*. In several annotations of the RcGTA structural gene cluster, it is included as part of the cluster (Lang & Beatty, 2000; Lang & Beatty, 2001; Lang *et al.*, 2002). The RcGTA structural gene cluster is organized in a head-to-tail manner, similar to the organization of other phages (Casjens *et al.*, 1992). The genes that typically follow directly after the large "host specificity protein" gene (e.g. lambda *J*) are tail fibres. Insertional disruption of *rcc01699* had no effect on RcGTA activity (Figure 3-2). As this ORF is not co-regulated with the RcGTA gene cluster or required for gene transfer activity, I conclude it is not an RcGTA gene. Another notable gene not on my list was *rcp00136*, encoding one of the proteins identified by proteomics as associated with RcGTA particles (Chen *et al.*, 2008). This gene was not identified in any of my transcriptome pair-wise comparison lists and is also absent from the RcGTA overproducer strain DE442, which lacks the plasmid carrying this gene (Hynes *et al.*, 2012).

Within the class *Rhodobacterales*, the RcGTA structural cluster is widely conserved (Lang & Beatty, 2007; Lang *et al.*, 2012; Biers *et al.*, 2008). I examined the genome sequences of representatives from four separate genera known to contain complete RcGTA structural gene clusters for evidence of these nine ORFs which may be additional components of the RcGTA "genome". These included *Ruegeria pomeroyi* DSS-3, in which GTA-mediated gene transfer has been reported (Biers *et al.*, 2008) as well as *Paracoccus denitrificans* PD1222, *Rhodobacter sphaeroides* 2.4.1 and *Octadecabacter antarcticus* 307. In each of these, at least half of the additional loci were identified and they were always distributed throughout the genome, in some cases located

on different replicons (Figure 3-6). The tail fibre locus was not conserved, which is unsurprising because tail fibres mediate specific adsorption to the cells and are more likely to be conserved in unrelated phages infecting the same host than in related phages infecting different hosts (Haggård-Ljungquist *et al.*, 1992). The *rcc00555-rcc00556* lysis module and the *rcc01865-rcc01866* module were the only two to be conserved in every genome examined, although *rcc00171* was absent only from *O. antarcticus 307*. The homologue of *rcc00171* in *P. denitrificans* is of interest because it is located adjacent to a duplicated version of the RcGTA *orfg15* and a fragment of a transposase gene. This may reflect the ancestral state of the RcGTA cluster with *rcc00171* immediately downstream of *orfg15*, a position consistent with a role in interaction with recipient cells. This position downstream of *orfg15* is conserved in *Roseobacter denitrificans* phage RDJLΦ1 (Huang *et al.*, 2011), which also contains a gene with homology to the known RcGTA regulator *ctrA*.

Gene transfer agents are remarkable because of their inability to package sufficient DNA to transfer their "genome" to a recipient cell. For RcGTA, this genome was previously thought of as a ~15-kb gene cluster. I have now shown that RcGTA is actually encoded by at least two additional loci (*rcc00171* and *rcc01079-rcc01080*) and previous work (Hynes *et al.*, 2012) demonstrated dependence on a third (*rcc00555-rcc00556*). Homologues of these genes are also present in other organisms carrying related GTA gene clusters, and identification of these additional genes in other species may help to guide efforts to identify other active GTAs in the *Rhodobacterales*. In *R. capsulatus*, these additional RcGTA genes show the same transcriptional patterns as

the main structural gene cluster, demonstrating a conserved mechanism of regulation that has yet to be elucidated.

4 DNA Packaging Bias and Differential Expression of Gene Transfer Agent Genes Within a Population During Production and Release of the *Rhodobacter capsulatus* Gene Transfer Agent, RcGTA

4.1 Introduction

Gene transfer agents (GTAs) are phage-like particles encoded in prokaryotic genomes that transfer fragments of a producing cell's genome to recipient cells (Stanton, 2007). The process is analogous to transduction, but GTAs are differentiated from transducing phages by two main features: GTAs appear to always contain DNA from the cell's genome, and they always package smaller amounts of DNA than are known or predicted to encode the particles (Stanton, 2007; Lang & Beatty, 2007). The first GTA identified was RcGTA, discovered as a DNase-resistant and protease-sensitive mediator of gene transfer in cell-free filtrates of *Rhodobacter capsulatus* cultures (Marrs, 1974). RcGTA particles look like small tailed phages (Yen *et al.*, 1979), but cultures producing RcGTA do not exhibit detectable lysis (Solioz *et al.*, 1975). We know of no reported instance where a tailed phage particle has been found to exit a cell without lysis. Screening of a transposon library for mutants that had lost the ability to make functional RcGTA identified an approximately 15-kb gene cluster on the *R. capsulatus* chromosome encoding the RcGTA particle (Lang & Beatty, 2000), genes *rcc01682-1699* (GenBank accession numbers AF181080 and NC_014034). This RcGTA structural gene cluster has an organization conserved in tailed phages (Casjens *et al.*, 1992; Lang & Beatty, 2000).

Previous work with RcGTA showed that the particles contain ~4-kb of linear double-stranded DNA (dsDNA) (Solioz & Marrs, 1977; Yen *et al.*, 1979), which had similar GC content, as quantified using $CsCl_2/CsSO_4$ gradients (Solioz & Marrs, 1977), and complexity, as determined by hybridization kinetics and restriction analyses (Yen *et al.*, 1979), to that of the *R. capsulatus* genome. RcGTA packages DNA from all replicons in donor cells including introduced plasmids (Scolnik & Haselkorn, 1984). These findings have led to the assumption that RcGTA packages *R. capsulatus* DNA at random.

DNA packaging mechanisms in dsDNA phages are dependent on the action of an endonuclease complex known as the terminase. Terminases are responsible for initiation of packaging, translocation of the DNA into the particle, and the cutting of the DNA to complete packaging (Black, 1989). The activity of terminases can be categorized according to the nature of the ends of the packaged DNA, a grouping that corresponds closely with terminase phylogeny (Casjens *et al.*, 2005). The RcGTA gene cluster contains a recognizable large terminase-encoding gene, *orfg2* (Lang & Beatty, 2000). If RcGTA does indeed package DNA at random, its packaging must be by a "headful" packaging mechanism assisted by a non-sequence-specific terminase. The best-studied example of such packaging is that of phage T4, where each head is packed full with 1.02 genome lengths of T4 DNA (Streisinger *et al.*, 1967) and the large terminase appears to have no sequence specificity (Bhattacharyya & Rao, 1994).

We have performed quantitative analyses of the DNA packaged within RcGTA particles and single-cell RcGTA gene expression levels. The identification and disruption of a putative lysis gene involved in release of RcGTA particles from cells lead us to propose a mechanism by which the variation in RcGTA gene expression among cells in a

population explains an observed packaging bias and the lack of observable lysis within a RcGTA-producing culture.

4.2 Experimental procedures

Table 4-1: *R. capsulatus* strains and experimental plasmids used in this study

Strains and plasmids	Description	References
R. capsulatus strain		
SB1003	Genome-sequenced strain	(Yen and Marrs, 1976; Strnad *et al.*, 2010)
DE442	RcGTA overproducer, *crtD* (provenance uncertain; believed to be derived from RcGTA overproducer Y262)	(Yen *et al.*, 1979; Fogg *et al.*, 2011)
SB1685	SB1003 with KIXX insertion in *rcc01685*	This study
DE1685	DE442 with KIXX insertion in *rcc01685*	This study
SB2539	SB1003 with KIXX insertion in *rcc02539*	This study
DE2539	DE442 with KIXX insertion in *rcc02539*	This study
SB555	SB1003 with KIXX insertion in *rcc00555*	This study
DW5	SB1003 Δ*puhA*	(Wong *et al.*, 1996)
Plasmid		
pXPB	*pucB*'::'*lacZ* fusion	This study
pX2	RcGTA *orfg2'::'lacZ* fusion	This study
pX2NP	Promoterless RcGTA *orfg2'::'lacZ* fusion	This study
pX3	RcGTA *orfg3'::'lacZ* fusion	This study
pX3NP	Promoterless RcGTA *orfg3'::'lacZ* fusion	This study
pR555	*rcc00555* and 193 bp of 5' sequence in KpnI site of pRK767	This study

4.2.1 Bacterial strains, growth conditions, and plasmids

The strains of *R. capsulatus* used in this study are listed in Table 4-1. Strains carrying the kanamycin resistance marker were created as follows. The ORFs of interest were amplified by PCR and cloned in pGEM-T Easy (Promega, Madison, WI). The 1368-bp SmaI fragment of the KIXX cartridge (Barany, 1985), encoding resistance to kanamycin, was then ligated into a restriction enzyme cut site within the cloned PCR product. The primers used for the *rcc02539* construct were 5'-TTCCATGCCGAAATAGGCCGC-3' and 5'-GGCGCCGTCGTCGATCTGAAT-3', and the KIXX fragment was ligated into a SmaI site. The primers used for the *rcc01685* construct were 5'-AACGGGATGGGACTGAATTT-3' and 5'-ATGTCACCAGCGACACTTCC-3', and the KIXX fragment was ligated into an Eco47III site. The primers used for the *rcc00555* construct were 5'-AACGAGGTTTTCCTGGAGGT-3' and 5'-AACCTGTTCCGCAAGATCAC-3', and the KIXX fragment was ligated into a SmaI site. These plasmids were independently transferred into *R. capsulatus* by conjugation from *E. coli* C600 (pDPT51) (Taylor *et al.*, 1983), and the kanamycin resistance genes transferred to the chromosome of recipient *R. capsulatus* cells by RcGTA transfer (Scolnik & Haselkorn, 1984). Successful transfers of the gene were confirmed by PCR using the same primer pairs and template DNA from the resultant kanamycin resistant RcGTA recipients, which showed 1.4-kb larger products than the non-disrupted versions.

R. capsulatus cells were grown under anaerobic photoheterotrophic conditions in complex YPS medium (Wall *et al.*, 1975b) at 35°C for RcGTA production bioassays,

purification of particles for DNA isolations, and purification of RNA for microarray analysis. For all other purposes, they were grown aerobically at 30°C in RCV medium (Beatty & Gest, 1981).

The experimental plasmids are listed in Table 4-1. They consist of an in-frame fusion of the chosen ORF to *lacZ* in the promoter probe vector pXCA601 (Adams *et al.*, 1989), created using the BamHI and PstI sites in the vector and adding corresponding sites to the amplification primers for cloning. pXB was created using the primers 5'TGCCTGCAGAAAGATGCGTCTGGAACACC-3' and 5'-GGGGATCCCCATCGATCAGGTAGCTGTG-3'; pX3 and pX3NP were created using the forward primers 5'-CGGCTGCAGACCGATCCGG-3' and 5'-ATACTGCAGCATGGACATGGGGTTCAA-3', respectively, and the reverse primer, 5'-AGGATCCCCCGTGCGCATCAGACTGAC-3'; pX2 and pX2NP were created using the same forward primers used for pX3 and pX3NP, respectively, and the reverse primer, 5'-AGGATCCACGTCGCGCACCTGAT-3'; underlined bases represent the restriction sites added for cloning. All constructs were created from sequences amplified from the genome-sequenced strain SB1003, and the fusions were confirmed as in-frame by sequencing. We sequenced the same RcGTA upstream region amplified from the RcGTA overproducer DE442, confirming it was identical to the SB1003 sequence.

Complementation of the *rcc00555* mutant was carried out with *rcc00555* and its native promoter, as amplified by the primers 5'-ATGGTACCATGGTCGAGGGCACCTTT-3' and 5'-ATGGTACCCCAGGATCGTCCCGATC-3', ligated into the broad host-range vector pRK767 (Gill & Warren, 1988) using KpnI sites (underlined in the primer sequences).

4.2.2 RcGTA DNA isolation

Cultures of strain DE442 were grown for 48 h. The cells were then centrifuged at 5855 *g*, the supernatant filtered through a 0.45-µm polyvinylidene fluoride (PVDF) filter (Millipore, Bedford, MA), and the filtrate ultracentrifuged at 184000 *g* for 5 h. The resulting pellet was resuspended by shaking at 100 RPM in G buffer (Solioz & Marrs, 1977) overnight at 4°C. The resuspensions were treated with 2 units RNase-free DNase I (New England Biolabs, Pickering, Canada) and 1.2 units RNase A (Sigma-Aldrich, Oakville, Canada) in 1X DNase Buffer (New England Biolabs) at 37°C for 30 min to remove any free nucleic acids, and then incubated at 75°C in the presence of 5 mM EDTA (pH 8). DNA was purified by phenol:chloroform:isoamyl alcohol (25:24:1) extraction and ethanol precipitation. The sample was subjected to agarose gel electrophoresis and the ~4-kb RcGTA DNA band extracted using the QIAEX II Gel Extraction Kit (Qiagen, Mississauga, Canada).

4.2.3 Microarray analyses

The *R. capsulatus* microarrays are Affymetrix whole-genome expression arrays (Affymetrix, Santa Clara, CA) that contain oligonucleotide probes for 3635 ORFs (Mercer *et al.*, 2010). For the RNA analysis, cells were harvested 16 h after reaching stationary phase, as determined by monitoring culture turbidity, and RNA was extracted using the RNeasy Kit (Qiagen) as described (Mercer *et al.*, 2010). The RNA and DNA samples were processed for cDNA synthesis and fragmentation, respectively, and subsequent labeling and array hybridization at the Michael Smith Genome Science Center

(Vancouver, Canada) as described in the Affymetrix Expression Analysis Technical Manual for prokaryotic samples.

Raw data from the RcGTA DNA array were extracted using the MAS5 algorithm with detection calls (Pepper *et al.*, 2007) to generate signal intensity. Statistical analyses of the raw data were carried out using Minitab 15 (Minitab, State College, PA). Raw data from the RNA arrays were robust multi-array (RMA) normalized (Irizarry *et al.*, 2003) and normalized to the 50th percentile using GeneSpring 7.2 (Agilent Technologies, Santa Clara, CA).

The microarray data from this study have been deposited in the NCBI Gene Expression Omnibus (GEO) database (accession number GSE33176).

4.2.4 RcGTA activity bioassays

Cultures of the test strains were grown aerobically overnight, and then normalized for density and used to inoculate RcGTA-production bioassay cultures. These cultures were then grown for 48 h and filtrates were collected using 0.45-μm PVDF syringe filters (Millipore). Filtrates were assayed for RcGTA activity using strain DW5 as the recipient as follows. An overnight aerobic culture of DW5 was centrifuged and the cells resuspended in an equal volume of G buffer (Solioz *et al.*, 1975). Equal volumes of donor filtrate and recipient cells were mixed with four volumes of G buffer and incubated for 1 h at 35°C with shaking. Nine volumes of RCV medium were then added and the mixtures incubated for a further 3 h before plating. Each bioassay was plated in equal parts on YPS and YPS with kanamycin sulfate (10 μg ml^{-1}). The YPS plates were grown under anaerobic phototrophic conditions to select for transfer of the *puhA* gene while the

kanamycin-containing plates were incubated aerobically in the dark to select for transfer of the resistance marker. Colonies on the plates were counted after 2 days, and the ratios of transfer of kanamycin resistance to *puhA* in four independent assays were calculated. The transfer rates were compared by one-way ANOVA and Tukey HSD test. In lysis assays, a 1 ml portion of each culture was centrifuged and the cells resuspended in 30 µl of 20 mM Tris-HCl, 5 mM EDTA, 250 mM sucrose (pH 7.8) containing 0.5 mg ml^{-1} lysozyme (Sigma-Aldrich). After three freeze-thaw cycles in dry ice-ethanol, 1 ml of 20 mM Tris-HCl, 0.5 mM $MgCl_2$ (pH 7.8) containing 0.1 mg ml^{-1} DNase (Sigma-Aldrich) was added to the cells and the mixtures incubated for 5 min before filtration using a 0.45-µm PVDF filter (Millipore). The filtrates were then used for gene transfer bioassays as described above.

4.2.5 Western blots targeting the RcGTA major capsid protein

Cells and culture filtrates from the same cultures used in RcGTA activity bioassays were assayed for RcGTA capsid protein by western blotting. Cultures were centrifuged at 17000 *g*, the supernatant was removed and the cells re-suspended in an equal volume of TE buffer. For the different samples, 5 µl of the cell suspensions, 10 µl of culture filtrates, and 10 µl of the cell lysates were run. SDS-PAGE, blotting, and detection of the RcGTA major capsid protein were done as described (Mercer *et al.*, 2012) with the primary antibody AS08 365 (Agrisera, Vännäs, Sweden). Images were captured on a gel documentation system and subsequently inverted and adjusted for brightness and contrast using Adobe photoshop CS5.0 (Adobe Systems Inc, San Jose, CA).

4.2.6 RcGTA gene expression in single cells

R. capsulatus cultures containing the fusion constructs (Table 4-1) were grown until 4 h after reaching stationary phase, and analyzed for β-galactosidase activity. Cells were permeabilized by exposure to 15% (v v^{-1}) isopropyl alcohol for 15 min and then washed with Z buffer (60 mM Na_2HPO_4, 40 mM NaH_2PO_4, 10 mM KCl, 1 mM $MgSO_4$, 50 mM β-mercaptoethanol; pH 7) (Miller, 1992). Fluorescein di-β-D-galactopyranoside (FDG) (Sigma-Aldrich) in H_2O:DMSO:ethanol (8:1:1) was added to a final concentration of 0.1 mg ml^{-1}. The cells were incubated for 1 h and subsequently diluted 1:200 in Z buffer and analyzed by flow cytometry recording 100 000 events. These events were gated, according to forward and side scatter, to identify >90% of events as "cells". These assays were repeated three times with independently grown cultures.

4.2.7 Bioinformatic analyses

The RcGTA large terminase protein sequence was used to perform a BLAST search against the nr database (Wheeler *et al.*, 2007) by both psi-blast and blastp (Altschul *et al.*, 1990; Altschul *et al.*, 1997). The selected terminase protein sequences were aligned using Clustal X (Larkin *et al.*, 2007).

4.2.8 Treatments of RcGTA DNA for ligation experiments

RcGTA DNA (1 μg) was treated with 3 units T4 DNA polymerase (New England Biolabs) at 12°C for 15 min as per the manufacturer's recommendations. An equivalent sample was incubated with 2 μl M-MuLV RNase H^+ RT solution from the Phusion RT-PCR kit (Finnzymes, Espoo, Finland) supplemented with 0.5 mM dNTPs according to the

manufacturer's recommendation at 40°C for 30 min, and the enzyme then heat-inactivated by incubation at 85°C for 5 min. Both reactions, alongside a negative control that contained the same amount of RcGTA DNA in TE buffer in the same total volume, were cleaned using the QIAquick cleanup kit (Qiagen), and the DNA was eluted in 30 µl of elution buffer. 26 µl of these eluates were independently treated with 800 cohesive end units of T4 DNA Ligase (New England Biolabs) according to the manufacturer's recommendations at 16°C for 18 h. The samples were then subjected to agarose gel electrophoresis.

4.3 Results

4.3.1 Genome-wide quantification of DNA packaged in RcGTA particles

To quantify the packaging of each gene from the *R. capsulatus* genome, DNA extracted from RcGTA particles harvested from DE442, an RcGTA overproducer of uncertain ancestry (Table 4-1), was hybridized to a whole-genome microarray [NCBI Gene Expression Omnibus (GEO) database accession number GSE33176]. All 3645 chromosomal ORFs present on the microarray were present in the particles. The raw signal intensities varied from 265.6 to 838.9, and the average signal intensity was 545.7 ± 72.2. The microarrays are based on the genome-sequenced strain, SB1003 (Strnad *et al.*, 2010), which contains a ~130-kb plasmid. The data for the plasmid showed that 156 of 157 ORFs had very low signals, with an average signal intensity of 6.1 ± 7.8. There was one exception, *rcp00051*, which had an intensity of 671.7. We examined the genome

sequence and found that a 99% identical paralog of *rcp00051* is located on the chromosome, *rcc01445*. Therefore, other than the single ORF that is duplicated on the chromosome, no plasmid genes were detected in the RcGTA DNA and gel electrophoresis confirmed strain DE442 lacks the plasmid (data not shown).

Signal intensities for the ORFs on the array are shown as a histogram (Figure 4-1, bottom). The frequency distribution of the signal intensities was unimodal. A probability plot of the data (Figure 4-1, top) showed that variation from normal occurs in the top and bottom 1% of signal intensities. Too many genes were packaged infrequently (signal intensities <~378) and too few genes were packaged most frequently (signal intensities >~714) for a strictly normal distribution. We examined a variety of properties of the genes that were most and least often packaged: predicted gene function, orientation, location in the genome, GC content, and transcript levels. Such examinations of the 100 most frequently packaged and 100 least frequently packaged ORFs revealed no obvious patterns or trends. Plotting of the RcGTA packaging signal, GC content, and transcript levels against genome position (Figure 4-2) identified a region with a pronounced drop in packaging frequency that corresponded with a spike in transcript levels (Figure 4-2A and C; approximately position 1700). This corresponded to the RcGTA gene cluster, and this region had the lowest average packaging in the genome with the moving average window of 20 ORFs. The average packaging intensity of these genes was only 433.8 ± 66.8. This region also showed an obvious differential expression in the RcGTA overproducer strain relative to the wild type (Figure 4-2D).

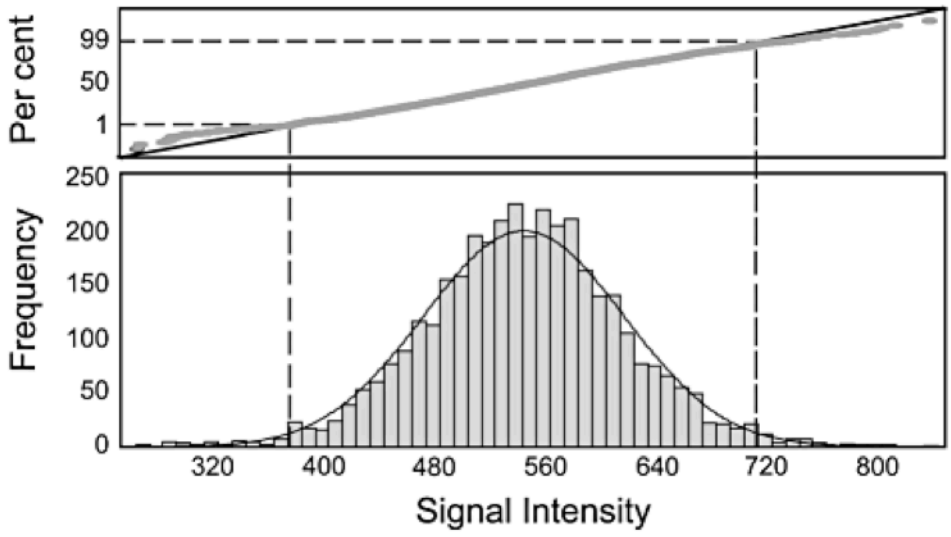

Figure 4-1: Distribution of signal intensities from hybridization of DNA packaged in RcGTA particles to an *R. capsulatus* microarray. A quantile-quantile plot is shown on the top and a frequency histogram on the bottom. On the bottom, genes are grouped within signal intensity ranges of 10 (e.g. 119 ORFs had a signal intensity between 470.0 and 479.9). The highest signal intensities represent genes packaged most often, the lowest those packaged least often. On the top, the percentages represent the fraction of total signal intensities that fall within a range (e.g. 98% of the signals were between 380 and 720). The top and bottom 1% of signal intensities are demarcated with dashed lines. The solid black lines represent normal distributions and points of departure are seen where the actual data in gray deviate from these lines.

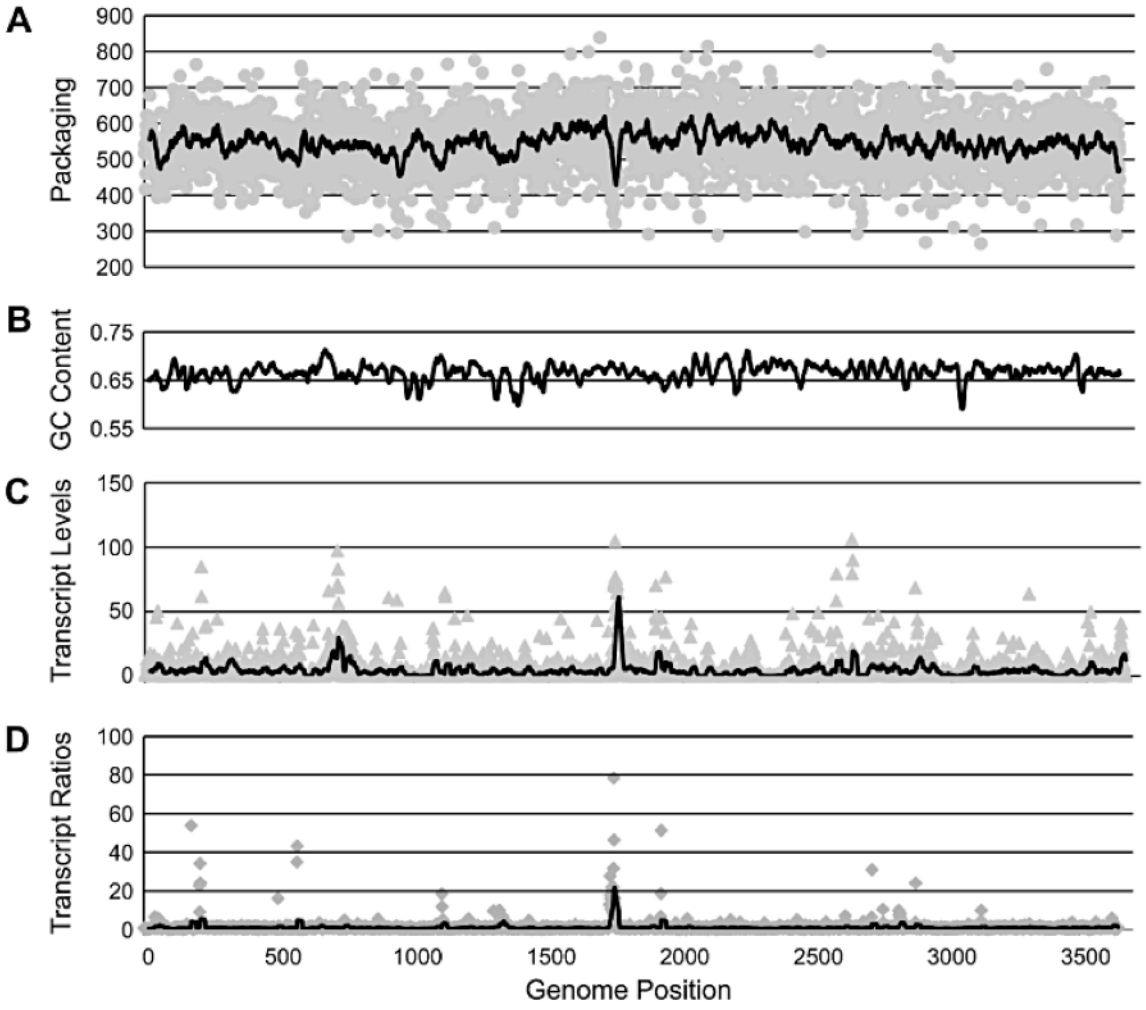

Figure 4-2: Relationships between chromosomal location and RcGTA packaging frequency, transcript levels, and GC content. The packaging signal intensity **(A)**, GC content **(B)**, normalized transcript signal intensity of the RcGTA overproducer **(C)**, and ratio of transcript levels in the RcGTA overproducer relative to wild type **(D)** are plotted versus *Rhodobacter capsulatus* genome position. In (A), (C), and (D) genome position represents the relative position of each ORF (i.e. position 500 is the 500th ORF from base 1) and each point represents one ORF; all trend lines represent a moving average with a window of 20 ORFs. In (B), genome position represents the moving average over 20 of the 1026-bp windows used to calculate GC content.

4.3.2 Transfer rates of genes from different locations in the *R. capsulatus* genome

To confirm the microarray-based observation of under-packaging of the RcGTA genes, we compared the transfer frequencies of a kanamycin resistance marker inserted in

the chromosome inside and outside the RcGTA gene cluster. The insertion inside the RcGTA gene cluster is located in a putative ORF (*rcc01685*) between the RcGTA genes encoding the predicted portal (*orfg3*, *rcc01684*) and protease (*orfg4*, *rcc01686*) proteins. This ORF had a packaging intensity of 409.5 on the RcGTA DNA array. The insertion outside of the cluster is located in gene *rcc02539*, predicted to encode a c-di-GMP signaling protein. This ORF had a packaging intensity of 582.8 on the RcGTA DNA array. Transfer rates were normalized to the transfer of *puhA* (*rcc00659*, packaging intensity of 674.6 on the RcGTA DNA array), the photosynthetic reaction centre H protein-encoding gene deleted in the transfer assay recipient strain, DW5. This approach was taken to normalize the kanamycin resistance transfer frequencies to the transfer of an independent marker representing the total RcGTA production by a strain in a given experiment. Neither kanamycin resistance marker insertion significantly affected RcGTA production, as measured by comparison of the *puhA* transfer rates. In both wild type (SB1003) and RcGTA overproducer (DE442) backgrounds, transfer of the marker outside of the RcGTA gene cluster occurred at ~40% of that of *puhA*, whereas transfer of the marker inside the RcGTA gene cluster occurred at ~20% (Figure 4-3). These differences were statistically significant (ANOVA, Tukey HSD test, $p<0.01$), while the differences between the rates of transfer from equivalent SB1003- and DE442-derived strains were not ($p>0.05$).

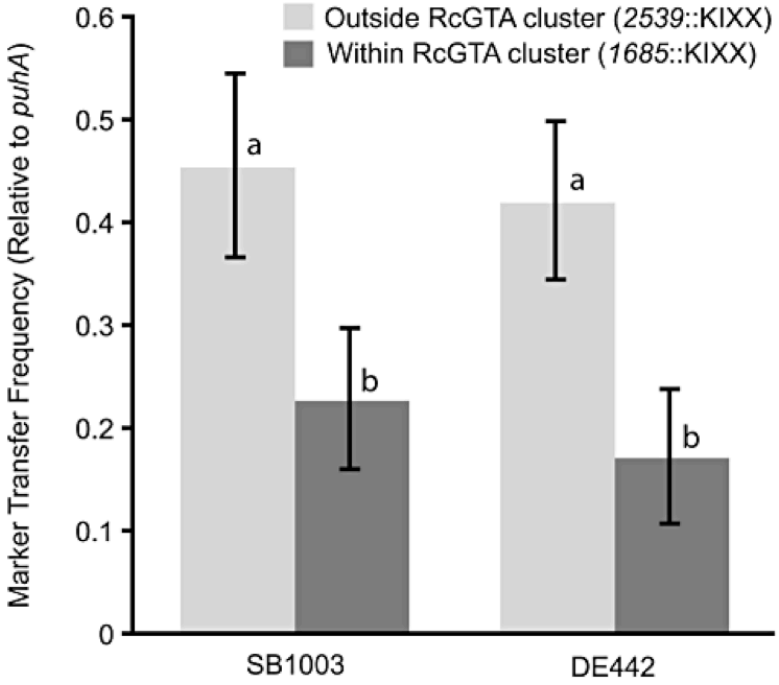

Figure 4-3: Frequency of RcGTA-mediated transfer of the kanamycin resistance marker when located inside (*rcc01685*::KIXX) or outside (*rcc02539*::KIXX) the RcGTA gene cluster. Assays were performed with the marker inserted into the genome of both the wild type (SB1003) and RcGTA overproducer (DE442) strains. The transfer of the kanamycin resistance marker in these locations was normalized to the transfer of the *puhA* gene by the same strains in the same assays. The data are shown as averages from four replicate gene transfer bioassay experiments and the bars represent the standard deviation. Each letter (a,b) indicates a group whose members are not statistically different from one another, but are different from the other group.

4.3.3 RcGTA gene expression in single cells

We hypothesized that the packaging bias against the RcGTA gene cluster might be linked to the coincident high level of transcription of these genes (Figure 4-2C). However, other regions showing similar localized high transcript levels (e.g. region at position ~700 on Figure 4-2A), or high ratios of transcript levels in DE442 versus SB1003 (e.g. region at position ~200 on Figure 4-2D) did not have a corresponding

decrease in packaging. We therefore hypothesized that the RcGTA genes might be differentially expressed within the population, and that perhaps only a subset of the population is responsible for all of the RcGTA expression. If so, population-wide expression arrays would underestimate transcript levels in the cells actually expressing the RcGTA genes. In these cells, high-occupancy of the RcGTA genes by the transcriptional machinery might limit access by the RcGTA packaging machinery, thereby causing a decrease in the packaging of these genes relative to other regions.

To test this hypothesis, we analyzed RcGTA gene expression at the single-cell level using plasmid-borne translational fusions to a *lacZ* reporter gene. Fusions were constructed to two different RcGTA genes, *orfg2* encoding the terminase protein and *orfg3* encoding the portal protein, because of previously observed differences in the transcript patterns of these genes by microarray analyses (Mercer *et al.*, 2010). Both fusions contained the same sequences upstream of the RcGTA gene cluster, and negative controls lacked the predicted promoter regions. An independent fusion to a photosynthesis gene, *pucB* encoding the light-harvesting complex 2 β protein, was constructed as a control. The β-galactosidase activities of the gene fusions were assayed in RcGTA overproducer (DE442) and wild type (SB1003) cells by flow cytometry (Figure 4-4). The *pucB* (Figure 4-4A) and *orfg2* (Figure 4-4B) fusions produced similar unimodal patterns in both SB1003 and DE442. The *orfg3* fusion, however, had considerably higher signals in some cells and showed a clearly multimodal distribution, with extended tails of cell counts with increased fluorescence (Figure 4-4C) not observed for the other fusions. Both SB1003 and DE442 had subsets of higher *orfg3* expression,

but the expression levels and numbers of highly expressing cells are greater in the overproducer.

4.3.4 Identification of the putative RcGTA lysis gene

The sub-population expression phenotype identified above might explain the lack of observed lysis in RcGTA-producing cultures. If only a small subset of cells is responsible for the majority of RcGTA production, these cells could lyse and release the particles. One of the genes up-regulated in the RcGTA overproducer relative to wild-type (Figure 4-2D), *rcc00555*, encodes a putative N-acetylmuramidase lysozyme protein that contains a variation (E-X_8-D-X_4-T) on the conserved catalytic residues present in the N terminus of many phage endolysins, E-X_8-D/C-X_5-T (Sun *et al.*, 2009). The downstream and overlapping gene, *rcc00556*, is similarly up-regulated in DE442 and may encode a holin protein, which would be required for such an endolysin to access the peptidoglycan. It is predicted to have three trans-membrane domains and a topology consistent with a lambda S-type holin (Young, 2002) but lacks the dual-start codons separated by a positively charged amino acid common to such proteins and has no homology to proteins of known function. There are no apparent phage-related genes in the genome near these two genes.

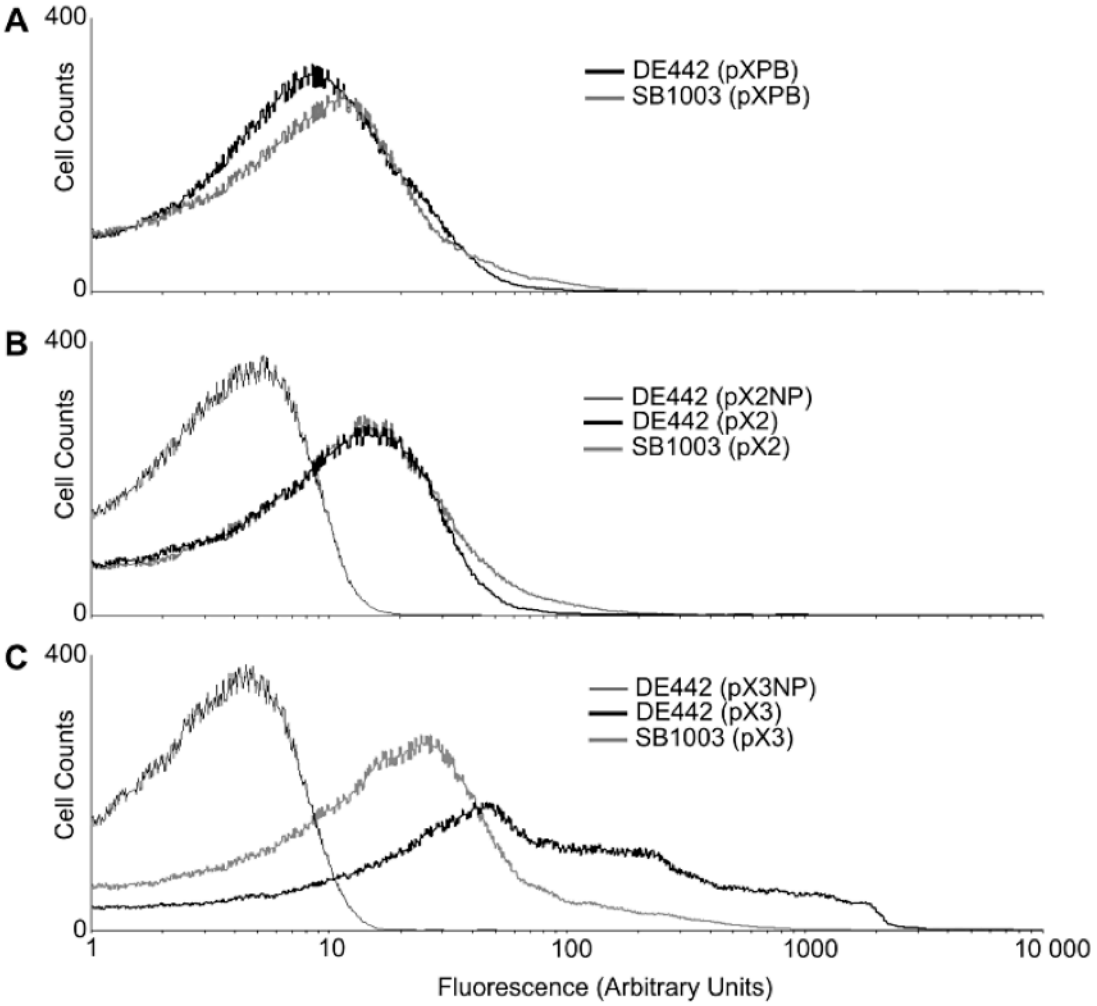

Figure 4-4: Population patterns in gene expression measured with reporter gene fusions. Gene expression was quantified for the different gene fusions within populations of cells using flow cytometry, recording 100 000 events. The assays were repeated independently three times and a representative experiment is shown for each set of strains. **A.** The control fusion of the photosynthesis gene *pucB* (pXPB) in DE442 and SB1003 (black and gray lines, respectively). The mean fluorescence values were 7.77 and 7.70 for DE442 and SB1003, respectively. **B.** The experimental fusion of the terminase-encoding *orfg2* (pX2) in DE442 and SB1003 (thick black and gray lines, respectively), and the promoterless control fusion (pX2NP) in DE442 (thin black line). The mean fluorescence values were 3.75, 10.18, and 10.55 for the promoterless, and experimental DE442 and SB1003, respectively. **C.** The experimental fusion of the portal-encoding *orfg3* (pX3) in DE442 and SB1003 (thick black and gray lines, respectively), and the promoterless control fusion (pX3NP) in DE442 (thin black line). The mean fluorescence values were 3.46, 50.3, and 16.11 for the promoterless, and experimental DE442 and SB1003, respectively.

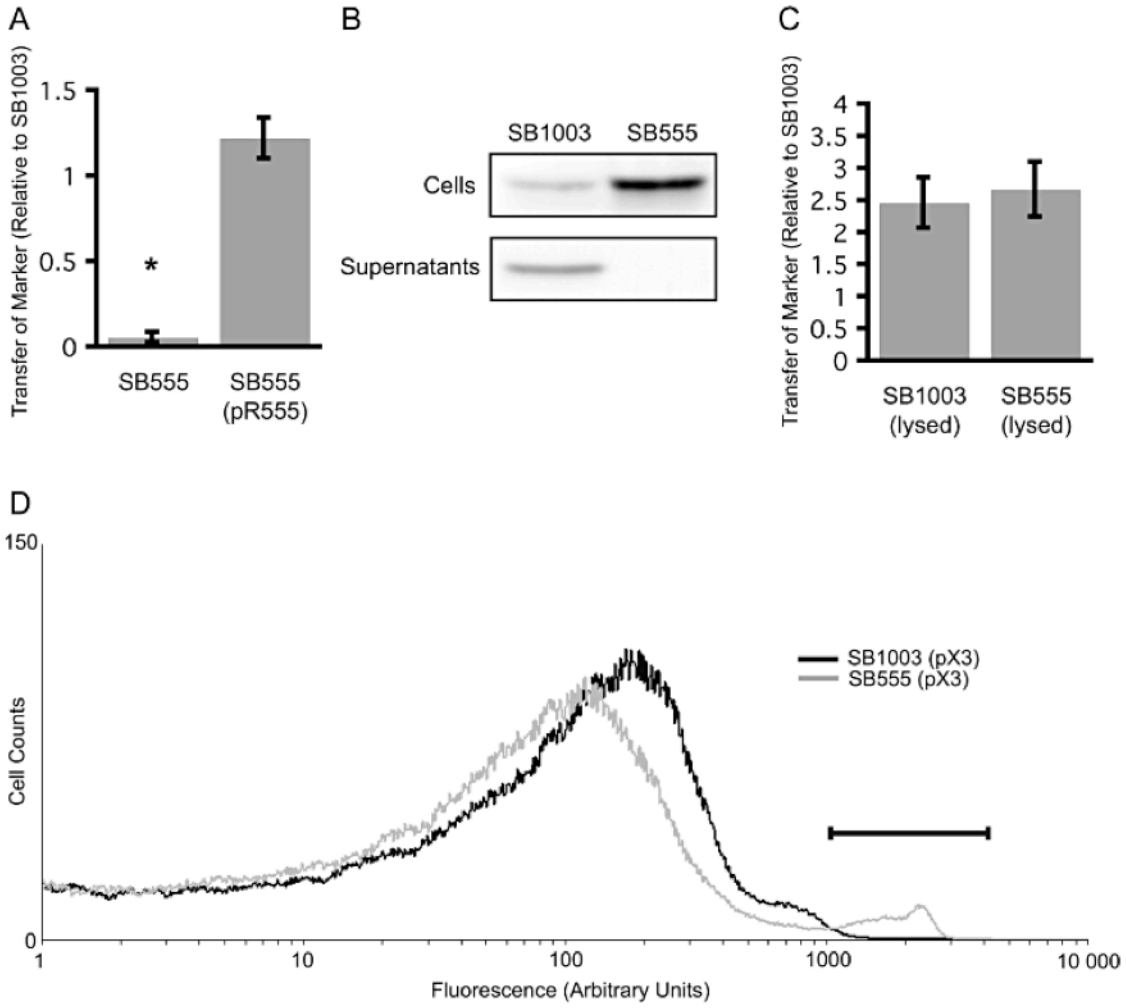

Figure 4-5: Identification of a putative endolysin required for release of RcGTA from cells. **A.** The frequency of gene transfer by the *rcc00555* mutant, SB555, and the plasmid-complemented strain. The gene transfer activity was determined as an average relative to SB1003 in five replicate bioassays and the bars represent the standard deviation. An asterisk (*) denotes RcGTA gene transfer levels that differed significantly from the wild type ($p<0.001$) determined by analysis of variance (ANOVA). The complemented strain did not differ from the wild type ($p=0.67$). **B.** The relative abundance, as measured by western blot, of RcGTA capsid protein in the cells (top) and culture supernatants (bottom) for SB1003 and SB555. Blots were performed on three replicate cultures and one representative set of blots is shown. **C.** Gene transfer activities in samples from artificially lysed SB1003 and SB555. The gene transfer activity was determined as an average relative to un-lysed SB1003 in three replicate experiments and the bars represent the standard deviation. ANOVA indicated these activities do not differ from one another ($p=0.78$). **D.** Expression of the *orfg3* fusion construct (pX3) in SB1003 and SB555 (black and gray lines, respectively). A black line above the peak in the SB555 strain demarcates the sub-population discussed in the text. Mean fluorescence values were 138.6 and 145.3

for the entire SB1003 and SB555 populations, respectively, and 1847 for the cells in the indicated sub-population. The assays were repeated independently three times and a representative is shown.

Insertional disruption of the putative endolysin resulted in a ~95% reduction in RcGTA gene transfer activity (Figure 4-5A). This decrease was associated with a decrease of RcGTA major capsid protein in the culture supernatants (Figure 4-5B) and an increase in intra-cellular capsid protein. To determine whether the intra-cellular capsid protein represented functional particles, the cells were artificially lysed and the lysates showed RcGTA gene transfer activity equivalent to the lysed wild-type strain (Figure 4-5C) confirming that the only RcGTA-related phenotype of the mutant is the inability to release functional particles trapped within the cells. Assay of the *orfg3* reporter construct in the *rcc00555* mutant revealed an accentuated sub-population of cells, 2.76% (± 0.53) of the population, expressing *orfg3* at a much higher level (9.02-fold ± 3.31) than the remainder of the population (Figure 4-5D).

4.3.5 Bioinformatic analysis of the RcGTA terminase

Using the RcGTA large terminase protein sequence for a BLAST search returned many high-scoring (score >500) sequences, all of which were from RcGTA-like elements in α-proteobacterial genomes and prophages. The top matches from phages were much weaker ($e>10^{-5}$), and these were all from γ-proteobacterial phages classified as "T4-like". Therefore, at the present time, the phage sequences most closely related to the RcGTA terminase are in the T4-like group, although the recognizable homology to these T4-like phage proteins is over only ~31% of the protein. There is no recognizable homology

between any other RcGTA protein and sequences from the T4-like group. An alignment of the homologous region for the RcGTA, phage Acj61 (the top phage BLAST match), and two other phages from the T4-like group whose packaging has been characterized (T4 and IME08), is shown (Figure 4-6). The RcGTA and Acj61 sequences are 28% identical (47% similar) over the aligned region, supporting an evolutionary connection between these sequences. We presume that the RcGTA *orfg1* encodes the small terminase subunit because of its location directly upstream of the large terminase, but it lacks recognizable sequence homology to any known phage sequence.

```
                 . : * :  **:.      *:: : * *:...* ::..  **        *::  .      *  :
RcGTA       25   ---PADAGPAHRVALVG---ETFDQVRDVMIFGESGILACSPPDRR-------PEWEAT------KRRLVW   75
Acj61      103   SKLSRQLGKTTAVAIFLAHYVCFNKAKSVGILAHKGSMSAEVLDRTKQAIELLPDFLQPGIVEWNKGNITL  169
T4         103   CNLSRQLGKTTVVAIFLAHFVCFNKDKAVGILAHKGSMSAEVLDRTKQAIELLPDFLQPGIVEWNKGSIEL  169
IME08      103   CNLSRQLGKTTVVAIFLAHFVCFNKDKAVGILAHKGSMSAEVLDRTKQAIELLPDFLQPGIVEWNKGSIEL  169

                   **::  **::. *:*:** .*   ::** *   .  : *  :* .:  *::.: ::****..:. :  *
RcGTA       76   WANGATAQAYSAQEPEALRGPQFDAAWVDELAKWRRAEETWDMLQFALRLGKHPQQVITTTPRNVGVLKAI  146
Acj61      170   LDNGSSIGAYASS-PDAVRGNSFSLIYIDEVAFVQNWEDCWLAIQPVISSGRHSKIVMTTTPNGMNHFYTI  239
T4         170   LDNGSSIGAYASS-PDAVRGNSFAMIYIDECAFIPNFHDSWLAIQPVISSGRRSKIIITTTPNGLNHFYDI  239
IME08      170   LDNGSSIGAYASS-PDAVRGNSFAMIYIDECAFIPNFLDSWLAIQPVISSGRRSKIIITTTPNGLNHFYDI  239
```

Figure 4-6: Alignment of a portion of the large terminase proteins from RcGTA to phages T4, IME08, and Acj61. The numbers indicate the amino acid residue positions in the original proteins. The presence of positively scoring positions is indicated above the aligned sequences as defined in CLUSTAL: "*" indicates a fully conserved residue, ":" indicates full conservation of a strong group and "." indicates full conservation of a weak group.

4.3.6 Characterization of the ends of the DNA molecules in RcGTA particles

We conducted ligation experiments to determine the structure of the ends of the DNA within RcGTA particles. Untreated RcGTA did not ligate with itself efficiently, and most of the DNA remained at the ~4-kb size (Figure 4-7). Treatment of the DNA with T4 DNA polymerase prior to ligation improved the efficiency considerably, while treatment with M-MuLV reverse transcriptase did not (Figure 4-7). As T4 DNA polymerase will

convert both 5' and 3' overhangs into blunt ends (Kucera & Nichols, 2008), the improvement in ligation efficiency indicates the ends of the DNA within the particles are neither blunt nor complementary overhangs. M-MuLV reverse transcriptase will fill 5' overhangs to make blunt ends but does not possess the 3'-5' exonuclease activity that would be required to make 3' overhangs blunt (Verma, 1975). Therefore, the ends of the DNA in RcGTA particles are 3' overhangs that are not consistent from particle to particle. Some ligation did occur in the absence of any end treatments (Figure 4-7).

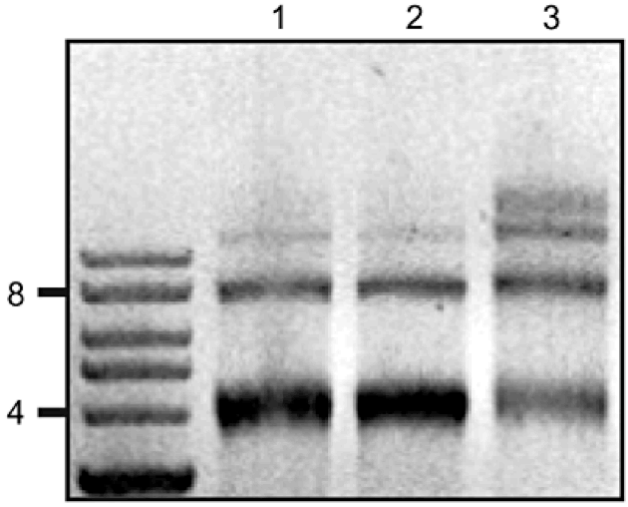

Figure 4-7: Ligation of DNA from RcGTA particles. Purified RcGTA DNA was treated with DNA ligase only (lane 1), M-MuLV reverse transcriptase followed by DNA ligase (lane 2), and T4 DNA Polymerase followed by DNA ligase (lane 3). A DNA ladder is shown on the left with the 4-kb and 8-kb bands indicated.

4.4 Discussion

Previous studies of the DNA inside RcGTA particles (Solioz & Marrs, 1977; Yen & Marrs, 1976) using low-resolution techniques suggested that RcGTA packages DNA from within the producing cell at random. Our data from hybridization of DNA from RcGTA particles to an *R. capsulatus* whole-genome microarray show that DNA

packaging by RcGTA is essentially random (Figure 4-1). The RcGTA particles contain every gene in the donor cell (Figure 4-2), and the terminase protein shows homology to known sequence-independent enzymes from phages in the T4-like group (Figure 4-6). Phage T4 is a well-characterized example of a phage that uses a non-sequence-specific headful packaging mechanism (Rao & Black, 2005). The limited, but recognizable, sequence homology between these terminases indicates a distant evolutionary connection between RcGTA and the T4-like phage proteins.

Sequence-independent headful packaging is thought to always result in blunt ends, as there is no requirement for cohesive end structures (Casjens & Gilcrease, 2009). Blunt ends have been demonstrated for phages P22 (Schmieger *et al.*, 1990), Mu (Morgan *et al.*, 2002), as well as T4 (Louie & Serwer, 1990). We expected to find similarly blunt-ended DNA within the RcGTA particles. However, ligation experiments with DNA from within RcGTA particles indicate the presence of 3' non-sequence-specific ends on the packaged DNA. The observation of a small amount of ligation in the absence of end-modifying treatments (Figure 4-7) most likely indicates that some matching cohesive ends are present at a low frequency and the 3' overhangs may be only several nucleotides in length. The discovery of non-matching end sequences and the random packaging data together support a model where the RcGTA terminase has no sequence specificity and the DNA molecules present inside producing cells act as "concatamer" substrates for headful packaging.

A putative *Bartonella* GTA capable of packaging all genes in the genome was recently identified and was found to preferentially package a chromosomal "high plasticity zone" (Berglund *et al.*, 2009). This region was associated with run-off

replication and therefore the packaging bias likely results from the increased relative copy number of certain sequences and not a packaging specificity. It might have been expected to find an increased packaging by RcGTA of the genes nearer to the origin of chromosome replication (*ori*), due to overall higher copy number in cells undergoing replication and division, but this was not observed (Figure 4-2, where *ori* is at position 0). However, RcGTA production is highest in the stationary phase of growth (Solioz *et al.*, 1975), which is also the phase at which the RcGTA particles were harvested, and so there would be little replication and division happening at this time in the culture. Our data from *R. capsulatus* fail to support the hypothesis that GTAs might preferentially package "cloud genes" (poorly conserved genes) (Kristensen *et al.*, 2010).

An examination of packaging frequency and a variety of other factors did not yield any obvious correlations. There was one notable exception, where the RcGTA gene cluster was the least frequently packaged region of the chromosome, at ~75% of the average. Therefore, RcGTA DNA packaging is not selective for RcGTA genes with occasional packaging of "host" DNA, as one might expect of a transducing prophage. The relative rate of transfer of markers inserted inside and outside of the RcGTA gene cluster confirmed the array-based quantification and extended this observation from the overproducer strain to the wild type (Figure 4-3). The higher transfer rates of the control marker *puhA* are most likely due to the smaller size of this marker, which requires transfer of 591 bp of non-homologous sequence to the recipient while the kanamycin resistance marker requires transfer of 1368 bp of non-homologous sequence to the recipient. This size difference would result in an increase in the frequency with which an intact copy of the *puhA* marker is packaged with sufficient flanking sequence to allow for

homologous recombination in the recipient cell, and may affect the efficiency of that recombination event.

The finding that under-packaging of RcGTA genes was correlated with localized high transcript levels (Figure 4-2) suggested there could be a link between gene expression and RcGTA packaging, and perhaps high occupancy of these genes by the cell's transcription machinery could protect them from packaging. However, there was no correlation between transcript levels and packaging intensity over the remainder of the chromosomal ORFs (Figure 4-2). The RNA expression microarray measures total transcript levels from the entire population and the results represent an average for each cell. Therefore, if only a subset of the population were transcribing the RcGTA genes at a high level, the arrays would yield an artificially low estimate of the transcript levels in those cells. Such sub-population expression patterns have been reported in other species (Avery, 2006; Lopez *et al.*, 2009). Analysis of translational reporter fusions to RcGTA *orfg3* by flow cytometry validated this hypothesis, as there was a multimodal distribution of gene expression levels in the population (Figure 4-4). The control fusion to the photosynthesis gene *pucB*, which would be expressed by all cells in these phototrophic growth conditions, showed a unimodal distribution. A fusion to RcGTA *orfg2* also was unimodal. It has previously been observed that *orfg2* and *orfg3* differ in their transcription patterns. Loss of the response regulator CtrA leads to loss of RcGTA production, but whereas no transcripts are detected for *orfg3*, some transcripts of *orfg2* are still detected in a *ctrA* mutant (Mercer *et al.*, 2010). This indicates that control of transcription and protein expression for different genes in the RcGTA gene cluster is more complex than previously realized.

The documentation of unequal RcGTA gene expression within a population may help explain the lack of observable cell lysis in cultures producing RcGTA. A tailed phage particle escaping from cells without lysis has never been reported, and it is presumably only the small subset of the population that is expressing the RcGTA genes at a higher level (Figure 4C) that are producing RcGTA and lysing to release the particles. In order to validate this hypothesis, we examined a list of genes co-regulated with RcGTA in the overproducer strain and identified a putative lysis gene with sequence homology to lysozyme proteins. Disruption of this gene resulted in a ~95% reduction in gene transfer activity (Figure 4-5A). This reduction is the result of lower levels of RcGTA in culture supernatants and an accumulation within the cells (Figure 4-5B). Manual lysis of this mutant released equivalent functional RcGTA to that from the lysed wild type strain (Figure 4-5C). These findings support the role of *rcc00555* as a gene involved in release of RcGTA. Furthermore, the presence of the *orfg3* reporter construct in the *rcc00555* mutant resulted in the appearance of a more pronounced sub-population of cells highly expressing *orfg3* (Figure 4-5D). This sub-population of ~3% of the total cells showed ~9-fold higher expression and is presumably responsible for almost all of the RcGTA activity in the culture (Figure 4-5A). The accentuation of this population in the *rcc00555* mutant must reflect lack of lysis in the highly expressing cells that would normally have lysed to release RcGTA particles. The lack of observed lysis in RcGTA-producing cultures is easily explained, given the small size of this sub-population responsible for release of the majority of the RcGTA particles.

All genes inside the producing cell are packaged inside RcGTA particles, although there is a slight but significant reduction in packaging of the RcGTA-encoding structural

gene cluster. The higher RcGTA gene expression in the subset of cells responsible for producing RcGTA particles could result in decreased access to these genes by the packaging machinery. There could be a selective advantage to this protection of the RcGTA genes in these cells, favoring their prolonged expression to maximize RcGTA production. The confirmation that cells are lysing to release RcGTA, an important cost for their production, is mitigated by the discovery that only ~3% of the cells in RcGTA-producing cultures are responsible for release of the majority of the particles.

5 Summary

There is very little known about any of the gene transfer agents, even the archetypal *Rhodobacter capsulatus* gene transfer agent (RcGTA). Prior to my research, we knew of its ability to package and transduce DNA seemingly at random (Solioz & Marrs, 1977; Yen *et al.*, 1979) at a high frequency (Yen & Marrs, 1976), of the existence of the "structural cluster" encoding the particles (Lang & Beatty, 2000), and of its regulation driven by entry into stationary phase (Solioz *et al.*, 1975) by means of host mechanisms including a complex phosphorelay (Lang & Beatty, 2000) and quorum-sensing (Schaefer *et al.*, 2002).

The non-replicative nature of RcGTA DNA transfer explains the absence of distinguishable replication, integration or excision genes in the structural gene cluster, but not the absences of a means of escaping the producing cell or of tail fibre-encoding genes. Given that GTAs have been found in Archaea (Bertani, 1999), spirochaetes (Matson *et al.*, 2005), α- (Marrs, 1974) and δ- (Rapp & Wall, 1987) proteobacteria, and that the α-proteobacterial lineage appears quite conserved (Lang & Beatty, 2007), GTAs present a very interesting evolutionary puzzle to which we have too few of the pieces. What is the cost of production? How many cells produce RcGTA, and when? What is the benefit to the individual cell producing RcGTA, or to its conspecifics?

To help answer these questions, I searched for the "missing" functions not encoded by the RcGTA structural cluster. Given that several satellite or "defective" phages are dependent on other phages for their "life" cycle (Christie & Dokland, 2012), the search began with the phages of *R. capsulatus* (Chapter 2). In purifying RcGTA

DNA, I observed the presence of a co-purifying DNA band much larger than the 4-kb expected for RcGTA-packaged DNA. The detection of this larger band was possible only in *R. capsulatus* strains capable of high levels of gene transfer activity, suggesting it was linked to RcGTA in some way. This larger band proved to consist of two ~40.3-kb phage genomes, those of RcapNL and RcapMu, which were sequenced and characterized. With the modification of my extraction protocols I was able to isolate a third, ~80.85-kb band, which I dubbed RcapH. It proved intractable to restriction enzyme digestion and sequencing, but appeared independently of RcGTA gene-transfer activity and so was of lesser interest. RcapMu mapped to the published *R. capsulatus* SB1003 genome sequence, and accounts for half of the ~80-kb "transposable phage cluster" annotated in Figure 1-1. RcapNL did not map to the published sequence, but I identified its integration site. Figure 5-1 depicts an updated map of the *R. capsulatus* genome's prophages.

I assayed a variety of RcGTA-affected mutants to investigate the link between RcGTA activity and phage production. I also knocked out both RcapMu and RcapNL, individually and in combination, and assayed the effects on RcGTA activity. They proved to have no effect on gene transfer activity, which establishes both that the link between RcGTA and the phages is not reciprocal and that the transduction activity of the phages is negligible in comparison to that of RcGTA. No obvious "missing" GTA functions appear to be encoded by these phages.

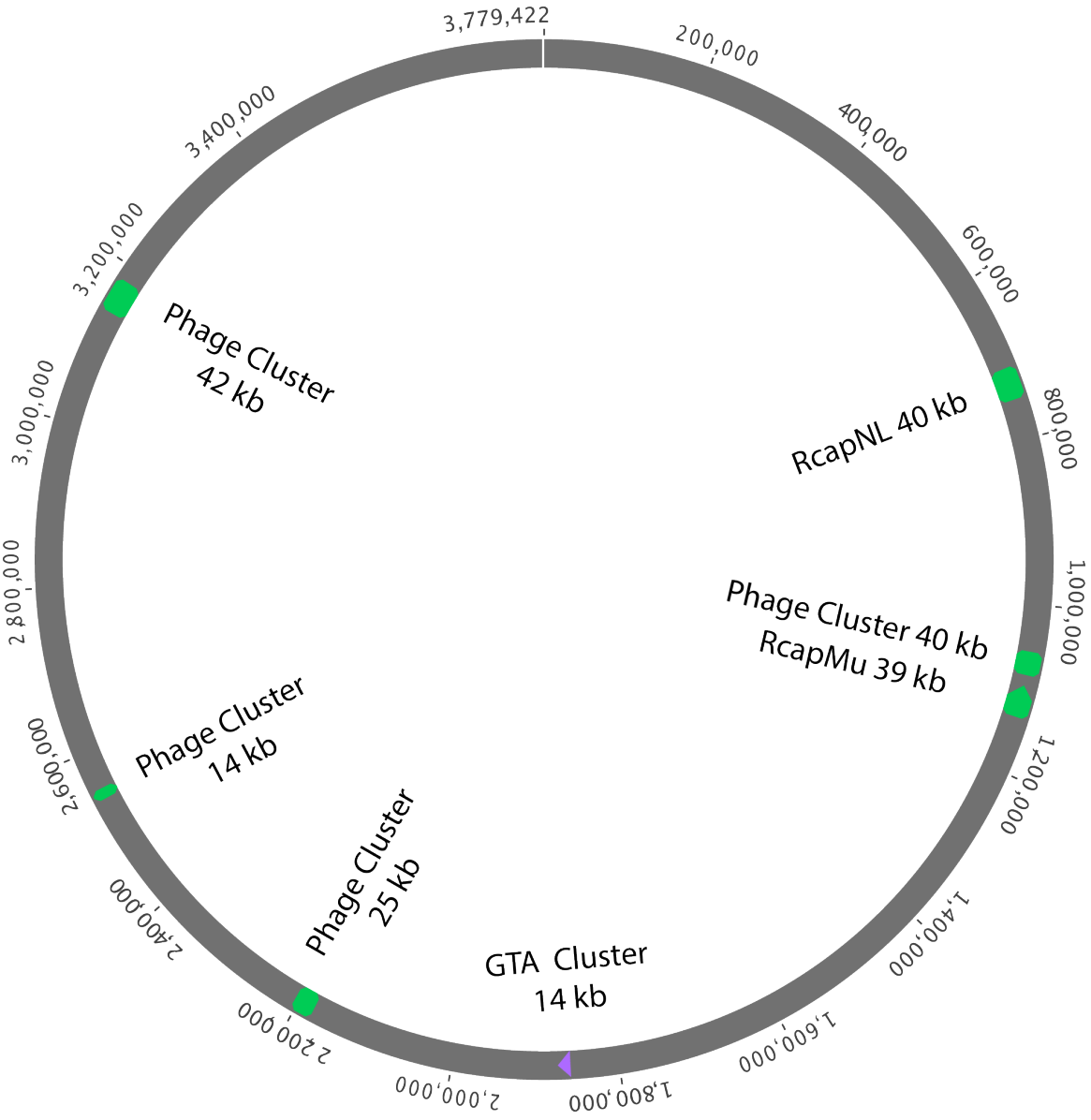

Figure 5-1: Prophage elements of *R. capsulatus* SB1003. The circle is a representation of the SB1003 chromosome, with base numbers from base 1 (white line) labeled. Each green region represents a cluster of phage-derived genes likely to be a prophage or prophage remnant, with the size of the cluster identified. The functional transposable prophage RcapMu forms roughly half of the previously annotated 80-kb "transposable phage cluster", while the prophage RcapNL is a new addition, absent from the published genome sequence. The structural cluster encoding the phage-like gene transfer agent (GTA) is highlighted in purple.

To search elsewhere for these "missing" functions, I turned to a transcriptomic analysis of the genome to find genes co-regulated with RcGTA activity (Chapter 3). In doing so, I identified 26 genes, 17 of which were the RcGTA structural cluster (Figure 5-2). The remaining 9, encoded in six additional loci, were investigated in detail. I found that *rcc01079/01080* encode tail fibre proteins, *rcc00171* is probably involved in non-reversible attachment, and *rcc00555/556* appear to be a lysin/holin pair (Chapter 4). The role of these additional elements has fleshed out many previously obscure aspects of the RcGTA "life" cycle.

The identification and characterization of these functions, essential for efficient RcGTA gene transfer activity, led me to search for these additional loci in other α-proteobacteria known to carry RcGTA structural clusters. They may serve as good additional indicators of functional GTAs. This analysis established that many of these additional loci are present in these other organisms, and they are consistently spread across multiple loci, in some cases across multiple replicons. Given that an unrelated GTA, VSH-1, is encoded by at least two separate loci, this may be a conserved feature of all GTAs.

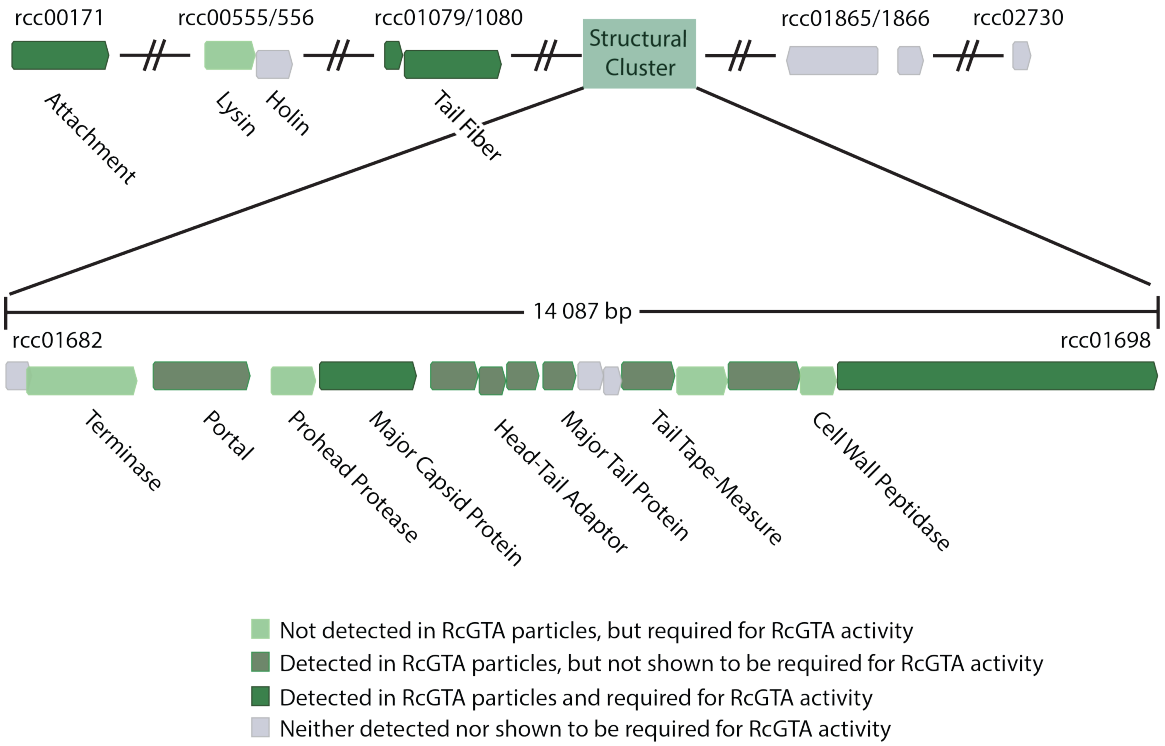

Figure 5-2: The RcGTA "genome". Depiction of the structural gene cluster and genes co-regulated with the structural cluster. The direction of transcription is indicated by the arrow depicting each gene. Vertical offset in neighbouring genes indicates different reading frames. All genes and gaps are to scale, with the exception of lines interrupted by two parallel lines, which indicate discontinuity. The RcGTA structural gene cluster spans from gene *rcc01682* to *rcc01698*, a total of 14 087-bp. Genes with known function or homology to a phage gene are annotated by function, and the first and last gene of each locus is identified by gene ID. Proteins detected in RcGTA particles are as reported by (Chen *et al.*, 2008), while those shown to be required for RcGTA activity are as reported by (Lang & Beatty, 2000; Fogg *et al.*, 2012; Hynes *et al.*, 2012) and Chapter 3.

The search for additional genes involved in RcGTA production presented me with a very important piece of the evolutionary puzzle: the involvement of *rcc00555* in lysing the RcGTA-producing cells to release the particles into the media. I was now in a position to establish the cost, and potentially the benefit, of RcGTA production.

I began by assaying which genes were being transferred by the particles to search for any biases that might indicate a system that evolved to transfer specific genes preferentially (Chapter 4), as conjugation systems (plasmids), transduction (phage) and transformation (specific uptake sequences) do. By analysis of the data from hybridizing DNA extracted from RcGTA particles to a whole-genome micro-array, I established that overall packaging trends appear random with one notable exception: the structural cluster responsible for RcGTA production is significantly under-packaged. This is precisely the opposite one would expect for a selfish or specialized transfer apparatus.

I was unable to find evidence for a mechanism actively depressing RcGTA structural cluster packaging. I proposed a mechanistic explanation in which only a subset of cells are responsible for the RcGTA production and that these cells are transcribing the genes involved in production at such a high level that the transcription machinery limits access by the packaging machinery to the structural cluster region. To confirm this, I constructed *lacZ* fusions to RcGTA structural cluster genes and monitored single-cell expression. I found evidence of a non-uniform distribution of RcGTA gene transcription in the population, which was greatly accentuated in a lysis-deficient background. These experiments implicated that 3% of the cells were transcribing the genes >10 fold more than the population average and are responsible for 95% of RcGTA activity. This sub-population of cells is difficult to observe in the wild type due to lysis.

Although I am still far from answering the evolutionary question: “why is RcGTA production conserved?”, I’ve gained considerable insight into the workings of RcGTA. I found two phages dependent on RcGTA production but on whose production RcGTA does not depend. I found several new loci involved in RcGTA production that leads me to

believe in a conserved "non-contiguous" RcGTA genome, consistently encoded across multiple loci. I discovered that RcGTA DNA packaging avoids transferring the genes responsible for its own production. I found that the cost to an individual RcGTA producing cell is lysis, and to a population is the lysis of ~3% of its members, at least in the conditions used in my experiments.

The only "benefit" of RcGTA production that has been uncovered to date is the gene transfer activity facilitating the horizontal exchange of DNA. It is not linked to bacteriocin activity (Wall *et al.*, 1975a), unlike PBSX in *B. subtilis* (Wood *et al.*, 1990). Whether facilitated horizontal gene transfer is a benefit at all is a topic of some contention (Redfield, 2001; Vos, 2009; Redfield, 1988; Redfield *et al.*, 1997). The cost of RcGTA production, however, is more easily quantified. The overproduction phenotype is very unstable, and frequently results in dramatic decrease in RcGTA production within a few sub-culturing events. While seemingly altruistic behaviours such as this could be explained by kin-selection, the existence of non-RcGTA producing cells (and strains) which are capable of receiving RcGTA complicate the issue. These "cheaters", given the cost of production, would rapidly take over the population unless a) there is a very strong selection for maintenance of RcGTA production or b) there is a means of re-introducing RcGTA production to "cheating" cells. The latter is significantly hampered by the fact that in addition to being unable to package the entirety of its structural cluster within a particle, RcGTA is encoded across multiple loci and cannot possibly, even if it were carried by a much larger transducing element, be re-introduced into "cheaters".

It is my belief that I am still missing a crucial part of the equation: some benefit conferred by RcGTA that far outweighs that of horizontal gene transfer. Examples might

include a function (directly or indirectly) such as an abortive infection (abi) mechanism that competes with infecting phages (or triggers the production of RcapNL and RcapMu, which compete) to reduce burst sizes, a role as a bacteriocin against a to-date unidentified bacterium competing for a similar ecological niche, or an effort by the host cell to preserve its DNA in harsh environments, akin to sporulation.

While I leave many important questions unanswered, the work elaborated here has helped elucidate many features of the RcGTA "life" cycle. The discovery of several proteins involved in release and attachment should be a boon for evolutionary and functional studies of RcGTA, allowing us to identify the elusive RcGTA receptor and, by disrupting it and working with strains that produce non-functional GTA, assess evolutionary costs in the absence of benefits as well as test relative fitness of "cheaters". The characterization of two new phages has also increased the available information on α-proteobacterial phages a full 20%, as well as given us the tools to, hopefully, identify interactions and shared histories between phages and RcGTA.

6 Bibliography

Ackermann, H.-W., (2007) 5500 phages examined in the electron microscope. *Arch Virol* **152**: 227-243.

Ackermann, H.-W., (2011) Bacteriophage taxonomy. *Viruses of Prokaryotes. Tom I.*

Adams, C. W., M. E. Forrest, S. N. Cohen & J. T. Beatty, (1989) Structural and functional analysis of transcriptional control of the *Rhodobacter capsulatus* puf operon. *J Bacteriol* **171**: 473-482.

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

Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller & D. J. Lipman, (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389-3402.

Anderson, L. M. & K. F. Bott, (1985) DNA packaging by the *Bacillus subtilis* defective bacteriophage PBSX. *J Virol* **54**: 773-780.

Arber, W., (1965) Host-Controlled Modification of Bacteriophage. *Annu Rev Microbiol* **19**: 365-378.

Avery, S. V., (2006) Microbial cell individuality and the underlying sources of heterogeneity. *Nat Rev Microbiol* **4**: 577-587.

Aziz, R., D. Bartels, A. Best, M. DeJongh, T. Disz, R. Edwards, K. Formsma, S. Gerdes, E. Glass, M. Kubal, F. Meyer, G. Olsen, R. Olson, A. Osterman, R. Overbeek, L. McNeil, D. Paarmann, T. Paczian, B. Parrello, G. Pusch, C. Reich, R. Stevens, O. Vassieva, V. Vonstein, A. Wilke & O. Zagnitko, (2008) The RAST Server: Rapid Annotations using Subsystems Technology. *BMC Genomics* **9**: 75.

Barany, F., (1985) Single-stranded hexameric linkers: a system for in-phase insertion mutagenesis and protein engineering. *Gene* **37**: 111-123.

Barrangou, R., C. Fremaux, H. Deveau, M. Richards, P. Boyaval, S. Moineau, D. A. Romero & P. Horvath, (2007) CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. *Science* **315**: 1709-1712.

Beatty, J. T. & H. Gest, (1981) Generation of succinyl-coenzyme A in photosynthetic bacteria. *Arch Microbiol* **129**: 335-340.

Benzer, S., (1955) Fine structures of a genetic recombination region in bacteriophage. *Proc Natl Acad Sci USA* **41**: 344-354.

Berglund, E. C., A. C. Frank, A. Calteau, O. Vinnere Pettersson, F. Granberg, A.-S. Eriksson, K. Näslund, M. Holmberg, H. Lindroos & S. G. E. Andersson, (2009) Run-off replication of host-adaptability genes is associated with gene transfer agents in the genome of mouse-infecting *Bartonella grahamii*. *PLoS Genet* **5**: e1000546.

Bertani, G., (1999) Transduction-like gene transfer in the methanogen *Methanococcus voltae*. *J Bacteriol* **181**: 2992-3002.

Bertani, G. & J. J. Weigle, (1953) Host controlled variation in bacterial viruses. *J Bacteriol* **65**: 113-121.

Besemer, J. & M. Borodovsky, (2005) GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses. *Nucleic Acids Res* **33**: W451-W454.
Bhattacharyya, S. P. & V. B. Rao, (1994) Structural analysis of DNA cleaved in vivo by bacteriophage T4 terminase. *Gene* **146**: 67-72.
Biers, E. J., K. Wang, C. Pennington, R. Belas, F. Chen & M. A. Moran, (2008) Occurrence and expression of gene transfer agent genes in marine bacterioplankton. *Appl Environ Microbiol* **74**: 2933-2939.
Black, L. W., (1989) DNA packaging in dsDNA bacteriophages. *Annu Rev Microbiol* **43**: 267-292.
Bolotin, A., B. Quinquis, A. Sorokin & S. D. Ehrlich, (2005) Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology-UK* **151**: 2551-2561.
Boyd, E. F. & H. Brüssow, (2002) Common themes among bacteriophage-encoded virulence factors and diversity among the bacteriophages involved. *Trends Microbiol* **10**: 521-529.
Brimacombe, C. A., A. Stevens, D. Jun, R. Mercer, A. S. Lang & J. T. Beatty, (2013) Quorum-sensing regulation of a capsular polysaccharide receptor for the *Rhodobacter capsulatus* gene transfer agent (RcGTA). *Mol Microbiol* **87**: 802-817.
Brüssow, H., C. Canchaya & W.-D. Hardt, (2004) Phages and the evolution of bacterial pathogens: from genomic rearrangements tolysogenic conversion. *Microbiol Mol Biol R* **68**: 560-602.
Brüssow, H. & R. W. Hendrix, (2002) Phage genomics: small is beautiful. *Cell* **108**: 13-16.
Brüssow, H. & E. Kutter, (2005) Phage ecology. In: Bacteriophages: Biology and Application. E. Kutter & A. Sulakvelidze (eds). Boca Raton, Florida: CRC Press, pp. 129-164.
Calendar, R., E. Ljungquist, G. Deho, D. C. Usher, R. Goldstein, P. Youderian, G. Sironi & E. W. Six, (1981) Lysogenization by satellite phage P4. *Virology* **113**: 20-38.
Campbell, A., (1961) Sensitive mutants of bacteriophage λ. *Virology* **14**: 22-32.
Casjens, S., (2003) Prophages and bacterial genomics: what have we learned so far? *Mol Microbiol* **49**: 277-300.
Casjens, S., G. Hatfull & R. Hendrix, (1992) Evolution of dsDNA tailed-bacteriophage genomes. *Sem Virol* **3**: 383-397.
Casjens, S. & J. King, (1975) Virus assembly. *Annu Rev Biochem* **44**: 555-611.
Casjens, S. R. & E. B. Gilcrease, (2009) Determining DNA packaging strategy by analysis of the termini of the chromosomes in tailed-bacteriophage virions. In: Bacteriophages. pp. 91-111.
Casjens, S. R., E. B. Gilcrease, D. A. Winn-Stapley, P. Schicklmaier, H. Schmieger, M. L. Pedulla, M. E. Ford, J. M. Houtz, G. F. Hatfull & R. W. Hendrix, (2005) The generalized transducing *Salmonella* bacteriophage ES18: Complete genome sequence and DNA packaging strategy. *J Bacteriol* **187**: 1091-1104.
Chen, F., A. Spano, B. E. Goodman, K. R. Blasier, A. Sabat, E. Jeffery, A. Norris, J. Shabanowitz, D. F. Hunt & N. Lebedev, (2008) Proteomic analysis and

identification of the structural and regulatory proteins of the *Rhodobacter capsulatus* gene transfer agent. *J Proteome Res* **8**: 967-973.
Chen, I. & D. Dubnau, (2004) DNA uptake during bacterial transformation. *Nat Rev Microbiol* **2**: 241-249.
Chen, Y., I. Golding, S. Sawai, L. Guo & E. C. Cox, (2005) Population fitness and the regulation of *Escherichia coli* genes by bacterial viruses. *PLoS Biol* **3**: e229.
Chibani-Chennoufi, S., A. Bruttin, M.-L. Dillmann & H. Brüssow, (2004) Phage-host interaction: an ecological perspective. *J Bacteriol* **186**: 3677-3686.
Chiura, H. X., K. Kogure, S. Hagemann, A. Ellinger & B. Velimirov, (2011) Evidence for particle-induced horizontal gene transfer and serial transduction between bacteria. *FEMS Microbiol Ecol* **76**: 576-591.
Christie, G. E. & T. Dokland, (2012) Pirates of the Caudovirales. *Virology* **434**: 210-221.
Clarke, B. R., F. Esumeh & I. S. Roberts, (2000) Cloning, expression, and purification of the K5 capsular polysaccharide lyase (KflA) from coliphage K5A: evidence for two distinct K5 lyase enzymes. *J Bacteriol* **182**: 3761-3766.
Claverys, J.-P., B. Martin & P. Polard, (2009) The genetic transformation machinery: composition, localization, and mechanism. *FEMS Microbiol Rev* **33**: 643-656.
Crick, F., L. Barnett, S. Brenner & R. Watts-Tobin, (1961) General nature of the genetic code for proteins. *Nature* **192**: 1227-1232.
d'Hérelle, F., (1917) Sur un microbe invisible antagoniste des bacilles dysentériques. *Comptes Rendus de l'Académie des Sciences, Paris*.
d'Hérelle, F., (1949) The Bacteriophage. *Sci News* **14**: 44-59.
Dagan, T., Y. Artzy-Randrup & W. Martin, (2008) Modular networks and cumulative impact of lateral transfer in prokaryote genome evolution. *Proc Natl Acad Sci USA* **105**: 10039-10044.
Datsenko, K. A., K. Pougach, A. Tikhonov, B. L. Wanner, K. Severinov & E. Semenova, (2012) Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. *Nat Commun* **3**: 945.
Daw, M. A. & F. R. Falkiner, (1996) Bacteriocins: nature, function and structure. *Micron* **27**: 467-479.
Del Campillo-Campbell, A., G. Kayajanian, A. Campbell & S. Adhya, (1967) Biotin-requiring mutants of *Escherichia coli* K-12. *J Bacteriol* **94**: 2065-2066.
Delcher, A. L., K. A. Bratke, E. C. Powers & S. L. Salzberg, (2007) Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* **23**: 673-679.
Doolittle, W. F. & E. Bapteste, (2007) Pattern pluralism and the tree of life hypothesis. *Proc Natl Acad Sci USA* **104**: 2043-2049.
Drummond, A., B. Ashton, S. Buxton, M. Cheung, A. Cooper, C. Duran, M. Field, J. Heled, M. Kearse, S. Markowitz, R. Moir, S. Stones-Havas, S. Sturrock, T. Thierer & A. Wilson, (2011) Geneious v5.4.
Dubey, G. P. & S. Ben-Yehuda, (2011) Intercellular nanotubes mediate bacterial communication. *Cell* **144**: 590-600.
Duckworth, D. H., (1976) "Who discovered bacteriophage?". *Bacteriol Rev* **40**: 793-802.
Ebel-Tsipis, J. & D. Botstein, (1971) Superinfection exclusion by P22 prophage in lysogens of *Salmonella typhimurium*: I. Exclusion of generalized transducing particles. *Virology* **45**: 629-637.

Ebel-Tsipis, J., D. Botstein & M. S. Fox, (1972) Generalized transduction by phage P22 in *Salmonella typhimurium*: I. Molecular origin of transducing DNA. *J Mol Biol* **71**: 433-448.

Edlin, G., L. E. O. Lin & R. Kudrna, (1975) λ Lysogens of *E. coli* reproduce more rapidly than non-lysogens. *Nature* **255**: 735-737.

Fau, M. H. & M. K. Cowman, (1986) Combined alcian blue and silver staining of glycosaminoglycans in polyacrylamide gels: application to electrophoretic analysis of molecular weight distribution. *Anal Biochem* **155**: 275-285.

Fiers, W., R. Contreras, F. Duerinck, G. Haegema, D. Iserentant, J. Merregaert, W. Min Jou, F. Molemans, A. Raeymaekers, A. Van Den Berghe, G. Volckaert & M. Ysebaert, (1976) Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene. *Nature* **260**: 500-507.

Fineran, P. C., T. R. Blower, I. J. Foulds, D. P. Humphreys, K. S. Lilley & G. P. C. Salmond, (2009) The phage abortive infection system, ToxIN, functions as a protein-RNA toxin-antitoxin pair. *Proc Natl Acad Sci USA* **106**: 894-899.

Fineran, P. C. & E. Charpentier, (2012) Memory of viral infections by CRISPR-Cas adaptive immune systems: Acquisition of new information. *Virology* **434**: 202-209.

Finkel, S. E. & R. Kolter, (2001) DNA as a nutrient: Novel role for bacterial competence gene homologs. *J Bacteriol* **183**: 6288-6293.

Florizone, S. M., (2006) Studies in the regulation of the gene transfer agent (GTA). In: Biology. Vancouver: University of British Columbia.

Fogg, P. C. M., A. P. Hynes, E. Digby, A. S. Lang & J. T. Beatty, (2011) Characterization of a newly discovered Mu-like bacteriophage, RcapMu, in *Rhodobacter capsulatus* strain SB1003. *Virology* **421**: 211-221.

Fogg, P. C. M., A. B. Westbye & J. T. Beatty, (2012) One for all or all for one: heterogeneous expression and host cell lysis are key to gene transfer agent activity in *Rhodobacter capsulatus*. *PLoS ONE* **7**: e43772.

Fonstein, M., S. Zheng & R. Haselkorn, (1992) Physical map of the genome of *Rhodobacter capsulatus* SB 1003. *J Bacteriol* **174**: 4070-4077.

Forde, A. & G. Fitzgerald, (1999) Bacteriophage defence systems in lactic acid bacteria. *Antonie van Leeuwenhoek* **76**: 89-113.

Fuqua, W. C., S. C. Winans & E. P. Greenberg, (1994) Quorum sensing in bacteria: the LuxR-LuxI family of cell density-responsive transcriptional regulators. *J Bacteriol* **176**: 269-275.

Ge, F., L.-S. Wang & J. Kim, (2005) The cobweb of life revealed bygenome-scale estimates of horizontal gene transfer. *PLoS Biol* **3**: e316.

Genthner, F. J. & J. D. Wall, (1984) Isolation of a recombination-deficient mutant of *Rhodopseudomonas capsulata*. *J Bacteriol* **160**: 971-975.

Georgopoulos, C. P., R. W. Hendrix, S. R. Casjens & A. D. Kaiser, (1973) Host participation in bacteriophage lambda head assembly. *J Mol Biol* **76**: 45-50.

Gill, P. R., Jr. & G. J. Warren, (1988) An iron-antagonized fungistatic agent that is not required for iron assimilation from a fluorescent rhizosphere pseudomonad. *J Bacteriol* **170**: 163-170.

Goldstein, R., J. Sedivy & E. Ljungquist, (1982) Propagation of satellite phage P4 as a plasmid. *Proc Natl Acad Sci USA* **79**: 515-519.
Grissa, I., G. Vergnaud & C. Pourcel, (2007) CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res* **35**: W52-W57.
Haggård-Ljungquist, E., C. Halling & R. Calendar, (1992) DNA sequences of the tail fiber genes of bacteriophage P2: evidence for horizontal transfer of tail fiber genes among unrelated bacteriophages. *J Bacteriol* **174**: 1462-1477.
Haselkorn, R., A. Lapidus, Y. Kogan, C. Vlcek, J. Paces, V. Paces, P. Ulbrich, T. Pecenkova, D. Rebrekov, A. Milgram, M. Mazur, R. Cox, N. Kyrpides, N. Ivanova, V. Kapatral, T. Los, A. Lykidis, N. Mikhailova, G. Reznik, O. Vasieva & M. Fonstein, (2001) The *Rhodobacter capsulatus* genome. *Photosynth Res* **70**: 43-52.
Hershey, A. D., (1946) Mutation of bacteriophage with respect to type of plaque. *Genetics* **31**: 620-640.
Hershey, A. D. & M. Chase, (1952) Independent functions of viral protein and nucleic acid in growth of bacteriophage. *J Gen Physiol* **36**: 39-56.
Hershey, A. D., J. Dixon & M. Chase, (1953) Nucleic acid economy in bacteria infected with Bacteriophage T2. *J. Gen. Physiol* **36**: 777-789.
Housby, J. N. & N. H. Mann, (2009) Phage therapy. *Drug Discovery Today* **14**: 536-540.
Howe, M. M., (1973) Transduction by bacteriophage MU-1. *Virology* **55**: 103-117.
Huang, S., Y. Zhang, F. Chen & N. Jiao, (2011) Complete genome sequence of a marine roseophage provides evidence into the evolution of gene transfer agents in alphaproteobacteria. *Virology J* **8**: 124.
Humphrey, S. B., T. B. Stanton, N. S. Jensen & R. L. Zuerner, (1997) Purification and characterization of VSH-1, a generalized transducing bacteriophage of *Serpulina hyodysenteriae*. *J Bacteriol* **179**: 323-329.
Hynes, A. P., R. G. Mercer, D. E. Watton, C. B. Buckley & A. S. Lang, (2012) DNA packaging bias and differential expression of gene transfer agent genes within a population during production and release of the *Rhodobacter capsulatus* gene transfer agent, RcGTA. *Mol Microbiol* **85**: 314-325.
Ikeda, H. & J. I. Tomizawa, (1965) Transducing fragments in generalized transduction by phage P1. I. Molecular origin of the fragments. *J Mol Biol* **14**: 85-109.
Imhoff, J. & U. Bias-lmhoff, (2004) Lipids, quinones and fatty acids of anoxygenic phototrophic bacteria. In: Anoxygenic Photosynthetic Bacteria. R. E. Blankenship, M. T. Madigan & C. E. Bauer (eds). Springer Netherlands, pp. 179-205.
Imhoff, J. F., H. G. Trüper & N. Pfennig, (1984) Rearrangement of the species and genera of the phototrophic "purple nonsulfur bacteria". *Int J Syst Bacteriol* **34**: 340-343.
Irizarry, R. A., B. M. Bolstad, F. Collin, L. M. Cope, B. Hobbs & T. P. Speed, (2003) Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res* **31**: e15-.
Jabrane, A., A. Sabri, P. Compère, P. Jacques, I. Vandenberghe, J. Van Beeumen & P. Thonart, (2002) Characterization of serracin P, a phage-tail-like bacteriocin, and its activity against *Erwinia amylovora*, the fire blight pathogen. *Appl Environ Microbiol* **68**: 5704-5710.

Jacob, F. & J. Monod, (1961) Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol* **3**: 318-356.

Jansen, R., J. D. A. v. Embden, W. Gaastra & L. M. Schouls, (2002) Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol* **43**: 1565-1575.

Johnsborg, O., V. Eldholm & L. S. Håvarstein, (2007) Natural genetic transformation: prevalence, mechanisms and function. *Res Microbiol* **158**: 767-778.

Juhala, R. J., M. E. Ford, R. L. Duda, A. Youlton, G. F. Hatfull & R. W. Hendrix, (2000) Genomic sequences of bacteriophages HK97 and HK022: pervasive genetic mosaicism in the lambdoid bacteriophages. *J Mol Biol* **299**: 27-51.

Kristensen, D. M., A. R. Mushegian, V. V. Dolja & E. V. Koonin, (2010) New dimensions of the virus world discovered through metagenomics. *Trends Microbiol* **18**: 11-19.

Kucera, R. B. & N. M. Nichols, (2008) DNA-dependent DNA polymerases. *Curr Protoc Mol Biol*: 3-5.

Kurland, C. G., B. Canback & O. G. Berg, (2003) Horizontal gene transfer: a critical view. *Proc Natl Acad Sci USA* **100**: 9658-9662.

Kutter, E. M. & J. S. Wiberg, (1968) Degradation of cytosine-containing bacterial and bacteriophage DNA after infection of *Escherichia coli* B with bacteriophage T4D wild type and with mutants defective in genes 46, 47 and 56. *J Mol Biol* **38**: 395-411.

Labrie, S. J., J. E. Samson & S. Moineau, (2010) Bacteriophage resistance mechanisms. *Nat Rev Microbiol* **8**: 317-327.

Lang, A. S. & J. T. Beatty, (2000) Genetic analysis of a bacterial genetic exchange element: the gene transfer agent of *Rhodobacter capsulatus*. *Proc Natl Acad Sci USA* **97**: 859-864.

Lang, A. S. & J. T. Beatty, (2001) The gene transfer agent of *Rhodobacter capsulatus* and "constitutive transduction" in prokaryotes. *Arch Microbiol* **175**: 241-249.

Lang, A. S. & J. T. Beatty, (2007) Importance of widespread gene transfer agent genes in α-proteobacteria. *Trends Microbiol* **15**: 54-62.

Lang, A. S., T. A. Taylor & J. T. Beatty, (2002) Evolutionary implications of phylogenetic analyses of the gene transfer agent (GTA) of *Rhodobacter capsulatus*. *J Mol Evol* **55**: 534-543.

Lang, A. S., O. Zhaxybayeva & J. T. Beatty, (2012) Gene transfer agents: phage-like elements of genetic exchange. *Nat Rev Micro* **10**: 472-482.

Larkin, M. A., G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson & D. G. Higgins, (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* **23**: 2947-2948.

Leung, M. M., C. A. Brimacombe, G. B. Spiegelman & J. T. Beatty, (2012) The GtaR protein negatively regulates transcription of the gtaRI operon and modulates gene transfer agent (RcGTA) expression in *Rhodobacter capsulatus*. *Mol Microbiol* **83**: 759-774.

Lima-Mendez, G., A. Toussaint & R. l. Leplae, (2007) Analysis of the phage sequence space: The benefit of structured information. *Virology* **365**: 241-249.

Lindqvist, B. H., G. Dehò & R. Calendar, (1993) Mechanisms of genome propagation and helper exploitation by satellite phage P4. *Microbiol Rev* **57**: 683-702.
Lindsay, J. A., A. Ruzin, H. F. Ross, N. Kurepina & R. P. Novick, (1998) The gene for toxic shock toxin is carried by a family of mobile pathogenicity islands in *Staphylococcus aureus*. *Mol Microbiol* **29**: 527-543.
Longchamp, P. F., C. Mauël & D. Karamata, (1994) Lytic enzymes associated with defective prophages of *Bacillus subtilis*: Sequencing and characterization of the region comprising the N-acetylmuramoyl-L-alanine amidase gene of prophage PBSX. *Microbiology-UK* **140**: 1855-1867.
Lopez, D., H. Vlamakis & R. Kolter, (2009) Generation of multiple cell types in *Bacillus subtilis*. *FEMS Microbiol Rev* **33**: 152-163.
Louie, D. & P. Serwer, (1990) Blunt-ended ligation can be used to produce DNA ladders with rung spacing as large as 0.17 Mb. *Nucleic Acids Res* **18**: 3090.
Luria, S. E. & M. Delbrück, (1943) Mutations of bacteria from virus sensitivity to virus resistance. *Genetics* **28**: 491–511.
Luria, S. E. & M. L. Human, (1952) A nonhereditary, host-induced variation of bacterial viruses. *J Bacteriol* **64**: 557-569.
Lwoff, A., (1953) Lysogeny. *Bacteriol Rev* **17**: 269–337.
Madigan, M. T., (1995) Microbiology of nitrogen fixation by anoxygenic photosynthetic bacteria. In: Anoxygenic photosynthetic bacteria. C. E. Bauer (ed). Boston: Kluwer Academic Publishers, pp. 915-928.
Makarova, K. S., D. H. Haft, R. Barrangou, S. J. J. Brouns, E. Charpentier, P. Horvath, S. Moineau, F. J. M. Mojica, Y. I. Wolf, A. F. Yakunin, J. van der Oost & E. V. Koonin, (2011) Evolution and classification of the CRISPR-Cas systems. *Nat Rev Micro* **9**: 467-477.
Marchler-Bauer, A. & S. H. Bryant, (2004) CD-Search: protein domain annotations on the fly. *Nucleic Acids Res* **32**: W327-W331.
Margulies, M., M. Egholm, W. E. Altman, S. Attiya, J. S. Bader, L. A. Bemben, J. Berka, M. S. Braverman, Y.-J. Chen, Z. Chen, S. B. Dewell, L. Du, J. M. Fierro, X. V. Gomes, B. C. Godwin, W. He, S. Helgesen, C. H. Ho, G. P. Irzyk, S. C. Jando, M. L. I. Alenquer, T. P. Jarvie, K. B. Jirage, J.-B. Kim, J. R. Knight, J. R. Lanza, J. H. Leamon, S. M. Lefkowitz, M. Lei, J. Li, K. L. Lohman, H. Lu, V. B. Makhijani, K. E. McDade, M. P. McKenna, E. W. Myers, E. Nickerson, J. R. Nobile, R. Plant, B. P. Puc, M. T. Ronan, G. T. Roth, G. J. Sarkis, J. F. Simons, J. W. Simpson, M. Srinivasan, K. R. Tartaro, A. Tomasz, K. A. Vogt, G. A. Volkmer, S. H. Wang, Y. Wang, M. P. Weiner, P. Yu, R. F. Begley & J. M. Rothberg, (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**: 376-380.
Marraffini, L. A. & E. J. Sontheimer, (2008) CRISPR interference limits horizontal gene transfer in *Staphylococci* by targeting DNA. *Science* **322**: 1843-1845.
Marrs, B. L., (1974) Genetic recombination in *Rhodopseudomonas capsulata*. *Proc Natl Acad Sci USA* **71**: 971-973.
Marvin, D. A. & B. Hohn, (1969) Filamentous bacterial viruses. *Bacteriol Rev* **33**: 172.
Mashburn-Warren, L. M. & M. Whiteley, (2006) Special delivery: vesicle trafficking in prokaryotes. *Mol Microbiol* **61**: 839-846.

Matson, E. G., M. G. Thompson, S. B. Humphrey, R. L. Zuerner & T. B. Stanton, (2005) Identification of genes of VSH-1, a prophage-like gene transfer agent of *Brachyspira hyodysenteriae*. *J Bacteriol* **187**: 5885-5892.

McDonnell, G. E., H. Wood, K. M. Devine & D. J. McConnell, (1994) Genetic control of bacterial suicide: regulation of the induction of PBSX in *Bacillus subtilis*. *J Bacteriol* **176**: 5820-5830.

Mercer, R. G., S. J. Callister, M. S. Lipton, L. Pasa-Tolic, H. Strnad, V. Paces, J. T. Beatty & A. S. Lang, (2010) Loss of the response regulator CtrA causes pleiotropic effects on gene expression but does not affect growth phase regulation in *Rhodobacter capsulatus*. *J Bacteriol* **192**: 2701-2710.

Mercer, R. G., M. Quinlan, A. R. Rose, S. Noll, J. T. Beatty & A. S. Lang, (2012) Regulatory systems controlling motility and gene transfer agent production and release in *Rhodobacter capsulatus*. *FEMS Microbiol Lett* **331**: 53-62.

Meselson, M. & J. J. Weigle, (1961) Chromosome breakage accompanying genetic recombination in bacteriophage. *Proc Natl Acad Sci USA* **47**: 857–868.

Miller, J. H., (1992) *A short course in bacterial genetics: a laboratory manual and handbook for Escherichia coli and related bacteria,* p. 456. Cold Spring Harbor Laboratory Press, Plainview, NY.

Mizuuchi, K. & R. Craigie, (1986) Mechanism of bacteriophage mu transposition. *Annu Rev Genet* **20**: 385-429.

Moineau, S. & C. Lévesque, (2005) Control of bacteriophages in industrial ferments. In: Bacteriophages: Biology and Application. E. Kutter & A. Sulakvelidze (eds). Boca Raton, Florida: CRC Press, pp. 285-296.

Mojica, F. J. M., C. Díez-Villaseñor, J. García-Martínez & C. Almendros, (2009) Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology-UK* **155**: 733-740.

Mojica, F. M., C. Díez-Villaseñor, J. García-Martínez & E. Soria, (2005) Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol* **60**: 174-182.

Moon, S., Y. Byun, H.-J. Kim, S. Jeong & K. Han, (2004) Predicting genes expressed via -1 and +1 frameshifts. *Nucleic Acids Res* **32**: 4884-4892.

Morgan, G. J., G. F. Hatfull, S. Casjens & R. W. Hendrix, (2002) Bacteriophage Mu genome sequence: analysis and comparison with Mu-like prophages in *Haemophilus*, *Neisseria* and *Deinococcus*. *J Mol Biol* **317**: 337-359.

Morrison, D. A. & M. S. Lee, (2000) Regulation of competence for genetic transformation in *Streptococcus pneumoniae*: a link between quorum sensing and DNA processing genes. *Res Microbiol* **151**: 445-451.

Morse, M. L., E. M. Lederberg & J. Lederberg, (1956) Transduction in *Escherichia coli* K-12. *Genetics* **41**: 142.

Moyer, A. J., (1948) Method for production of Penicillin. S. o. Agriculture (ed). U.S.A.

Nagaraja Rao, R., (1968) Bacteriophage P22 controlled exclusion in *Salmonella typhimurium*. *J Mol Biol* **35**: 607-622.

Nakamura, Y., T. Itoh, H. Matsuda & T. Gojobori, (2004) Biased biological functions of horizontally transferred genes in prokaryotic genomes. *Nat Genet* **36**: 760-766.

Nealson, K. H. & J. W. Hastings, (1979) Bacterial bioluminescence: its control and ecological significance. *Microbiol Rev* **43**: 496.
Ochman, H., J. G. Lawrence & E. A. Groisman, (2000) Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**: 299-304.
Okamoto, K., J. A. Mudd, J. Mangan, W. M. Huang, T. V. Subbaiah & J. Marmur, (1968) Properties of the defective phage of *Bacillus subtilis*. *J Mol Biol* **34**: 413-428.
Okazaki, R., T. Okazaki, K. Sakabe, K. Sugimoto & A. Sugino, (1968) Mechanism of DNA chain growth, I. Possible discontinuity and unusual secondary structure of newly synthesized chains. *Proc Natl Acad Sci USA* **59**: 598-605.
Olendzenski, L. & J. P. Gogarten, (2009) Evolution of genes and organisms. *Ann NY Acad Sci* **1178**: 137-145.
Paul, J. H., (2008) Prophages in marine bacteria: dangerous molecular time bombs or the key to survival in the seas? *ISME J* **2**: 579-589.
Pemberton, J. M., I. M. Horne & A. G. McEwan, (1998) Regulation of photosynthetic gene expression in purple bacteria. *Microbiology-UK* **144**: 267-278.
Pepper, S., E. Saunders, L. Edwards, C. Wilson & C. Miller, (2007) The utility of MAS5 expression summary and detection call algorithms. *BMC Bioinformatics* **8**: 273.
Pourcel, C., G. Salvignol & G. Vergnaud, (2005) CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology-UK* **151**: 653-663.
Prentki, P. & H. M. Krisch, (1984) In vitro insertional mutagenesis with a selectable DNA fragment. *Gene* **29**: 303-313.
Ptashne, M., (1967) Specific binding of the λ phage repressor to λ DNA. *Nature* **214**: 232-234.
Ptashne, M., (2004) A genetic switch : Phage lambda revisited.
Puigbò, P., Y. I. Wolf & E. V. Koonin, (2010) The tree and net components of prokaryote evolution. *Genome biology and evolution* **2**: 745.
Ram, G., J. Chen, K. Kumar, H. F. Ross, C. Ubeda, P. K. Damle, K. D. Lane, J. R. Penadés, G. E. Christie & R. P. Novick, (2012) Staphylococcal pathogenicity island interference with helper phage reproduction is a paradigm of molecular parasitism. *Proc Natl Acad Sci USA* **109**: 16300-16305.
Rao, V. B. & L. W. Black, (2005) DNA packaging in bacteriophage T4. In: Viral genome packaging machines: genetics, structure, and mechanism. Springer, pp. 40-58.
Rapp, B. J. & J. D. Wall, (1987) Genetic transfer in *Desulfovibrio desulfuricans*. *Proc Natl Acad Sci USA* **84**: 9128-9130.
Redfield, R., (2001) Do bacteria have sex? *Nat Rev Genet* **2**: 634-639.
Redfield, R. J., (1988) Evolution of bacterial transformation: is sex with dead cells ever better than no sex at all? *Genetics* **119**: 213-221.
Redfield, R. J., (2002) Is quorum sensing a side effect of diffusion sensing? *Trends Microbiol* **10**: 365-370.
Redfield, R. J., M. R. Schrag & A. M. Dean, (1997) The evolution of bacterial transformation: Sex with poor relations. *Genetics* **146**: 27-38.
Riley, M. & M. Lizotte-Waniewski, (2009) Population genomics and the bacterial species concept. In: Horizontal Gene Transfer. M. Gogarten, J. Gogarten & L. Olendzenski (eds). Humana Press, pp. 367-377.

Rohwer, F., (2003) Global Phage Diversity. *Cell* **113**: 141.
Rosenshine, I., R. Tchelet & M. Mevarech, (1989) The mechanism of DNA transfer in the mating system of an archaebacterium. *Science* **245**: 1387-1389.
Ruska, H., (1940) Die sichtbarmachung der Bakteriophagen lyse im Übermikroskop. *Naturwissenschaften* **28**: 45-46.
Sanger, F., A. R. Coulson, T. Friedmann, G. M. Air, B. G. Barrell, N. L. Brown, J. C. Fiddes, C. A. Hutchison, P. M. Slocombe & M. Smith, (1978) The nucleotide sequence of bacteriophage φX174. *J Mol Biol* **125**: 225-246.
Sarabhai, A. S., A. O. W. Stretton, S. Brenner & A. Bolle, (1964) Co-linearity of the gene with the polypeptide chain. *Nature* **201**: 13-17.
Schaefer, A. L., T. A. Taylor, J. T. Beatty & E. P. Greenberg, (2002) Long-chain acyl-homoserine lactone quorum-sensing regulation of *Rhodobacter capsulatus* Gene Transfer Agent production. *J Bacteriol* **184**: 6515-6521.
Schmidt, L. S., H.-C. Yen & H. Gest, (1974) Bioenergetic aspects of bacteriophage replication in the photosynthetic bacterium *Rhodopseudomonas capsulata*. *Arch Biochem Biophys* **165**: 229-239.
Schmieger, H., K. M. Taleghani, A. Meierl & L. Weiß, (1990) A molecular analysis of terminase cuts in headful packaging of Salmonella phage P22. *Mol Gen Genet* **221**: 199-202.
Schwinghamer, E. A., C. E. Pankhurst & P. R. Whitfeld, (1973) A phage-like bacteriocin of *Rhizobium trifolii*. *Can J Microbiol* **19**: 359-368.
Scolnik, P. A. & R. Haselkorn, (1984) Activation of extra copies of genes coding for nitrogenase in *Rhodopseudomonas capsulata*. *Nature* **307**: 289-292.
Seaman, E., E. Tarmy & J. Marmur, (1964) Inducible Phages of *Bacillus subtilis**. *Biochemistry-US* **3**: 607-613.
Seed, K. D., D. W. Lazinski, S. B. Calderwood & A. Camilli, (2013) A bacteriophage encodes its own CRISPR/Cas adaptive response to evade host innate immunity. *Nature* **494**: 489-491.
Shimada, K., R. A. Weisberg & M. E. Gottesman, (1972) Prophage lambda at unusual chromosomal locations: I. Location of the secondary attachment sites and the properties of the lysogens. *J Mol Biol* **63**: 483-503.
Shore, D., G. Dehò, J. Tsipis & R. Goldstein, (1978) Determination of capsid size by satellite bacteriophage P4. *Proc Natl Acad Sci USA* **75**: 400-404.
Simon, R., U. Priefer & A. Pühler, (1983) A broad host range mobilization system for in vivo genetic engineering: transposon mutagenesis in Gram-negative bacteria. *Bio/Tech* **1**: 37-45.
Six, E. W., (1963) A defective phage depending on phage P2.
Six, E. W. & C. A. C. Klug, (1973) Bacteriophage P4: a satellite virus depending on a helper such as prophage P2. *Virology* **51**: 327-344.
Smith, H. O., C. A. Hutchison, C. Pfannkoch & J. C. Venter, (2003) Generating a synthetic genome by whole genome assembly: φX174 bacteriophage from synthetic oligonucleotides. *Proc Natl Acad Sci USA* **100**: 15440-15445.
Solioz, M., (1975) The gene transfer agent of *Rhodopseudomonas capsulata*. Saint Louis, MO: Saint Louis University.

Solioz, M. & B. Marrs, (1977) The gene transfer agent of *Rhodopseudomonas capsulata*: Purification and characterization of its nucleic acid. *Arch Biochem Biophys* **181**: 300-307.
Solioz, M., H.-C. Yen & B. Marrs, (1975) Release and uptake of gene transfer agent by *Rhodopseudomonas capsulata*. *J Bacteriol* **123**: 651-657.
Stanton, T. B., (2007) Prophage-like gene transfer agents--Novel mechanisms of gene exchange for *Methanococcus*, *Desulfovibrio*, *Brachyspira*, and *Rhodobacter* species. *Anaerobe* **13**: 43-49.
Stanton, T. B., S. B. Humphrey, D. O. Bayles & R. L. Zuerner, (2009) Identification of a divided genome for VSH-1, the prophage-like gene transfer agent of *Brachyspira hyodysenteriae*. *J. Bacteriol.* **191**: 1719-1721.
Stent, G., (1992) Max Delbrück. In: Phage and the origins of molecular biology (expanded edition). John Cairns, Gunther S. Stent & J. D. Watson. (eds). Cold Spring Harbor: Cold Spring Harbor Laboratory Press, pp. 351-366.
Strauch, E., H. Kaspar, C. Schaudinn, P. Dersch, K. Madela, C. Gewinner, S. Hertwig, J. r. Wecke & B. Appel, (2001) Characterization of enterocoliticin, a phage tail-like bacteriocin, and its effect on pathogenic *Yersinia enterocolitica* strains. *Appl Environ Microbiol* **67**: 5634-5642.
Streisinger, G., J. Emrich & M. M. Stahl, (1967) Chromosome structure in phage T4, III. Terminal redundancy and length determination. *Proc Natl Acad Sci USA* **57**: 292-295.
Strnad, H., A. Lapidus, J. Paces, P. Ulbrich, C. Vlcek, V. Paces & R. Haselkorn, (2010) Complete genome sequence of the photosynthetic purple nonsulfur bacterium *Rhodobacter capsulatus SB 1003*. *J Bacteriol* **192**: 3545-3546.
Summers, W. C., (2001) Bacteriophage therapy. *Annu Rev Microbiol* **55**: 437-451.
Sun, Q., G. F. Kuty, A. Arockiasamy, M. Xu, R. Young & J. C. Sacchettini, (2009) Regulation of a muralytic enzyme by dynamic membrane topology. *Nat Struct Mol Biol* **16**: 1192-1194.
Susskind, M. M. & D. Botstein, (1980) Superinfection exclusion by λ prophage in lysogens of *Salmonella typhimurium*. *Virology* **100**: 212-216.
Susskind, M. M., D. Botstein & A. Wright, (1974) Superinfection exclusion by P22 prophage in lysogens of *Salmonella typhimurium*: III. Failure of superinfecting phage DNA to enter sieA+ lysogens. *Virology* **62**: 350-366.
Susskind, M. M., A. Wright & D. Botstein, (1971) Superinfection exclusion by P22 prophage in lysogens of *Salmonella typhimurium*: II. Genetic evidence for two exclusion systems. *Virology* **45**: 638-652.
Suttle, C. A., (1994) The significance of viruses to mortality in aquatic microbial communities. *Microb Ecol* **28**: 237-243.
Swarts, D. C., C. Mosterd, M. W. J. van Passel & S. J. J. Brouns, (2012) CRISPR interference directs strand specific spacer acquisition. *PLoS ONE* **7**: e35888.
Tatum, E. L. & J. Lederberg, (1947) Gene recombination in the bacterium *Escherichia coli*. *J Bacteriol* **53**: 673-684.
Taylor, D. P., S. N. Cohen, W. G. Clark & B. L. Marrs, (1983) Alignment of the genetic and restriction maps of the photosynthetic region of the *Rhodopseudomonas*

capsulata chromosome by a conjugation-mediated marker rescue technique. *J Bacteriol* **154**: 580-590.
Thaler, J. O., S. Baghdiguian & N. Boemare, (1995) Purification and characterization of xenorhabdicin, a phage tail-like bacteriocin, from the lysogenic strain F1 of *Xenorhabdus nematophilus*. *Appl Environ Microbiol* **61**: 2049-2052.
Twort, F., (1915) An investigation on the nature of ultra-microscopic viruses. *Lancet* **11**: 1241.
Úbeda, C., E. Maiques, P. Barry, A. Matthews, M. Á. Tormo, Í. Lasa, R. P. Novick & J. R. Penadés, (2008) SaPI mutations affecting replication and transfer and enabling autonomous replication in the absence of helper phage. *Mol Microbiol* **67**: 493-503.
Vale, P. F. & T. J. Little, (2010) CRISPR-mediated phage resistance and the ghost of coevolution past. *P Roy Soc B-Biol Sci*.
Van Helvoort, T., (1992) Bacteriological and physiological research styles in the early controversy on the nature of the bacteriophage phenomenon. *Med Hist* **36**: 243–270.
Vercoe, R. B., J. T. Chang, R. L. Dy, C. Taylor, T. Gristwood, J. S. Clulow, C. Richter, R. Przybilski, A. R. Pitman & P. C. Fineran, (2013) Cytotoxic chromosomal targeting by CRISPR/Cas systems can reshape bacterial genomes and expel or remodel pathogenicity islands. *PLoS Genet* **9**: e1003454.
Verma, I. M., (1975) Studies on reverse transcriptase of RNA tumor viruses III. Properties of purified Moloney murine leukemia virus DNA polymerase and associated RNase H. *J. Virol.* **15**: 843-854.
Volkin, E. & L. Astrachan, (1956) Phosphorus incorporation in *Escherichia coli* ribonucleic acid after infection with bacteriophage T2. *Virology* **2**: 149-161.
Vos, M., (2009) Why do bacteria engage in homologous recombination? *Trends Microbiol* **17**: 226-232.
Wall, J. D. & K. Braddock, (1984) Mapping of *Rhodopseudomonas capsulata nif* genes. *J Bacteriol* **158**: 404-410.
Wall, J. D., P. F. Weaver & H. Gest, (1975a) Gene transfer agents, bacteriophages, and bacteriocins of *Rhodopseudomonas capsulata*. *Arch Microbiol* **105**: 217-224.
Wall, J. D., P. F. Weaver & H. Gest, (1975b) Genetic transfer of nitrogenase-hydrogenase activity in *Rhodopseudomonas capsulata*. *Nature* **258**: 630-631.
Walsh, J. & G. G. Meynell, (1967) The isolation of non-excluding mutants of Phage p22. *J Gen Virol* **1**: 581-582.
Wang, X., Y. Kim, Q. Ma, S. H. Hong, K. Pokusaeva, J. M. Sturino & T. K. Wood, (2010) Cryptic prophages help bacteria cope with adverse environments. *Nat Commun* **1**: 147.
Weaver, P. F., J. D. Wall & H. Gest, (1975) Characterization of *Rhodopseudomonas capsulata*. *Arch Microbiol* **105**: 207-216.
Wheeler, D. L., T. Barrett, D. A. Benson, S. H. Bryant, K. Canese, V. Chetvernin, D. M. Church, M. DiCuccio, R. Edgar, S. Federhen, L. Y. Geer, Y. Kapustin, O. Khovayko, D. Landsman, D. J. Lipman, T. L. Madden, D. R. Maglott, J. Ostell, V. Miller, K. D. Pruitt, G. D. Schuler, E. Sequeira, S. T. Sherry, K. Sirotkin, A. Souvorov, G. Starchenko, R. L. Tatusov, T. A. Tatusova, L. Wagner & E.

Yaschenko, (2007) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* **35**: D5-12.
Wilhelm, S. W. & C. A. Suttle, (1999) Viruses and nutrient cycles in the sea. *BioScience* **49**: 781-788.
Wollman, E., (1938) Recherches sur le phénoméne de Twort-d'Hérelle (Bactériophagie ou autolyse hérédo-Contagieuse). *Ann. Inst. Pasteur Paris* **60**: 13-57.
Wong, D. K.-H., W. J. Collins, A. Harmer, T. G. Lilburn & J. T. Beatty, (1996) Directed mutagenesis of the *Rhodobacter capsulatus puhA* gene and pleiotropic effects on photosynthetic reaction center and light-harvesting I complexes. *J Bacteriol* **178**: 2334-2342.
Wood, H. E., M. T. Dawson, K. M. Devine & D. J. McConnell, (1990) Characterization of PBSX, a defective prophage of *Bacillus subtilis*. *J Bacteriol* **172**: 2667-2674.
Yen, H.-C., N. T. Hu & B. L. Marrs, (1979) Characterization of the gene transfer agent made by an overproducer mutant of *Rhodopseudomonas capsulata*. *J Mol Biol* **131**: 157-168.
Yen, H.-C. & B. Marrs, (1976) Map of genes for carotenoid and bacteriochlorophyll biosynthesis in *Rhodopseudomonas capsulata*. *J Bacteriol* **126**: 619-629.
Young, R. Y., (2002) Bacteriophage holins: deadly diversity. *J Mol Microb Biotech* **4**: 21-36.
Zhaxybayeva, O., J. P. Gogarten, R. L. Charlebois, W. F. Doolittle & R. T. Papke, (2006) Phylogenetic analyses of cyanobacterial genomes: Quantification of horizontal gene transfer events. *Genome Res* **16**: 1099-1108.
Zinder, N. D. & J. Lederberg, (1952) Genetic exchange in *Salmonella*. *J Bacteriol* **64**: 679-699.

Appendix 1 – Supplemental tables for Chapter 3

Table A-1: Transcript level signal intensity ratios for 26 genes identified as co-regulated with RcGTA production in four transcriptome comparisons.

Gene	SB1003/ALS1	SB1003/SBRM1	DE442/SB1003	SB1003/Midlog
rcc00171	2.77	8.84	53.86	6.37
rcc00555	2.36	2.98	35.00	2.79
rcc00556	2.47	4.87	43.19	4.92
rcc01079	5.72	8.99	18.52	7.68
rcc01080	3.36	10.89	11.87	5.93
rcc01682	2.18	3.68	12.83	5.05
rcc01683	2.67	3.47	13.26	6.80
rcc01684	6.54	20.91	27.74	8.16
rcc01685	6.77	38.70	17.88	11.12
rcc01686	3.76	15.06	16.06	4.28
rcc01687	6.09	25.03	10.84	6.31
rcc01688	6.28	13.76	18.11	5.34
rcc01689	11.09	4.98	16.47	5.45
rcc01690	8.82	10.92	21.90	5.82
rcc01691	4.09	29.86	18.33	9.98
rcc01692	4.58	9.86	15.37	5.33
rcc01693	2.53	5.51	8.94	4.59
rcc01694	5.37	15.42	18.91	6.32
rcc01695	5.82	16.98	15.66	10.47
rcc01696	4.77	15.70	22.13	7.15
rcc01697	3.72	3.76	31.64	3.10
rcc01698	2.10	2.39	78.49	2.25
rcc01865	2.22	3.09	18.70	3.15
rcc01866	2.92	3.12	51.34	2.69
rcc02623	4.66	15.88	30.98	5.64
rcc02730	2.18	2.79	5.79	3.93

Genes highlighted in grey are in the RcGTA structural cluster

Table A-2: Genes co-regulated with the RcGTA structural cluster in three of four comparisons.

Gene Name	Annotation
	Genes co-regulated with RcGTA cluster in all but ALS1
rcc00042	PAS/PAC sensor domain protein
rcc01058	gas vesicle synthesis protein GvpL/GvpF
rcc01061	conserved hypothetical protein
rcc02724	RNA polymerase sigma factor, sigma-70 family, ECF subfamily
rcc02725	conserved hypothetical protein
rcc02726	conserved hypothetical protein
rcc02727	conserved hypothetical protein
rcc02728	ATPase, AAA family
rcc02729	conserved hypothetical protein
rcc02731	conserved hypothetical protein
rcc02733	conserved hypothetical protein
rcc02734	exodeoxyribonuclease V
rcc03521	flagellar protein export ATPase FliI
	Genes co-regulated with RcGTA cluster in all but DE442
rcc01666	membrane protein, putative
rcc02241	taurine ABC transporter, periplasmic taurine-binding protein TauA
rcc02274	proline iminopeptidase
rcc03023	tripartite ATP-independent periplasmic transporter, DctQ component
rcc03098	DNA protecting protein DprA
*rcp00126**	conserved hypothetical protein

*This gene is not present in DE442

Appendix 2 – Other Scientific Contributions

Fogg, P. C. M., A. P. Hynes, E. Digby, A. S. Lang & J. T. Beatty, (2011) Characterization of a newly discovered Mu-like bacteriophage, RcapMu, in *Rhodobacter capsulatus* strain SB1003. *Virology* 421: 211-221

Hynes, A.P., & A.S. Lang, *Rhodobacter capsulatus* Gene Transfer Agent (RcGTA) Activity Bioassays, in bio-protocol.org

Virology 421 (2011) 211–221

Contents lists available at SciVerse ScienceDirect

Virology

journal homepage: www.elsevier.com/locate/yviro

Characterization of a newly discovered Mu-like bacteriophage, RcapMu, in *Rhodobacter capsulatus* strain SB1003

Paul C.M. Fogg [a], Alexander P. Hynes [b], Elizabeth Digby [a], Andrew S. Lang [b], J. Thomas Beatty [a,*]

[a] University of British Columbia, Department of Microbiology and Immunology, Vancouver, British Columbia, Canada
[b] Memorial University of Newfoundland, Department of Biology, St. John's, Newfoundland, Canada

ARTICLE INFO

Article history:
Received 20 July 2011
Returned to author for revision
25 August 2011
Accepted 28 September 2011
Available online 22 October 2011

Keywords:
Rhodobacter capsulatus
Bacteriophage Mu
Transposing phage
Induction
Siphoviridae

ABSTRACT

The α-proteobacterium *Rhodobacter capsulatus* is a model organism for the study of bacterial photosynthesis and the bacteriophage-like gene transfer agent. Characterization of phages that infect *Rhodobacter* is extremely rare, and scarce for the α-proteobacteria in general. Here, we describe the discovery of the only functional Mu-like transposing phage to have been identified in the α-proteobacteria, RcapMu, resident in the genome-sequenced *R. capsulatus* SB1003 strain. RcapMu packages ~42 kb of total DNA, including <3 kb of host DNA with no conserved motifs, indicative of replicative transposition with little insertion site preference. The phage genome contains 58 ORFs with comparable organization to known transposable phages. Shotgun proteomics of purified RcapMu particles detected all proteins with predicted structural functions as well as seven hypothetical proteins. Overall, comparison of RcapMu to enterobacteria phage Mu and other Mu-like phages revealed only regional homology to these phages, providing further evidence for the promiscuous, modular nature of bacteriophage evolution.

Introduction

Viruses, including bacteriophages, are estimated to be the most abundant biological entities on the planet (Suttle, 2005), and they are known to have an immense impact on microbial populations ranging from lytic dampening or augmentation of microalgae population blooms (Baudoux et al., 2006; Bratbak et al., 1996; Tarutani et al., 2000), to enhancement of pathogen persistence and virulence (Allison, 2007; Coleman et al., 1989; Karaolis et al., 1999; Saunders et al., 2001). Numerous phages have been extremely well characterized, mostly in gram positive bacteria and γ-proteobacteria, e.g. lambda (Ptashne, 2004), T4 (Miller et al., 2003), Mu (Howe and Bade, 1975), phiETA (Yamaguchi et al., 2000), and *Mycobacteria* (Hatfull et al., 2010) and *Lactococcus* phages (Brøndsted et al., 2001; Chopin et al., 2001). However, coverage of members of the α-proteobacteria is far sparser, limited mainly to *Caulobacter* phages (Poindexter, 1981), and a few *Roseobacter* (Zhao et al., 2009), *Azospirillum* (Boyer et al., 2008) and *Rhodobacter sphaeroides* phages (Abeliovich and Kaplan, 1974; Donohue et al., 1985; Duchrow and Giffhorn, 1987).

Rhodobacter capsulatus is a photosynthetic α-proteobacterium, well known for possession of diverse metabolic capabilities (Madigan and Gest, 1979; van Niel, 1944; Weaver et al., 1975). However, *R. capsulatus* is perhaps best known for an unusual horizontal gene transfer (HGT) mechanism known as the gene transfer agent (GTA) (Lang and Beatty, 2007; Marrs, 1974). Very little is known about bacteriophages of *R. capsulatus*. A general investigation was made of the infection profiles of a bank of 16 virulent viruses against 33 *R. capsulatus* stains (Wall et al., 1975), and one of these phages, RC1, was subjected to further examination (Schmidt et al., 1974). Wall et al. (1975) focused mainly on establishing the host range, investigating transduction ability and potential interaction with GTA production/release, without a detailed analysis of any one phage. Schmidt et al. (1974) looked in more depth at the bioenergetics of RC1 infection, concluding that it is energetically taxing to support infection, and also determined a *Myoviridae* morphology.

Transposition by a large group of mobile genetic elements known as insertion sequences plays an important role in the reorganization and evolution of bacterial genomes (Craig, 1995; Mahillon and Chandler, 1998; Taylor, 1963). Transposing phages belong to this group but are under-represented in the literature in comparison to their non-transposing equivalents. Enterobacteria phage Mu, henceforth referred to simply as Mu, is the archetypal transposing phage, first discovered in *Escherichia coli* K-12 (Taylor, 1963). Mu carries out two different types of transposition, conservative and replicative, at different points during the phage life cycle (Symonds et al., 1987). Conservative transposition occurs after initial infection of a cell leading to integration of the Mu genome into a single location in the host chromosome (Au et al., 2006; Liebart et al., 1982; Miller et al., 1984). Replicative transposition occurs prior to lytic release with insertion of Mu prophage copies at multiple sites in the genome (Chaconas et al., 1981), with little target sequence specificity (Ge et al., 2011; Haapa-Paananen et al., 2002; Mizuuchi and Mizuuchi, 1993).

Both forms of transposition result in the formation of a transpososome comprised of four transposase A subunits (MuA) bound to two

* Corresponding author. Fax: +1 604 822 6041.
E-mail address: jbeatty@mail.ubc.ca (J.T. Beatty).

0042-6822/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.virol.2011.09.028

attachment sites (*attL* and *attR*) and an internal enhancer region internal to the phage genome (Chaconas and Harshey, 2002; Chaconas et al., 1996). In addition, replicative transposition also requires the transposase B subunit (MuB), which binds to the host chromosome and is directly involved in the selection of insertion sites and immunity to self-integration (Ge and Harshey, 2008; Ge et al., 2010). Although on a superficial level the location of MuB binding sites and Mu insertion sites appears to be random, there are distinct peaks and troughs of the distribution of both types of site (Ge et al., 2011). The sites do not co-localize, rather integration occurs at sites adjacent to where MuB binds (Ge and Harshey, 2008; Ge et al., 2011; Mizuuchi and Mizuuchi, 1993). The phage genome is then packaged into the capsid by a headful mechanism (Mizuuchi and Craigie, 1986), which in combination with the lack of transposition site stringency, results in the almost random packaging of flanking host DNA (Bukhari and Taylor, 1975; Bukhari and Zisper, 1972).

Here, we describe and characterize a temperature-inducible transposing bacteriophage that infects *R. capsulatus*, designated RcapMu. We determined the morphology of RcapMu by electron microscopy, sequenced the DNA packaged into particles and characterized the replicative transposition sites, analysed the ORF makeup of the phage, and characterized the structural proteome. The lytic induction profile and host range were also evaluated. RcapMu represents the first well-characterized phage in *Rhodobacter* and the first active transposable phage found in the α-proteobacteria. There are no known close relatives of RcapMu, and where phages containing homologous regions do exist they are spread throughout diverse species of bacteria.

Results and discussion

RcapMu genome description

A new bacteriophage, RcapMu was initially discovered independently by two different methods. It was discovered empirically by the sequencing of a BamHI clone library of an >20 kb DNA band present in an RcGTA DNA purification from *R. capsulatus* Y262, and in silico as an apparently intact prophage cluster in the *R. capsulatus* SB1003 genome sequence that was subsequently amplified and verified experimentally. Phage DNA was isolated and sequenced (see Materials and methods), producing 1956 reads that covered 98% of the RcapMu phage genome. The remaining gaps were closed by targeted PCR and subsequent sequencing, yielding a contiguous 39,283 base sequence. Similar to the organization of Mu and its related phages, the majority of RcapMu ORFs were oriented in the same direction with the exception of the *c* repressor and adjacent genes (Fig. 1). Comparison of the RcapMu sequence with the prophage sequence in the *R. capsulatus* SB1003 genome (NC_014034) revealed several minor discrepancies: the start codons of several ORFs (RcapMu6, 15, 17, 20, 34 and 50) were adjusted, three point mutations, a two bp mutation, and a single base insertion were located in putative coding regions. Three additional ORFs designated RcapMu1 and RcapMu33 (hypothetical proteins), and RcapMu40 (a MuZ homologue, internal to RcapMu39) were also identified. Three potential −1 frameshifts were found (Moon et al., 2004), one of which occurs at position 28272 in a location where frameshifts are common in other phage (Levin et al., 1993; Summer et al., 2004). A shift would lead to a RcapMu ORFs 49/50 hybrid and would obviate the poor quality ribosome binding site of the downstream gene. However, the putative fusion protein does not have any close homologues in the database and thus no function can be assigned. Furthermore, all coding regions were assigned putative functions based on blastp hits to proteins of known function in the non-redundant (nr) database (Altschul et al., 1990, 1997; Wheeler et al., 2007), and homology to the type strain determined by direct comparison to Mu (Table 1). Interestingly, 6 ORFs within RcapMu, including two predicted to encode putative tail fiber proteins, were found to be homologous to similarly organized ORFs in a prophage cluster elsewhere in the genome: RcapMu58 (rcc02903; $2e^{-124}$), RcapMu57 (rcc02905; $2e^{-36}$), RcapMu55 (rcc02907; $1e^{-15}$), RcapMu52 (rcc02910; 4.1), RcapMu48 (rcc02914; $5e^{-31}$) and RcapMu45 (rcc02917; $2e^{-13}$).

Twenty-one sequencing reads spanned the phage-to-host DNA junctions (Fig. 2, Table S1), 13 from the left *c* repressor end of the genome (*attL*) and 8 from the right structural end (*attR*). The location of each read was mapped to the *R. capsulatus* genome, which allowed the phage ends to be definitively identified (Fig. 2A). DNA sequence and annotation data produced for RcapMu revealed several novel features of the phage ends, compared to other transposing phages. Most Mu-like phages have a short non-coding region between the start of phage sequence and the end of the *c* gene, e.g. *Pseudomonas* phages B3 (123 bp) (Braid et al., 2004) and D3112 (173 bp) (Wang et al., 2004), Mu (338 bp) (Morgan et al., 2002) and *Burkholderia* BcepMu (364 bp) (Summer et al., 2004). Consequently, annotation of Mu-like prophages is typically carried out under the assumption that *c* is always the first gene, however, the RcapMu sequence began 2429 bp downstream of the *c* gene and included five ORFs of unknown function (Fig. 1). RcapMu also shares TG/CA conserved end residues in common with Mu, BcepMu and other transposing phages. In Mu, these residues are known to be essential for the formation of a stable transpososome complex (Lee and Harshey, 2001, 2003), and mutations in these nucleotides result in varying degrees of assembly aberrations (Coros and Chaconas, 2001; Surette et al., 1991).

When Mu-like phage DNA is packaged, the terminase recognizes a pac site early in the left end of the prophage sequence and begins packaging from ~100 bp into the flanking host DNA (Allet, 1978; Daniell et al., 1975; Mizuuchi and Craigie, 1986; Summer et al., 2004). Packaging continues to <2 kb beyond the distal end of the prophage sequence until the phage head is full (Daniell et al., 1975; Mizuuchi and Craigie, 1986). As a result, a mature Mu particle contains its own genome flanked by host sequences derived from the regions on either side of the transposition insertion points. Such asymmetrical packaging of host DNA was also observed here, the left flanking sequences packaged by RcapMu were all approximately 33 bp in length whilst the largest right flanking sequence read was 797 bp in length (Table S1). The restriction profiles produced by 3 different endonucleases (BamHI, EcoRV and SalI) (Fig. 3) were a close match to those predicted for a linear genome in silico, although there were notable exceptions for fragments that included the *attR* region. An EcoRV fragment, expected to band at 3.6 kb plus host sequence, produced a smear up to 6.5 kb, indicating that up to 2.9 kb of host-derived DNA was packaged (Fig. 3). ScaI and SspI produced similar results (data not shown), whilst the 1.8 kb BamHI *attR* band is absent completely with faint smearing barely visible surrounding the 3.6 kb band. The uncut genome was estimated to be ~42 kb by pulsed field electrophoresis (data not shown), which corroborated the restriction fragment sizing data. In addition, all of the sequenced RcapMu::host DNA junctions contained unique host DNA from dispersed genomic locations (Fig. 2A and Table S1). The data summarized above, in conjunction with the absence of any highly conserved sequence motifs among the host sequences and the heterologous DNA at *attR*, implies that RcapMu transposes to seemingly random locations, consistent with the Mu model (Bukhari and Taylor, 1975; Bukhari and Zisper, 1972). Whilst Mu does not have a conserved insertion site, it does display a bias towards certain pentameric sequences and regional topological features. Originally, a preference for the degenerate sequence 5′-NYSRN-3′ was discovered (Mizuuchi and Mizuuchi, 1993), and this was subsequently refined by a study of all possible pentamer combinations (Haapa-Paananen et al., 2002). The top 10% most frequent integration events occurred into 5′-CYSRG-3′, whereas the lowest 10% integrated into the almost converse 5′-NRWYN-3′ sequence (Haapa-Paananen et al., 2002). We were unable to detect any such bias for RcapMu in the sequencing data set, but future in vitro assays may reveal analogous preferences.

In characterized transposing phages, Mu-like phage *att* sites are recognized by the transposase during integration and replication, and there are two characterized *att* configurations in transposing

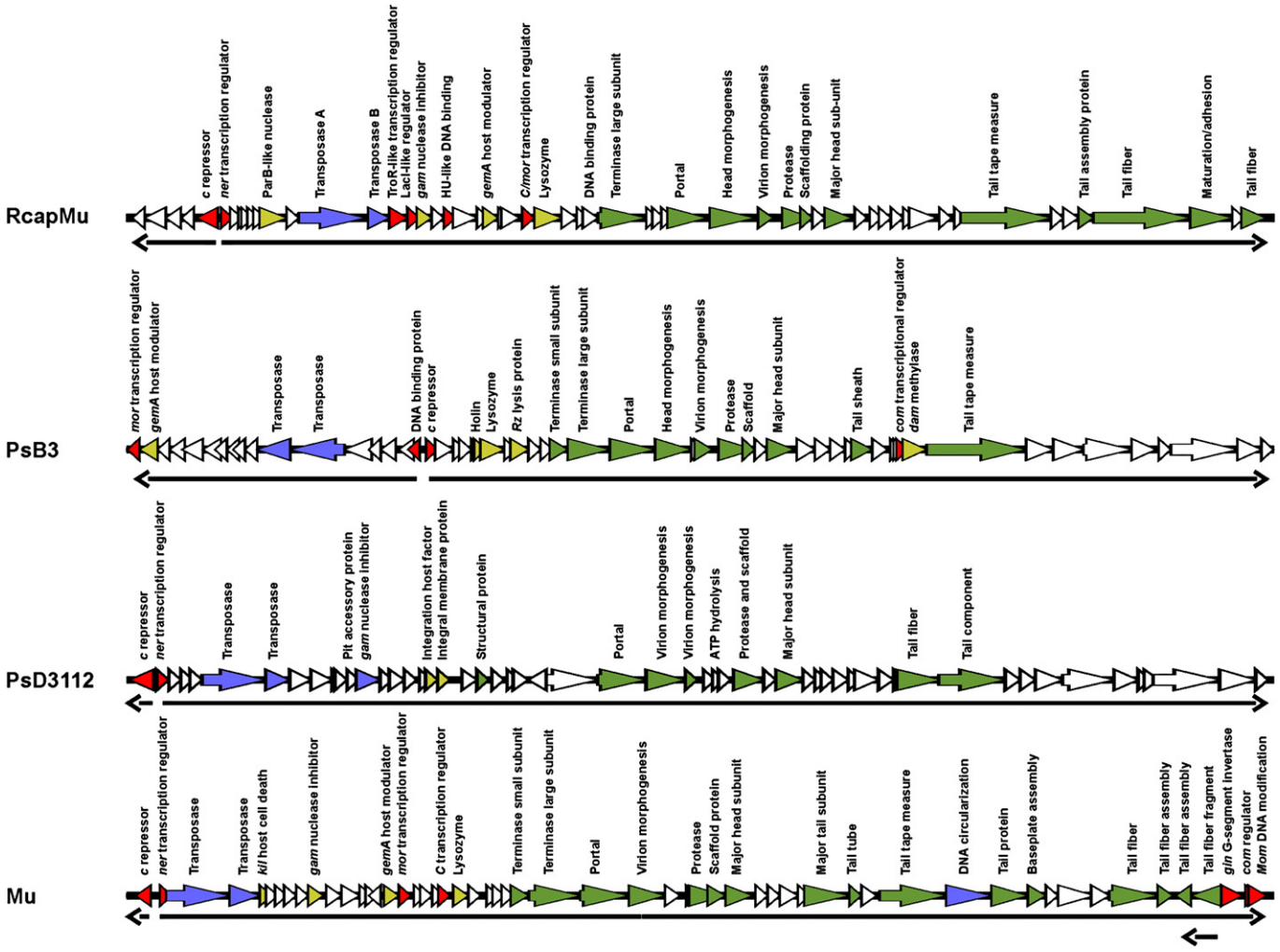

Fig. 1. Comparative schematic of the RcapMu, Mu, B3 and D3112 phage genomes. All ORFs are represented by colored arrows, of which those with predicted structural function are green, phage regulation are red, integration and replication are blue, host interaction are yellow, and hypothetical are white. ORFs encoding proteins with a predicted function are labeled above, and the direction of transcription is displayed as arrows below each map. RcapMu shares the organizational layout with the model phages *E. coli* Mu and *Pseudomonas* phage D3112, whereas *Pseudomonas* phage B3 differs slightly. Both B3 and D3112 produce *Siphoviridae* phage particles similar to RcapMu, whereas the *E. coli* Mu has a *Myoviridae* structure.

prophages. The first was initially identified in Mu (Kahmann and Kamp, 1979), and consists of three ~17 bp imperfect repeats spread throughout the first and last 100 bp of the phage genome (*attL*1–3 and *attR*1–3). The second, which resembles transposon Tn552 and was identified in BcepMu (Summer et al., 2004), consists of a pair of imperfect direct 23 bp repeats two bases from the phage ends. The RcapMu DNA sequences indicate sites more similar to the BcepMu configuration, with 24 bp motifs that start 2 bp from the phage ends but repeated only in the *attR* region (consensus: YRGVSSKBGAHYASCTTYCGGACG). MuA transposase binding to each of the *att* subsites is variable in both affinity and importance: in vitro only the *attL1*, *attR1* and *attR2* sites are essential structural binding sites in the transpososome (Craigie et al., 1984; Lavoie et al., 1991; Mizuuchi et al., 1991). Single site deletions or substitutions of the remaining sites (*attL2*, *attL3* and *attR3*) dramatically reduced integration efficiency but this was partially mitigated by addition of IHF, indicating that these sites are non-essential but not completely expendable (Allison and Chaconas, 1992). The three 24 bp motifs identified in RcapMu correspond to the three essential MuA binding sites, and whilst we were unable to identify any of the non-essential MuA binding sites in silico there may not be a strong conservation of these sequences. Of note is that, unlike in Mu and many Mu-like phages, the RcapMu transposase A and c proteins share no significant sequence similarity (Harshey et al., 1985). Typically, the repressor contains a DNA-binding motif homologous to that of the transposase, and thus prevents replication by blocking access to the *att* sites (Harshey et al., 1985).

No left-flanking and right-flanking host sequences were contiguous in our data, but as Mu is known to transpose to 50 to 200 sites per cell during replication (Symonds et al., 1987), detection of contiguous host DNA is unlikely without massive sequencing coverage. However, when the DNA sequences flanking the RcapMu prophage in the published *R. capsulatus* SB1003 genome sequence (NC_014034) (Strnad et al., 2010) were joined together (i.e. with the intervening phage sequence deleted) the reconstituted gene had almost 100% DNA sequence identity to a second ORF (rcc02097) in the chromosome predicted to encode a transposase (Fig. 2B), which does not appear to be part of any bacteriophage or IS element (Mahillon and Chandler, 1998). The only discrepancy was a 5 bp duplication of host sequence at the site of phage integration (Fig. 2B). Mu transposition in *E. coli* is known to produce duplications of host DNA during lysogenization, resulting in a prophage flanked by identical 5 bp direct repeats (Allet, 1979). The presence of these direct repeats in RcapMu suggests that the phage may have initially integrated within a transposase gene identical to rcc02097 in SB1003 and created a 5 bp duplication at the site of integration, although additional examples would be necessary to unequivocally determine this.

Table 1
RcapMu ORFs and BLAST hits.

Name	Synonym	AA length	Predicted product	Mu homologue (by function)	Best phage hit (E-value)	Id[a], Sim[b], Cov[c]
RcapMu1	N/A	45	Hypothetical protein			
RcapMu2	rcc01015	67	Hypothetical protein			
RcapMu3	rcc01014	230	Hypothetical protein			
RcapMu4	rcc01013	159	Conserved hypothetical protein			
RcapMu5	rcc01012	140	Hypothetical protein			
RcapMu6	rcc01011[d]	239	c-repressor	(Mup1) c	*Pseudomonas* phage DMS3 (3e−18)[e]	31%,47%,97%
RcapMu7	rcc01010	98	ner-like transcriptional regulator	(Mup2) ner	*Escherichia* phage D108 (0.001)[e]	32%,60%,63%
RcapMu8	rcc01009	106	Hypothetical protein			
RcapMu9	rcc01008	45	Hypothetical protein			
RcapMu10	rcc01007	42	Hypothetical protein			
RcapMu11	rcc01006	97	Hypothetical protein			
RcapMu12	rcc01005	64	Hypothetical protein			
RcapMu13	rcc01004	284	ParB-like nuclease			
RcapMu14	rcc01003	156	Conserved transposable prophage hyp.			
RcapMu15	rcc01002[d]	763	Transposase A	(Mup3) A	*Pseudomonas* phage D3112 (8e−99)[e]	35%,51%,89%
RcapMu16	rcc01001	258	Transposition protein B	(Mup4) B	*Pseudomonas* phage D3112 (1e−04)[e]	22%,46%,81%
RcapMu17	rcc01000[d]	186	TroR domain transcriptional regulator			
RcapMu18	rcc00999	122	LacI-like regulatory protein		*Pseudomonas* phage MP38, (1e−08)[e]	33%,54%,87%
RcapMu19	rcc00998	181	Host-nuclease inhibitor Gam	(Mup10) gam	*Enterobacteria* phage Mu (2e−34)[e]	43%,68%,93%
RcapMu20	rcc00997[d]	78	Hypothetical protein			
RcapMu21	rcc00996	92	DNA-binding protein HU	(Mup17) mor		
RcapMu22	rcc00995	295	Hypothetical protein			
RcapMu23	rcc00994	72	Hypothetical protein			
RcapMu24	rcc00993	149	GemA-like (modulation of host genes)	(Mup16) gemA	*Burkholderia* phage phiE255 (3e−04)	29%,44%,89%
RcapMu25	rcc00992	36	Hypothetical protein			
RcapMu26	rcc00991	226	Conserved transposable phage hyp.		*Pseudomonas* phage B3 (2e−05)[e]	27%,42%,92%
RcapMu27	rcc00990	123	Putaitve transcriptional regulator	(Mup21) C	EBPR siphovirus 4 (4e−29)	54%,70%,94%
RcapMu28	rcc00989	310	Lysozyme	(Mup22) lys	Phage Gifsy-2 (6e−10)	39%,50%,46%
RcapMu29	rcc00988	171	Conserved hypothetical protein			
RcapMu30	rcc00987	89	Hypothetical protein			
RcapMu31	rcc00986	168	Putaitve DNA binding protein		*Pseudomonas* phage MP29 (1e−25)[e]	40%,58%,94%
RcapMu32	rcc00985	553	Large terminase	(Mup28) H	*Pseudomonas* phage MP38 (1e−146)[e]	52%,69%,91%
RcapMu33	N/A	50	Hypothetical protein			
RcapMu34	rcc00984[d]	113	Conserved hypothetical protein		*Burkholderia* phage Bups phi1 (1e−20)	53%,63%,93%
RcapMu35	rcc00983	75	Hypothetical protein			
RcapMu36	rcc00982	459	Putative portal protein	(Mup29) Portal	*Pseudomonas* phage B3 (5e−60)[e]	40%,57%,73%
RcapMu37	rcc00981	491	Head morphogenesis (muf-like)	(Mup30) F	*Pseudomonas* phage B3 (7e−39)[e]	39%,54%,45%
RcapMu38	rcc00980	185	Virion morphogenesis (head–tail joining)	(Mup31) G	*Pseudomonas* phage B3 (7e−21)[e]	35%,51%,91%
RcapMu39	rcc00979	371	Putative protease	(Mup32) I	*Burkholderia* phage phiE255 (4e−55)	38%,52%,96%
RcapMu40	N/A	152	Putative scaffolding protein	(Mup33) Z	*Pseudomonas* phage B3 (2e−08)[e]	42%,58%,58%
RcapMu41	rcc00978	115	Hypothetical protein			
RcapMu42	rcc00977	318	Major capsid protein	(Mup34) T	*Burkholderia* phage phiE255 (2e−40)	36%,54%,94%
RcapMu43	rcc00976	163	Hypothetical protein			
RcapMu44	rcc00975	97	Conserved hypothetical		*Aeromonas* phage 25 (0.070)	31%,55%,66%
RcapMu45	rcc00974	142	Conserved transposable prophage hyp.	(Mup36)	*Pseudomonas* phage PA1/KOR/2010 (3e−10)	35%,54%,90%
RcapMu46	rcc00973	144	Hypothetical protein			
RcapMu47	rcc00972	66	Hypothetical protein			
RcapMu48	rcc00971	309	Conserved transposable prophage hyp.		uncul. MedDCM-OCT-S09-C399 (2e−41)	34%,54%,95%
RcapMu49	rcc00970	126	Hypothetical protein			
RcapMu50	rcc00969[d]	96	Conserved transposable prophage hyp.		*Sodalis* phage SO-1 (0.021)	38%,55%,75%
RcapMu51	rcc00968	1028	Tape measure protein	(Mup42)	*S. epidermidis* RP62A phage SP-beta (8e−15)	44%,59%,11%
RcapMu52	rcc00967	141	Conserved transposable prophage hyp.		*Burkholderia* phage BcepGomr (0.13)	26%,39%,77%
RcapMu53	rcc00966	176	Conserved transposable prophage hyp.			
RcapMu54	rcc00965	147	Putative tail assembly protein	(Mup51) U′	*Cronobacter* phage ENT47670 (0.037)	29%,41%,76%
RcapMu55	rcc00964	1143	Tail fiber	(Mup52) S′	*Burkholderia* phage BcepGomr (3e−25)	24%,40%,54%
RcapMu56	rcc00963	444	Maturation/adhesion		*Salmonella* phage Vi06 (1e−09)	28%,41%,53%
RcapMu57	rcc00962	89	Conserved hypothetical protein			
RcapMu58	rcc00961	327	Putative tail fiber		*Synechococcus* phage S-SSM5 (5e−13)	30%,43%,53%

[a] Percentage identity.
[b] Percentage similarity.
[c] Percentage coverage.
[d] Indicates an adjusted start site in RcapMu vs rcc annotations.
[e] Indicates a phage known to be transposable.

RcapMu ORF sequence comparisons

There were clear homologues of many RcapMu proteins in characterized phages from diverse genera (Table 1), chiefly *Pseudomonas* and *Burkholderia*. Notably, the transposases (A and B) and several head assembly proteins were most similar to counterparts from *Pseudomonas* phages D3112 and B3, respectively, whilst capsid and tail proteins were related to *Burkholderia* phages phiE225 and BcepGomr, respectively. However, comparisons on a full-genome scale at both the nucleotide (Figs. S1A–D) and translated nucleotide levels (Fig. S2, Tables 1 and S2) revealed that similarities to other known phages were of limited extent. The nearest approximations to related phages are found integrated in the genomes of disparate members of the α-proteobacteria, e.g., *Agrobacterium vitis* S4, *Polymorphum gilvum* SL003B-26A1, and *Roseibium* sp. TrichSKD4 (Table S2). However, even these putative prophages lack homologues of many RcapMu proteins and the similarities between

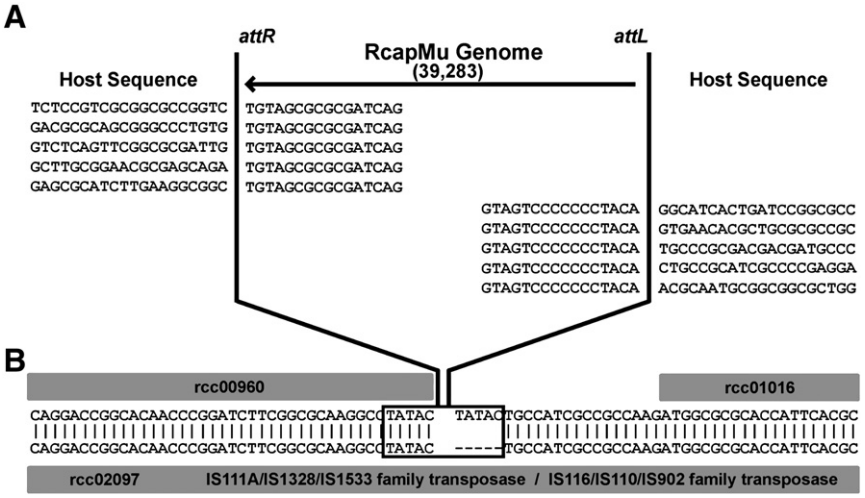

Fig. 2. Characterization of regions flanking the RcapMu genome. A. Each sequence represents a read from RcapMu packaged DNA that included *attL/R* sites, where one end mapped to RcapMu and the other mapped elsewhere in the SB1003 genome. The vertical bar denotes the transposition point. There appear to be no conserved motifs, locations or functions associated with the host flanking sequences. B. The sequence depicts an alignment of the regions flanking the integrated RcapMu prophage in the *R. capsulatus* chromosome joined together (top sequence), to another region within the chromosome predicted to encode a transposase (bottom sequence). Genes are represented by gray boxes and identified by their locus tag. The two sequences are identical, barring the 5 bp duplication highlighted by a black box.

these phages and RcapMu are limited to distinct regions and large swathes were not similar at all. This is further evidence for the well-characterized modular structure of bacteriophage genomes in general (Arbiol et al., 2010; Botstein, 1980). Interestingly, the *A. vitis* S4 prophage also appears to have been dissected, with one section containing the *c* repressor and the other the structural genes, with homologues of RcapMu proteins duplicated in both sections, possibly as a result of insertion, duplication or rearrangement events within the phage genome (Fig. S2).

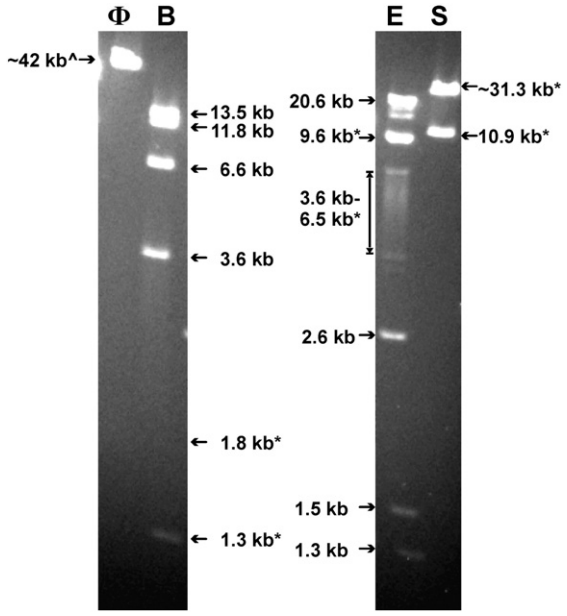

Fig. 3. Restriction enzyme digest analysis of RcapMu DNA. Purified phage DNA uncut (Φ), cut with BamHI (B), EcoRV (E) and SalI (S). Estimated sizes based on comparison to molecular weight markers (λ *Hin*d III, and NEB 1 kb ladder) with comparison to in silico predictions for fragments >23 kb (~). Bands from the RcapMu prophage termini (*) and uncut phage DNA (^) are indicated. All bands matched the predicted sizes barring the bands containing the *attR* terminus. The packaging of host DNA resulted in smearing of the 3.6 kb EcoRV band up to 6.5 kb, consistent with packaging of 2.9 kb of extra DNA, whereas the predicted 1.8 kb BamHI band is absent, likely due to a similar smearing (barely visible above and below the 3.6 kb band).

RcapMu host range

The host range of RcapMu was investigated using a variety of *R. capsulatus* strains and several different genera (Table 2). Spot infection assays were successful in producing visible lysis in only 2 of the assays, using *R. capsulatus* YW1 and YW2 as hosts, suggesting that RcapMu has a narrow host range. A lack of lytic infection was expected for the wild-type *R. capsulatus* strain B10 and the strains derived from it (SB1003, R121 and Y262), because RcapMu was isolated from SB1003 (i.e. these strains should contain RcapMu and hence be resistant to infection). Subsequent PCR screening of all strains tested using oligonucleotide primers specific to the capsid and *c* genes revealed 519 bp and 482 bp bands, respectively, that correlated with the apparent resistance of several potential hosts to infection. In addition to B10 and its

Table 2
Bacterial strains used in this study. All strains are designated to indicate the presence or absence of the RcapMu *c* repressor and capsid genes (determined by PCR), and whether plaques form when infected with a concentrated preparation of RcapMu.

Strain	Ref or source	*c*	Capsid	Plaque
Rhodobacter capsulatus				
SB1003	Yen and Marrs (1976)	+	+	−
B10	Weaver et al. (1975)	+	+	−
B6	Weaver et al. (1975)	−	−	−
SP18	Weaver et al. (1975)	−	−	−
SP36	Weaver et al. (1975)	−	−	−
H9	Weaver et al. (1975)	−	+	−
P12F1	Weaver et al. (1975)	−	−	−
YW1	Weaver et al. (1975)	−	−	+
YW2	Weaver et al. (1975)	−	−	+
YW1 RcapMu lysogen	This study	+	+	−
R121	Scolnik et al. (1980)	+	+	−
DE442	?[a]	+	+	−
Y262	Yen et al. (1979)	+	+	−
St. Louis	Weaver et al. (1975)	+	−	−
Rhodobacter sphaeroides 2.4.1	Mackenzie et al., (2001)	+	−	−
Paracoccus denitrificans	ATCC 17741	−	−	−
Rhodopseudomonas palustris CGA001	Harwood and Gibson (1986)	−	−	−
Rhodospirillum centenum	Favinger et al. (1989)	−	−	−
Escherichia coli BL21	Invitrogen	−	−	−

[a] Of uncertain provenance, a *crtD* mutant probably derived from Y262 (B. Marrs, personal communication).

derivatives, there were 3 strains that yielded an appropriately sized PCR product using *c* primers, and 2 strains using capsid primers (Table 2). Therefore, the absence of infection of *R. capsulatus* strains B10, SB1003, R121, DE442, Y262, H9, St. Louis, and *R. sphaeroides* 2.4.1 may be explained by homo-immunity due to an incumbent related prophage (Schumann, 1979). To confirm that an RcapMu lysogen confers resistance to superinfection to a homo-immune phage, an RcapMu lysogen of YW1 was infected with a high concentration RcapMu suspension (10^{10} pfu ml^{-1}). As anticipated, no lysis was evident, confirming the existence of a RcapMu lysogen-dependent immunity (data not shown).

Temperature-induction

In contrast to lambdoid bacteriophages, characterized transposing phages are relatively rare in the literature. Often, the primary method for induction of unknown prophages from a sample is stimulation of the lytic cycle with DNA damaging agents such as mitomycin C or UV light (Heinemann and Howard, 1964), yet only phages with a repressor protein analogous to that of bacteriophage lambda are known to be induced by this kind of treatment. There is no comparable method available for induction of high titres of the wild type Mu phage (Howe and Bade, 1975), which may partially account for the vast disparity between the relatively great number of characterized lambdoid phages compared to transposing phages. Instead, a variety of Mu temperature-sensitive mutants have been discovered that contain amino acid substitutions in the c repressor protein, which predispose it to loss of activity under elevated temperatures (Vogel et al., 1991). The RcapMu putative *c* gene is located in a position comparable to the Mu *c* but an alignment indicated only 15% amino acid identity and 25% similarity between these repressor proteins (Fig. 6A). Whilst this is clearly insufficient for definitive analysis, we observed that one amino acid occurs in the temperature-sensitive Mu cts62 form (a change of R47 to Q) in RcapMu (Vogel et al., 1991). To test whether RcapMu is induced by elevated temperatures, *R. capsulatus* SB1003 was grown at moderately elevated temperatures (37 °C and 40 °C). Elevated growth temperatures considerably increased phage production, with the specific plaque production (Pfu ml^{-1}·OD$_{660}^{-1}$) rising from 3×10^2 at 30 °C to 4×10^3 at 37 °C and to almost 2×10^4 at 40 °C, representing a >60-fold increase (Fig. 6B). Although the total phage titres were not as high as for the Mu *c* mutants in *E. coli*, the inductive effect was significant and enabled further characterization of the phage when combined with a multi-step enrichment process often used to concentrate phages from the environment. A similar temperature regimen methodology could easily be adapted to the identification of phages from stable lysogens insensitive to standard induction procedures, enabling the discovery of a greater range of bacteriophages.

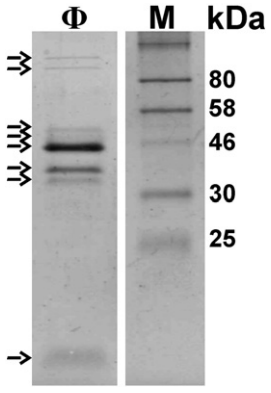

Fig. 4. SDS-PAGE gel of RcapMu particles. A sample of a concentrated, CsCl gradient-purified RcapMu suspension, run on a denaturing SDS-PAGE gel (Φ), with eight visible bands (identified by arrows). A broad range, pre-stained protein molecular weight marker is included for reference (M) with the molecular weights given on the right of the figure.

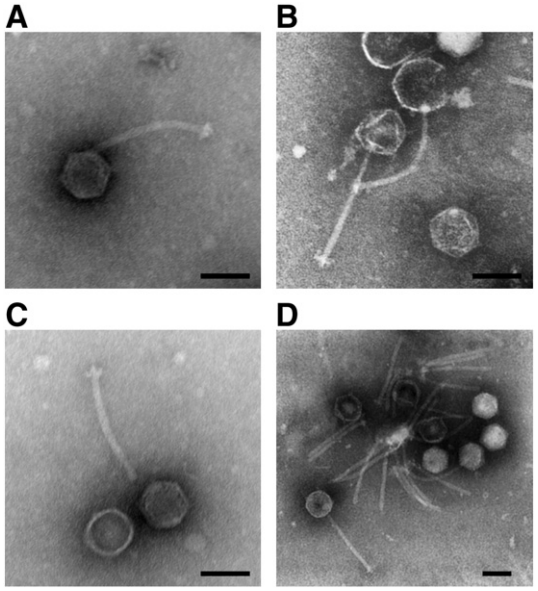

Fig. 5. Electron Micrographs of RcapMu. A. Intact RcapMu particle with a typical *Siphoviridae* morphology. B. An intact RcapMu particle beside several damaged phages. C. Dissociated head and tail sections of RcapMu particles. D. Numerous dissociated heads and tails, suggesting a labile nature of the phage. Scale bar in lower corner of each panel represents 100 nm.

Analysis of RcapMu proteins

The yields of RcapMu from culture supernatants were insufficient for most downstream applications, and thus it was necessary to carry out several rounds of concentration (see Materials and methods). Titres of 10^{11} to 10^{12} ml^{-1} were purified by CsCl gradient equilibrium ultracentrifugation, to reduce the presence of contaminants that may interfere with analysis. Two distinct white bands were produced when the CsCl gradient ultracentrifuge tube was illuminated from an acute angle with white light, and this double banding was observed in most but not all independent experiments. After collection, the buoyant densities of these bands were determined to be 1.48 and 1.38 g/ml by refractometry, with the 1.38 g/ml (upper) band appearing far brighter and retaining a higher infective titre than the lower, although both bands contained infective particles. There were no discernible differences in DNA restriction profiles between the bands (Fig. S3A), ruling out the possibility that the disparity in density was due to two different phages or empty phage particles.

Both bands were separately subjected to a second round of equilibrium centrifugation. The gradient tube in which the upper band was repurified produced a tight visible phage band containing 1.1×10^{10} pfu ml^{-1} with an absorbance at 280 nm of 0.151, however, infective particles were present throughout the gradient but at lower concentrations (typically 10^5–10^7 pfu ml^{-1}). The gradient tube in which the lower band was repurified contained a total of only 2.9×10^7 pfu ml^{-1} (sum of the visible band fraction and one immediately above), despite a comparable absorbance at 280 nm of 0.157. This indicates that the source of plaque forming units in the lower CsCl band was cross-contamination from the top band.

Concentrated CsCl phage samples were run in SDS-PAGE to evaluate the number and apparent sizes of the structural proteins. The samples contained six to eight visible bands of 30 to >80 kDa (Figs. 4 and S3B). To more fully determine the RcapMu structural proteome, a shotgun peptide identification approach was employed (Lavigne et al., 2006). There were no peptides that hit against any putative prophage regions within the SB1003 genome, indicating a pure RcapMu preparation. This shotgun method detected all proteins predicted to be present in mature phage particles, as well as a number of hypothetical proteins from the structural region of the genome (Table 3). In total an additional seven proteins were identified by mass spectrometry, compared to SDS-PAGE, which highlights the potency of this approach for the detection of low size/

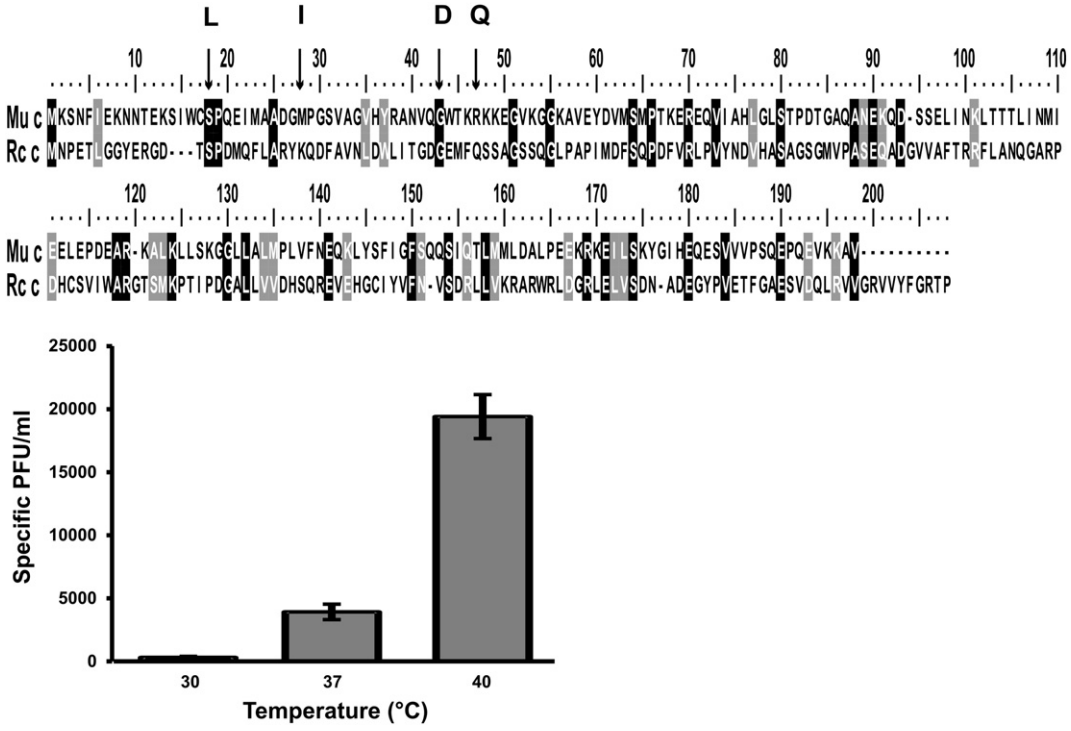

Fig. 6. The c repressor and temperature induction of the RcapMu lysogen. Top: Alignment of the wild-type *E. coli* Mu c repressor protein (Mu c) with the RcapMu c protein (Rc c). Four of the most potent temperature-sensitive mutations that have been described for Mu c are indicated above the sequence with arrows, with the amino acid substitutions indicated in single letter code. Identical sequence matches between Mu c and Rc c are boxed in black and similarities in gray. Bottom: Bar chart of the specific RcapMu pfu.ml^{-1} (normalized to OD_{660}) produced by *R. capsulatus* SB1003 cultures grown at 30 °C, 37 °C and 40 °C. In YPS rich, complex medium under aerobic (chemotrophic) conditions. Error bars represent standard deviation (n = 4).

abundance proteins in a complex sample. Both SDS-PAGE and the proteomic approach detected differences between two RcapMu bands of different densities produced by equilibrium centrifugation in CsCl gradients, primarily those proteins predicted to make up the phage tail.

Table 3
Structural proteome of RcapMu. Listed are all proteins that were identified by the proteomics of purified phage particles. Predicted molecular weights were calculated based on the amino acid sequence, whilst SDS-PAGE weights were assigned based on comparison to molecular weight standards. Initial confirmation of the identity of the major proteins was carried out by gel excision and analysis of the brightest bands (Band MS/MS) whereas a more all-encompassing analysis was used to identify all structural proteins present (Shotgun MS/MS). For each of these proteomic approaches the results are presented as the total score/number of peptide hits, where total score is a non-probabilistic value derived from the sum of the probabilities that each individual ion match is a random event (−10∗Log(P)).

Gene ID	Calculated MW (kDa)	SDS-PAGE MW (kDa)	Putative protein function	Band MS/MS	Shotgun MS/MS
RcapMu36	50.46		Portal		312/09
RcapMu37	55.53		Head Morphogenesis		85/03
RcapMu38	19.81		Virion Morphogenesis		34/01
RcapMu39	39.63	~20[b]	Protease	34/1	196/08
RcapMu41	11.20		Hypothetical		237/05
RcapMu42	34.69	45[a]	Major Head Subunit	5284/144	504/15
RcapMu45	14.72		Hypothetical		111/03
RcapMu46	15.21		Hypothetical		69/01
RcapMu48	32.54		Hypothetical		198/08
RcapMu51	107.07		Tail Tape Measure		285/07
RcapMu52	15.79		Hypothetical		74/02
RcapMu53	19.29		Hypothetical		63/01
RcapMu55	122.11		Tail Fiber		256/10
RcapMu56	46.85		Maturation/Adhesion		300/08
RcapMu58	32.76		Tail Fiber		45/01

[a] Size appeared larger than predicted for an unknown reason.
[b] All MS:MS hits mapped to the N-terminal half of the protein suggesting there may have been proteolytic modification.

Electron microscopy

Purified RcapMu from the upper (lower density) CsCl gradient band was also analysed by transmission electron microscopy. The structural genes present in the phage genome code for a putative tail assembly protein (RcapMu54), tail fibers (RcapMu55 and 61) and a relatively large putative tape measure protein (RcapMu51, 107 kDa), each of which were detected in the shotgun proteomics experiments (Table 3). The presence, type and size of these proteins indicated that RcapMu should have a *Siphoviridae* morphology with a long, flexible tail, and this proved to be the case (Figs. 5A–D). It has previously been demonstrated that the length of tape measure proteins is directly proportional to the length of the tail (Hendrix, 1988; Journet et al., 2003), and the RcapMu particles detected here possessed an average tail length somewhat larger than expected, >200 nm. Estimates using a 'protein ruler' in lambda found that each amino acid residue of the tape measure protein is approximately equivalent to 0.16 nm (Katsura, 1987), which in the case of RcapMu would predict a tail length of 164 nm (Figs. 5A–D). The RcapMu tail is longer than many other well-characterized *Siphoviridae* such as bacteriophage lambda, 150 nm (Hendrix, 1983), and *Pseudomonas* phage B3, 163 nm (Slayter et al., 1964), but comparable to several Mycobacterial *Siphoviridae* phages, <275 nm (Pedulla et al., 2003), and the myovirus BcepMu, 220 nm (Summer et al., 2004).

Unfortunately, imaging of intact RcapMu phage particles was difficult because they appeared to be extremely sensitive to the staining procedure, and so there was extensive capsid damage evident in samples using any of three negative stains (uranyl acetate, phosphotungstic acid and ammonium molybdate). Phages stained with uranyl acetate were relatively well preserved whilst phosphotungstic acid staining resulted in universally collapsed or exploded phage heads (Fig. 5B). It was also clear that the phage tails were readily dissociated from the capsid because the majority of phages observed were separated from their tails, with only a small proportion fully intact (Figs. 5C and D). The tails and capsids co-purified in the CsCl gradient and both were

present in approximately equal quantities in the EM images, indicating that separation most likely occurred after banding. Furthermore, all phages observed in the lower (greater density) band had no or short truncated tails (Figs. 5 and S3). The labile nature of the head–tail junction noted for the upper band (Figs. 5C and D) suggests that loss of the tail during purification may have been responsible for the altered buoyant densities within the RcapMu samples. It appears that the lack of a tail and host recognition proteins in the phages of the lower band is responsible for the differential banding and supports the aforementioned hypothesis that the plaques produced were due to cross-contamination between the two CsCl bands.

Transposing phages

Although, other than Mu, transposing phages are poorly characterized, Mu-like prophage regions have been found in many species including a number of pathogenic species such as *Haemophilus influenza* (FluMu), *Neisseria meningitides* (Pnm1) (Morgan et al., 2002), and several remnant elements in *Vibrio cholerae* (Heidelberg et al., 2000) and *Yersinia pestis* (Parkhill et al., 2001). Representatives of Mu-like transposing phages in multiple distantly related species indicate that they are widespread mobile genetic elements, although as yet relatively understudied. It is important to increase the breadth of knowledge of transposing phages, to increase the detection of novel phages and accumulate transferable data potentially applicable to diverse phage systems. In contrast to most species, a multitude of Mu-like bacteriophages able to infect *Pseudomonas aeruginosa* have been described (Akhverdian et al., 1984), and two of these *Pseudomonas* phages, D3112 and B3 (Figs. 1 and S1), are similar to RcapMu in at least two key areas. Both phages possess Mu-like replication modules and overall genome organization, but they also both contain structural genes that are more akin to *Siphoviridae* bacteriophages (Wang et al., 2004). Often bacteriophages are classified according to their morphological features, but with the advent of genome sequencing this has been shown to be superficial (Hendrix et al., 1999; Rohwer and Edwards, 2002). For example, Stx-encoding phages possess lambdoid genomes but are often structurally *Podoviridae* (Allison et al., 2003; Smith et al., 2007), whilst several publications, including this one, have described Mu-like phages with *Siphoviridae* morphology (Slayter et al., 1964; Wang et al., 2004). This evidence supports the modular theory of bacteriophage evolution and further highlights the mosaicism found amongst bacteriophages in general (Botstein, 1980; Hendrix et al., 1999).

Conclusion

RcapMu is the first bacteriophage infecting *R. capsulatus* to be characterized in depth, and the first transposing bacteriophage described that infects an α-proteobacterium, although even a cursory examination of the genomes of this class of bacteria reveals a plethora of potentially intact and remnant transposing prophages across a myriad of species. Our discovery of putative prophages in other α-proteobacteria that share some homology with RcapMu suggests that RcapMu may serve as an excellent starting point for the exploration of phage diversity in this class. Discovery and characterization of new lysogenic bacteriophages is currently hampered by the dearth of information available about atypical lytic induction and lifestyles of transposing phages, which can be remedied only by the intensive study of phylogentically diverse host species.

Materials and methods

Bacterial strains and growth conditions

The *R. capsulatus* strain Y262 (Yen et al., 1979) and the genome-sequenced strain SB1003 (Strnad et al., 2010) were initially analysed for bacteriophage production, and the latter was used for all subsequent phage production experiments. Strain YW1 (Weaver et al., 1975) containing the cosmid pLAFR1 (Friedman et al., 1982) was the strain used in plaque assays and propagation techniques. These and additional strains used for screening purposes are listed in Table 2, with growth conditions for *R. capsulatus* strains as indicated, and for other species as recommended by the supplier.

DNA purification for sequencing

Cells were removed from a 500 ml *R. capsulatus* SB1003 aerobic stationary phase culture grown at 37 °C in YPS medium (Wall et al., 1975) by centrifugation. The supernatant was treated with 1 $mg \cdot ml^{-1}$ DNAse I for 1 h at 37 °C. NaCl was added to a concentration of 1 M, and the solution was centrifuged again. The supernatant was brought to 10% (w/v) PEG8000, incubated at 28 °C for 2 h and the phage harvested by centrifugation at 6800 × *g* for 20 min. The pellets were resuspended in G buffer (Solioz and Marrs, 1977) overnight at 4 °C. Concentrated phage suspensions were treated with 2 U DNase I and 1.2 U RNase A at 37 °C for 30 min. Proteins were removed using phenol:chloroform: isoamyl alcohol (25:24:1) and DNA was ethanol-precipitated. In addition to the expected 4 kb RcGTA DNA band, a >20 kb DNA band was observed. A clone-library of the sample was created by BamHI digestion and subsequent ligation into pUC19, and uniquely sized inserts were sequenced using M13 primers. Preliminary sequencing included reads that mapped to the transposable prophage cluster (Fig. S1). The sample was then run on an agarose gel and phage DNA extracted by electro-elution and concentrated with a centrifugal filter unit (Millipore, MA). The concentration was determined by spectrophotometry and 280 ng was submitted to the Broad Institute (Cambridge, MA) for high-throughput (Roche 454) pyrosequencing (Margulies et al., 2005). Gaps in the sequence were closed by PCR amplification and sequencing across the gaps.

Bacteriophage induction

For induction experiments, *R. capsulatus* SB1003 was grown to stationary growth phase at 30 °C, 37 °C or 40 °C in YPS (Wall et al., 1975) or RCV medium (Beatty and Gest, 1981), as indicated. Cells were removed by centrifugation and the supernatant filtered through a 0.2 μm filter. For heat shock experiments, the cultures were grown to mid-exponential phase, shocked by transfer to 42 °C for 30 min and returned to 30 °C to recover for 2 h. Cells were then removed by centrifugation and the supernatant filtered through a 0.2 μm filter.

Phage infection assays

Phage titres were calculated by plaque assay using 300 μl of a mid-exponential YW1(pLAFR1) culture mixed with 500 μl serial dilutions of phage lysate, and incubated for 30 min at 37 °C. The mixture was then added to 8 ml of molten top agar (0.4% w/v agarose in YPS broth), poured onto YPS agar plates and incubated overnight at 37 °C. Host range spot assays were carried out by adding 300 μl of each host strain to 8 ml of molten top agar (0.4% w/v agarose in strain-specific growth medium) and pouring this mixture onto host-specific agar plates. Once set, 50 μl drops of a 1×10^{10} pfu ml^{-1} RcapMu suspension were added to the surface and allowed to dry. Plates were incubated overnight at 37 °C.

PCR amplification

100 ng of purified chromosomal DNA (Rapley et al., 2000) from each of the host range strains tested (Table 2) was amplified with Taq DNA polymerase (New England Biolabs) using primers specific for the RcapMu *c* (MuC F: ACATGCAGTTCCTTGCTCGC and MuC R: ATAGCCCTCGTCGGCATTGT) and the capsid (MuCap F: TCGATGCGCGTCCTGAAGTT and MuCap R: ATAGGCCGCCAGCATGTCAA).

Phage titre amplification

RcapMu titres were insufficient for most downstream applications and required amplification using a 3-step concentration method. In brief, stationary growth phase SB1003 cultures grown in YPS medium at 37 °C were centrifuged and the supernatant passed through a 0.2 μm filter. Phages were then precipitated using polyethylene glycol (PEG) according to a protocol from Rooks et al. (2010). For the second step, these precipitates were used to produce semi-confluent lysis in plates, as for the plaque assay, which were overlaid with 5 ml TBT buffer (100 mM $MgCl_2$, 100 mM NaCl, 100 mM Tris–HCl, pH 7.0) and rocked gently at 4 °C for 1 h. Both the overlay buffer and the top agar were recovered, cells and debris were removed by centrifugation, and the supernatant passed through a 0.2 μm filter. The final concentration step consisted of ultracentrifugation of the filtrate at 300,000×*g* for 4 h. Phage pellets were resuspended in 1 ml TBT, titred by plaque assay and corroborated by epifluorescence microscopy, according to the methods from Patel et al. (2007). Phage particles were ultimately purified by equilibrium centrifugation in CsCl at a starting density of 1.5 $g \cdot ml^{-1}$ in a swinging bucket rotor, as previously described (Clokie and Kropinski, 2009).

Restriction analysis

A 100 μl phage suspension was mixed vigorously with an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1), and centrifuged at 13,000×*g* for 5 min. The aqueous phase was removed to a fresh tube, the DNA ethanol-precipitated, washed in 70% ethanol, and the pellet dissolved in 10 mM Tris–HCl (pH 8.0). Restriction digests of RcapMu DNA were carried out according to manufacturer's instructions for each enzyme (New England Biolabs), and analysed using standard pulsed field gel electrophoresis conditions.

Protein analysis and identification

CsCl-purified phage suspensions were denatured by heat treatment at 95 °C for 5 min in Laemmli buffer and run on a denaturing 10% SDS-PAGE gel (Laemmli, 1970; Nandi and Lewis, 1970). The proteins were visualized by Coomassie staining (Fazekas de St Groth et al., 1963). Phage protein bands were excised from a Coomassie-stained SDS-PAGE gel and digested in-gel with trypsin (Kinter and Sherman, 2000), whilst whole particle shotgun samples (Lavigne et al., 2006) were first digested with trypsin and then treated with DNase I. All samples were then analysed on an API QSTAR PULSARi Hybrid LC/MS/MS (Applied Biosystems). The resulting data were used to search a custom *R. capsulatus* SB1003 Mascot database at the University of Victoria (Canada). Ion scores were calculated as $-10*\mathrm{Log}(P)$, where P is the probability that the observed match is a random event. Individual ion scores >26 indicate identity or a high degree of similarity ($p<0.05$). Protein scores are derived from ion scores as a non-probabilistic basis for ranking protein hits. All protein identification was carried out at the University of British Columbia (UBC) Centre for High-Throughput Biology (CHiBi).

Electron microscopy

CsCl-concentrated phage samples (10^{11} pfu ml^{-1}) were negatively stained with uranyl acetate, ammonium molybdate or phosphotungstic acid, and imaged in a Hitachi H7600 transmission electron microscope at the UBC Bioimaging Facility.

Sequence assembly and bioinformatics

Sequence assembly was performed using Geneious Pro 5.3.4 (Drummond et al., 2010) high sensitivity (gap free) assembly. Annotations from the SB1003 genome sequence were compared to automated glimmer 3.02 (Delcher et al., 2007) and genemark (Besemer and Borodovsky, 2005) annotations, and a and manual annotation based primarily on start codon preference, putative ribosome binding sites and ORF length. Any predicted ORF was then compared to existing CDS databases by blastp and PSI-BLAST (Altschul et al., 1990, 1997; Wheeler et al., 2007), and the start positions optimized by alignment with close homologues, where relevant. A search for convincing frameshift-encoded protein products was carried out using Frameshift Finder (Moon et al., 2004). The RcapMu sequence was submitted to GenBank (gb id: JN190960).

Supplementary materials related to this article can be found online at doi:10.1016/j.virol.2011.09.028.

Acknowledgments

We thank A. Chan and C. Suttle for advice, and assistance with microscopy. PCMF, ED and JTB thank the Canadian Institutes of Health Research for funding, and JTB thanks the Canada Council for the Arts for a Killam Research Fellowship. APH was supported by a SGS fellowship from Memorial University of Newfoundland and a PGS from the Natural Sciences and Engineering Research Council of Canada (NSERC). ASL thanks NSERC and the Newfoundland and Labrador Research & Development Corporation for funding. This research was funded in part by the Gordon and Betty Moore Foundation through a grant to the Broad Institute. Samples G9056 and G2263 were sequenced at the Broad Institute.

References

Abeliovich, A., Kaplan, S., 1974. Bacteriophages of *Rhodopseudomonas spheroides*: isolation and characterization of a *Rhodopseudomonas spheroides* bacteriophage. J. Virol. 13, 1392–1399.
Akhverdian, V.Z., Khrenova, E.A., Bogush, V.G., Gerasimova, T.V., Kirsanov, N.B., 1984. Wide distribution of transposable phages in natural *Pseudomonas aeruginosa* populations. Genetika 20, 1612–1619.
Allet, B., 1978. Nucleotide sequences at the ends of bacteriophage Mu DNA. Nature 274, 553–558.
Allet, B., 1979. Mu insertion duplicates a 5 base pair sequence at the host inserted site. Cell 16, 123–129.
Allison, H.E., 2007. Stx-phages: drivers and mediators of the evolution of STEC and STEC-like pathogens. Future Microbiol. 2, 165–174.
Allison, R.G., Chaconas, G., 1992. Role of the A protein-binding sites in the *in vitro* transposition of mu DNA. A complex circuit of interactions involving the mu ends and the transpositional enhancer. J. Biol. Chem. 267, 19963–19970.
Allison, H.E., Sergeant, M.J., James, C.E., Saunders, J.R., Smith, D.L., Sharp, R.J., Marks, T.S., McCarthy, A.J., 2003. Immunity profiles of wild-type and recombinant Shiga-like toxin-encoding bacteriophages and characterization of novel double lysogens. Infect. Immun. 71, 3409–3418.
Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410.
Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J.H., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.
Arbiol, C., Comeau, A.M., Kutateladze, M., Adamia, R., Krisch, H.M., 2010. Mobile regulatory cassettes mediate modular shuffling in T4-type phage genomes. Genome Biol. Evol. 2, 140–152.
Au, T.K., Agrawal, P., Harshey, R.M., 2006. Chromosomal integration mechanism of infecting mu virion DNA. J. Bacteriol. 188, 1829–1834.
Baudoux, A.-C., Noordeloos, A.A.M., Veldhuis, M.J.W., Brussaard, C.P.D., 2006. Virally induced mortality of *Phaeocystis globosa* during two spring blooms in temperate coastal waters. Aquat. Microb. Ecol. 44, 207–217.
Beatty, J.T., Gest, H., 1981. Generation of succinyl-coenzyme A in photosynthetic bacteria. Arch. Microbiol. 129, 335–340.
Besemer, J., Borodovsky, M., 2005. GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses. Nucleic Acids Res. 33, W451–W454.
Botstein, D., 1980. A theory of modular evolution for bacteriophages. Ann. N. Y. Acad. Sci. 354, 484–490.
Boyer, M., Haurat, J., Samain, S., Segurens, B., Gavory, F., Gonzalez, V., Mavingui, P., Rohr, R., Bally, R., Wisniewski-Dye, F., 2008. Bacteriophage prevalence in the genus *Azospirillum* and analysis of the first genome sequence of an *Azospirillum brasilense* integrative phage. Appl. Environ. Microbiol. 74, 861–874.
Braid, M.D., Silhavy, J.L., Kitts, C.L., Cano, R.J., Howe, M.M., 2004. Complete genomic sequence of bacteriophage B3, a Mu-Like phage of *Pseudomonas aeruginosa*. J. Bacteriol. 186, 6560–6574.
Bratbak, G., Wilson, W., Heldal, M., 1996. Viral control of *Emiliania huxleyi* blooms? J. Mar. Syst. 9, 75–81.
Brøndsted, L., Østergaard, S., Pedersen, M., Hammer, K., Vogensen, F.K., 2001. Analysis of the complete DNA sequence of the temperate bacteriophage TP901-1: evolution, structure, and genome organization of Lactococcal bacteriophages. Virology 283, 93–109.
Bukhari, A.I., Taylor, A.L., 1975. Influence of insertions on packaging of host sequences covalently linked to bacteriophage Mu DNA. Proc. Natl. Acad. Sci. U. S. A. 72, 4399–4403.

Bukhari, A.I., Zisper, D., 1972. Random insertion of Mu-1 DNA within a single gene. Nat. New Biol. 236, 240–243.
Chaconas, G., Harshey, R.M., 2002. Transposition of phage Mu DNA. In: Craig, N.L., Craigie, R., Gellert, M., Lambowitz, A.M. (Eds.), Mobile DNA II. ASM Press, Washington, DC, pp. 384–402.
Chaconas, G., Harshey, R.M., Sarvetnick, N., Bukhari, A.I., 1981. Predominant end-products of prophage Mu DNA transposition during the lytic cycle are replicon fusions. J. Mol. Biol. 150, 341–359.
Chaconas, G., Lavoie, B.D., Watson, M.A., 1996. DNA transposition: jumping gene machine, some assembly required. Curr. Biol. 6, 817–820.
Chopin, A., Bolotin, A., Sorokin, A., Ehrlich, S.D., Chopin, M., 2001. Analysis of six prophages in *Lactococcus lactis* IL1403: different genetic structure of temperate and virulent phage populations. Nucleic Acids Res. 29, 644–651.
Clokie, M.R.J., Kropinski, A.M., 2009. Bacteriophages Methods and Protocols. Humana, Totowa, N.J.
Coleman, D.C., Sullivan, D.J., Russel, R.J., Arbuthnott, J.P., Carey, B.F., Pomeroy, H.M., 1989. *Staphylococcus aureus* bacteriophages mediating the simultaneous lysogenic conversion of β-lysin, staphylokinase and enterotoxin A: molecular mechanism of triple conversion. J. Gen. Microbiol. 135, 1679–1697.
Coros, C.J., Chaconas, G., 2001. Effect of mutations in the Mu-host junction region on transpososome assembly. J. Mol. Biol. 310, 299–309.
Craig, N.L., 1995. Unity in transposition reactions. Science 270, 253–254.
Craigie, R., Mizuuchi, M., Mizuuchi, K., 1984. Site-specific recognition of the bacteriophage Mu ends by the Mu A protein. Cell 39, 387–394.
Daniell, E., Kohne, D.E., Abelson, J., 1975. Characterization of the inhomogeneous DNA in virions of bacteriophage Mu by DNA reannealing kinetics. J. Virol. 15, 739–743.
Delcher, A.L., Bratke, K.A., Powers, E.C., Salzberg, S.L., 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. Bioinformatics 23, 673–679.
Donohue, T.J., Chory, J., Goldsand, T.E., Lynn, S.P., Kaplan, S., 1985. Structure and physical map of *Rhodopseudomonas sphaeroides* bacteriophage RS1 DNA. J. Virol. 55, 147–157.
Drummond, A.J., Ashton, B., Buxton, S., Cheung, M., Cooper, A., Heled, J., Kearse, M., Moir, R., Stones-Havas, S., Sturrock, S., Thierer, T., Wilson, A., 2010. Geneious v5.3. Available from http://www.geneious.com.
Duchrow, M., Giffhorn, F., 1987. Physical map of the *Rhodobacter sphaeroides* bacteriophage phi RsG1 genome and location of the prophage on the host chromosome. J. Bacteriol. 169, 4410–4414.
Favinger, J., Stadtwald, R., Gest, H., 1989. *Rhodospirillum centenum*, sp. nov., a thermotolerant cyst-forming anoxygenic photosynthetic bacterium. Antonie Van Leeuwenhoek 55, 291–296.
Fazekas de St Groth, S., Webster, R.G., Datyner, A., 1963. Two new staining procedures for quantitative estimation of proteins on electrophoretic strips. Biochim. Biophys. Acta 71, 377–391.
Friedman, A.M., Long, S.R., Brown, S.E., Buikema, W.J., Ausubel, F.M., 1982. Construction of a broad host range cosmid cloning vector and its use in the genetic analysis of *Rhizobium* mutants. Gene 18, 289–296.
Ge, J., Harshey, R.M., 2008. Congruence of in vivo and in vitro insertion patterns in hot *E. coli* gene targets of transposable element Mu: opposing roles of MuB in target capture and integration. J. Mol. Biol. 380, 598–607.
Ge, J., Lou, Z., Harshey, R.M., 2010. Immunity of replicating Mu to self-integration: a novel mechanism employing MuB protein. Mob. DNA 1, 8.
Ge, J., Lou, Z., Cui, H., Shang, L., Harshey, R.M., 2011. Analysis of phage Mu DNA transposition by whole-genome *Escherichia coli* tiling arrays reveals a complex relationship to distribution of target selection protein B, transcription and chromosome architectural elements. J. Biosci. 36, 587–601.
Haapa-Paananen, S., Rita, H., Savilahti, H., 2002. DNA transposition of bacteriophage Mu. A quantitative analysis of target site selection in vitro. J. Biol. Chem. 277, 2843–2851.
Harshey, R.M., Getzoff, E.D., Baldwin, D.L., Miller, J.L., Chaconas, G., 1985. Primary structure of phage mu transposase: homology to mu repressor. Proc. Natl. Acad. Sci. U. S. A. 82, 7676–7680.
Harwood, C.S., Gibson, J., 1986. Uptake of benzoate by *Rhodopseudomonas palustris* grown anaerobically in light. J. Bacteriol. 165, 504–509.
Hatfull, G.F., Jacobs-Sera, D., Lawrence, J.G., Pope, W.H., Russell, D.A., Ko, C.C., Weber, R.J., Patel, M.C., Germane, K.L., Edgar, R.H., Hoyte, N.N., Bowman, C.A., Tantoco, A.T., Paladin, E.C., Myers, M.S., Smith, A.L., Grace, M.S., Pham, T.T., O'Brien, M.B., Vogelsberger, A.M., Hryckowian, A.J., Wynalek, J.L., Donis-Keller, H., Bogel, M.W., Peebles, C.L., Cresawn, S.G., Hendrix, R.W., 2010. Comparative genomic analysis of 60 Mycobacteriophage genomes: genome clustering, gene acquisition, and gene size. J. Mol. Biol. 397, 119–143.
Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwinn, M.L., Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Umayam, L., Gill, S.R., Nelson, K.E., Read, T.D., Tettelin, H., Richardson, D., Ermolaeva, M.D., Vamathevan, J., Bass, S., Qin, H., Dragoi, I., Sellers, P., McDonald, L., Utterback, T., Fleishmann, R.D., Nierman, W.C., White, O., Salzberg, S.L., Smith, H.O., Colwell, R.R., Mekalanos, J.J., Venter, J.C., Fraser, C.M., 2000. DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. Nature 406, 477–483.
Heinemann, B., Howard, A.J., 1964. Induction of lambda-bacteriophage in *Escherichia coli* as a screening test for potential antitumor agents. Appl. Microbiol. 12, 234–239.
Hendrix, R.W., 1983. Lambda II. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
Hendrix, R.W., 1988. Tail length determination in double-stranded DNA bacteriophages. Curr. Top. Microbiol. Immunol. 136, 21–29.
Hendrix, R.W., Smith, M.C., Burns, R.N., Ford, M.E., Hatfull, G.F., 1999. Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. Proc. Natl. Acad. Sci. U. S. A. 96, 2192–2197.
Howe, M.M., Bade, E.G., 1975. Molecular biology of bacteriophage mu. Science 190, 624–632.
Journet, L., Agrain, C., Broz, P., Cornelis, G.R., 2003. The needle length of bacterial injectisomes is determined by a molecular ruler. Science 302, 1757–1760.
Kahmann, R., Kamp, D., 1979. Nucleotide sequences of the attachment sites of bacteriophage Mu DNA. Nature 280, 247–250.
Karaolis, D.K.R., Somara, S., Maneval, D.R., Johnson, J.A., Kaper, J.B., 1999. A bacteriophage encoding a pathogenicity island, a type-IV pilus and a phage receptor in cholera bacteria. Nature 399, 375–379.
Katsura, I., 1987. Determination of bacteriophage lambda tail length by a protein ruler. Nature 327, 73–75.
Kinter, M., Sherman, N.E., 2000. Protein Sequencing and Identification Using Tandem Mass Spectrometry. John Wiley, New York; Chichester.
Laemmli, U.K., 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227, 680–685.
Lang, A.S., Beatty, J.T., 2007. Importance of widespread gene transfer agent genes in alpha-proteobacteria. Trends Microbiol. 15, 54–62.
Lavigne, R., Noben, J.P., Hertveldt, K., Ceyssens, P.J., Briers, Y., Dumont, D., Roucourt, B., Krylov, V.N., Mesyanzhinov, V.V., Robben, J., Volckaert, G., 2006. The structural proteome of *Pseudomonas aeruginosa* bacteriophage phiKMV. Microbiology 152, 529–534.
Lavoie, B.D., Chan, B.S., Allison, R.G., Chaconas, G., 1991. Structural aspects of a higher order nucleoprotein complex: induction of an altered DNA structure at the Mu-host junction of the Mu type 1 transpososome. EMBO J. 10, 3051–3059.
Lee, I., Harshey, R.M., 2001. Importance of the conserved CA dinucleotide at Mu termini. J. Mol. Biol. 314, 433–444.
Lee, I., Harshey, R.M., 2003. The conserved CA/TG motif at Mu termini: T specifies stable transpososome assembly. J. Mol. Biol. 330, 261–275.
Levin, M.E., Hendrix, R.W., Casjens, S.R., 1993. A programmed translational frameshift is required for the synthesis of a bacteriophage λ tail assembly protein. J. Mol. Biol. 234, 124–139.
Liebart, J.C., Ghelardini, P., Paolozzi, L., 1982. Conservative integration of bacteriophage Mu DNA into pBR322 plasmid. Proc. Natl. Acad. Sci. U. S. A. 79, 4362–4366.
Mackenzie, C., Choudhary, M., Larimer, F., Predki, P., Stilwagen, S., Armitage, J., Barber, R., Donohue, T., Hosler, J., Newman, J., Shapleigh, J., Sockett, R., Zeilstra-Ryalls, J., Kaplan, S., 2001. The home stretch, a first analysis of the nearly completed genome of *Rhodobacter sphaeroides* 2.4.1. Photosynth. Res. 70, 19–41.
Madigan, M.T., Gest, H., 1979. Growth of the photosynthetic bacterium *Rhodopseudomonas capsulata* chemoautotrophically in darkness with H_2 as the energy source. J. Bacteriol. 137, 524–530.
Mahillon, J., Chandler, M., 1998. Insertion sequences. Microbiol. Mol. Biol. Rev. 62, 725–774.
Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.-J., Chen, Z., Dewell, S.B., Du, L., Fierro, J.M., Gomes, X.V., Godwin, B.C., He, W., Helgesen, S., Ho, C.H., Irzyk, G.P., Jando, S.C., Alenquer, M.L.I., Jarvie, T.P., Jirage, K.B., Kim, J.-B., Knight, J.R., Lanza, J.R., Leamon, J.H., Lefkowitz, S.M., Lei, M., Li, J., Lohman, K.L., Lu, H., Makhijani, V.B., McDade, K.E., McKenna, M.P., Myers, E.W., Nickerson, E., Nobile, J.R., Plant, R., Puc, B.P., Ronan, M.T., Roth, G.T., Sarkis, G.J., Simons, J.F., Simpson, J.W., Srinivasan, M., Tartaro, K.R., Tomasz, A., Vogt, K.A., Volkmer, G.A., Wang, S.H., Wang, Y., Weiner, M.P., Yu, P., Begley, R.F., Rothberg, J.M., 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature 437, 376–380.
Marrs, B., 1974. Genetic recombination in *Rhodopseudomonas capsulata*. Proc. Natl. Acad. Sci. U. S. A. 71, 971–973.
Miller, J.L., Anderson, S.K., Fujita, D.J., Chaconas, G., Baldwin, D.L., Harshey, R.M., 1984. The nucleotide sequence of the B gene of bacteriophage Mu. Nucleic Acids Res. 12, 8627–8638.
Miller, E.S., Kutter, E., Mosig, G., Arisaka, F., Kunisawa, T., Ruger, W., 2003. Bacteriophage T4 genome. Microbiol. Mol. Biol. Rev. 67, 86–156.
Mizuuchi, K., Craigie, R., 1986. Mechanism of bacteriophage mu transposition. Annu. Rev. Genet. 20, 385–429.
Mizuuchi, M., Mizuuchi, K., 1993. Target site selection in transposition of phage Mu. Cold Spring Harb. Symp. Quant. Biol. 58, 515–523.
Mizuuchi, M., Baker, T.A., Mizuuchi, K., 1991. DNase protection analysis of the stable synaptic complexes involved in Mu transposition. Proc. Natl. Acad. Sci. U. S. A. 88, 9031–9035.
Moon, S., Byun, Y., Kim, H.-J., Jeong, S., Han, K., 2004. Predicting genes expressed via −1 and +1 frameshifts. Nucleic Acids Res. 32, 4884–4892.
Morgan, G.J., Hatfull, G.F., Casjens, S., Hendrix, R.W., 2002. Bacteriophage Mu genome sequence: analysis and comparison with Mu-like prophages in *Haemophilus*, *Neisseria* and *Deinococcus*. J. Mol. Biol. 317, 337–359.
Nandi, M., Lewis, G.P., 1970. Thin-layer acrylamide gel electrophoresis. J. Clin. Pathol. 23, 727–729.
Parkhill, J., Wren, B.W., Thomson, N.R., Titball, R.W., Holden, M.T.G., Prentice, M.B., Sebaihia, M., James, K.D., Churcher, C., Mungall, K.L., Baker, S., Basham, D., Bentley, S.D., Brooks, K., Cerdeno-Tarraga, A.M., Chillingworth, T., Cronin, A., Davies, R.M., Davis, P., Dougan, G., Feltwell, T., Hamlin, N., Holroyd, S., Jagels, K., Karlyshev, A.V., Leather, S., Moule, S., Oyston, P.C.F., Quail, M., Rutherford, K., Simmonds, M., Skelton, J., Stevens, K., Whitehead, S., Barrell, B.G., 2001. Genome sequence of *Yersinia pestis*, the causative agent of plague. Nature 413, 523–527.
Patel, A., Noble, R.T., Steele, J.A., Schwalbach, M.S., Hewson, I., Fuhrman, J.A., 2007. Virus and prokaryote enumeration from planktonic aquatic environments by epifluorescence microscopy with SYBR Green I. Nat. Protoc. 2, 269–276.
Pedulla, M.L., Ford, M.E., Houtz, J.M., Karthikeyan, T., Wadsworth, C., Lewis, J.A., Jacobs-Sera, D., Falbo, J., Gross, J., Pannunzio, N.R., Brucker, W., Kumar, V., Kandasamy, J., Keenan, L., Bardarov, S., Kriakov, J., Lawrence, J.G., Jacobs Jr., W.R., Hendrix, R.W., Hatfull, G.F., 2003. Origins of highly mosaic mycobacteriophage genomes. Cell 113, 171–182.
Poindexter, J.S., 1981. The Caulobacters: ubiquitous unusual bacteria. Microbiol. Rev. 45, 123–179.

Ptashne, M., 2004. A Genetic Switch: Phage Lambda Revisited, 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
Rapley, R., Chachaty, E., Saulnier, P., 2000. Isolating Chromosomal DNA from Bacteria, The Nucleic Acid Protocols Handbook. Humana Press, pp. 29–32.
Rohwer, F., Edwards, R., 2002. The Phage Proteomic Tree: a genome-based taxonomy for phage. J. Bacteriol. 184, 4529–4535.
Rooks, D.J., Yan, Y., McDonald, J.E., Woodward, M.J., McCarthy, A.J., Allison, H.E., 2010. Development and validation of a qPCR-based method for quantifying Shiga toxin-encoding and other lambdoid bacteriophages. Environ. Microbiol. 12, 1194–1204.
Saunders, J.R., Allison, H.E., James, C.E., McCarthy, A.J., Sharp, R., 2001. Phage-mediated transfer of virulence genes. J. Chem. Technol. Biotechnol. 76, 662–666.
Schmidt, L.S., Yen, H.C., Gest, H., 1974. Bioenergetic aspects of bacteriophage replication in the photosynthetic bacterium *Rhodopseudomonas capsulata*. Arch. Biochem. Biophys. 165, 229–239.
Schumann, W., 1979. Cloning and biological characterization of the immunity region of *Escherichia coli* phage Mu. Gene 5, 275–290.
Scolnik, P.A., Walker, M.A., Marrs, B.L., 1980. Biosynthesis of carotenoids derived from neurosporene in *Rhodopseudomonas capsulata*. J. Biol. Chem. 255, 2427–2432.
Slayter, H.S., Holloway, B.W., Hall, C.E., 1964. The structure of *Pseudomonas aeruginosa* phages B3, E79, and F116. J. Ultrastruct. Res. 11, 274–281.
Smith, D.L., Wareing, B.M., Fogg, P.C., Riley, L.M., Spencer, M., Cox, M.J., Saunders, J.R., McCarthy, A.J., Allison, H.E., 2007. Multilocus characterization scheme for shiga toxin-encoding bacteriophages. Appl. Environ. Microbiol. 73, 8032–8040.
Solioz, M., Marrs, B., 1977. The gene transfer agent of *Rhodopseudomonas capsulata*. Purification and characterization of its nucleic acid. Arch. Biochem. Biophys. 181, 300–307.
Strnad, H., Lapidus, A., Paces, J., Ulbrich, P., Vlcek, C., Paces, V., Haselkorn, R., 2010. Complete genome sequence of the photosynthetic purple nonsulfur bacterium *Rhodobacter capsulatus* SB 1003. J. Bacteriol. 192, 3545–3546.
Summer, E.J., Gonzalez, C.F., Carlisle, T., Mebane, L.M., Cass, A.M., Savva, C.G., LiPuma, J., Young, R., 2004. *Burkholderia cenocepacia* phage BcepMu and a family of Mu-like phages encoding potential pathogenesis factors. J. Mol. Biol. 340, 49–65.
Surette, M.G., Harkness, T., Chaconas, G., 1991. Stimulation of the Mu A protein-mediated strand cleavage reaction by the Mu B protein, and the requirement of DNA nicking for stable type 1 transpososome formation. In vitro transposition characteristics of mini-Mu plasmids carrying terminal base pair mutations. J. Biol. Chem. 266, 3118–3124.
Suttle, C.A., 2005. Viruses in the sea. Nature 437, 356–361.
Symonds, N., Toussaint, A., Van de Putte, A.P., Howe, M.M., 1987. Phage Mu. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
Tarutani, K., Nagasaki, K., Yamaguchi, M., 2000. Viral impacts on total abundance and clonal composition of the harmful bloom-forming phytoplankton *Heterosigma akashiwo*. Appl. Environ. Microbiol. 66, 4916–4920.
Taylor, A.L., 1963. Bacteriophage-induced mutation in *Escherichia coli*. Proc. Natl. Acad. Sci. 50, 1043–1051.
van Niel, C.B., 1944. The culture, general physiology, morphology, and classification of the non-sulfur purple and brown bacteria. Microbiol. Mol. Biol. Rev. 8, 1–118.
Vogel, J.L., Li, Z.J., Howe, M.M., Toussaint, A., Higgins, N.P., 1991. Temperature-sensitive mutations in the bacteriophage Mu c repressor locate a 63-amino-acid DNA-binding domain. J. Bacteriol. 173, 6568–6577.
Wall, J.D., Weaver, P.F., Gest, H., 1975. Gene transfer agents, bacteriophages, and bacteriocins of *Rhodopseudomonas capsulata*. Arch. Microbiol. 105, 217–224.
Wang, P.W., Chu, L., Guttman, D.S., 2004. Complete sequence and evolutionary genomic analysis of the *Pseudomonas aeruginosa* transposable bacteriophage D3112. J. Bacteriol. 186, 400–410.
Weaver, P.F., Wall, J.D., Gest, H., 1975. Characterization of *Rhodopseudomonas capsulata*. Arch. Microbiol. 105, 207–216.
Wheeler, D.L., Barrett, T., Benson, D.A., Bryant, S.H., Canese, K., Chetvernin, V., Church, D.M., DiCuccio, M., Edgar, R., Federhen, S., Geer, L.Y., Kapustin, Y., Khovayko, O., Landsman, D., Lipman, D.J., Madden, T.L., Maglott, D.R., Ostell, J., Miller, V., Pruitt, K.D., Schuler, G.D., Sequeira, E., Sherry, S.T., Sirotkin, K., Souvorov, A., Starchenko, G., Tatusov, R.L., Tatusova, T.A., Wagner, L., Yaschenko, E., 2007. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 35, D5–D12.
Yamaguchi, T., Hayashi, T., Takami, H., Nakasone, K., Ohnishi, M., Nakayama, K., Yamada, S., Komatsuzawa, H., Sugai, M., 2000. Phage conversion of exfoliative toxin A production in *Staphylococcus aureus*. Mol. Microbiol. 38, 694–705.
Yen, H.C., Marrs, B., 1976. Map of genes for carotenoid and bacteriochlorophyll biosynthesis in *Rhodopseudomonas capsulata*. J. Bacteriol. 126, 619–629.
Yen, H.C., Hu, N.T., Marrs, B.L., 1979. Characterization of the gene transfer agent made by an overproducer mutant of *Rhodopseudomonas capsulata*. J. Mol. Biol. 131, 157–168.
Zhao, Y., Wang, K., Jiao, N., Chen, F., 2009. Genome sequences of two novel phages infecting marine roseobacters. Environ. Microbiol. 11, 2055–2064.

Rhodobacter capsulatus Gene Transfer Agent (RcGTA) Activity Bioassays

Molecular Biology > DNA > Introduction of DNA into cells
Microbiology > Microbial genetics
Bacteria > Rhodobacter capsulatus > Nucleic acid isolation & anal.
Authors: Alexander P. Hynes and Andrew S. Lang, 1/20/2013, 534 views, 0 Q&A

[Abstract] RcGTA, a small phage-like particle produced by Rhodobacter capsulatus, was initially identified in culture filtrates as a DNase-resistant form of DNA transfer between R. capsulatus cells. This gene transfer assay has been used to identify RcGTA-producing strains and to help determine the roles of genes thought to be responsible for RcGTA production.

Materials and Reagents

1. GTA donor strain (See Notes 1)
2. Recipient strain (See Notes 1)
3. RCV broth (Beatty and Gest, 1981)
4. YPS broth (Wall et al., 1975)
5. YPS agar (2 plates per recipient-donor combination, one plate per donor, and one plate per recipient)
6. 0.22 µm filtered GTA Buffer (Solioz and Marrs, 1977)
7. 0.45 µm low protein binding (PVDF) syringe filters (e.g. Millipore catalogue # SLHV033RB)
8. 1 ml pipettes
9. 1.5 ml Microcentrifuge tubes
10. Plate spreader (Including 95% Ethanol + Flame)
11. 1 ml syringes
12. Test tubes for aerobic culturing (e.g. Fisherbrand catalogue # 14-961-30) w/caps
13. Culture tubes for anaerobic/photosynthetic culturing (e.g. Fisherbrand catalogue # 14-959-37A w/screw caps).
14. Polypropylene sterile culture tubes (e.g. Simport catalogue # T405-2A)
15. Optional (If transferring photosynthesis marker): Chambers/packs for anaerobic plate growth

Equipment

1. Shaking incubator
2. Microcentrifuge
3. Incandescent light-box or light incubator

Procedure

1. Preparation:
    1. 3 days prior to the assay, inoculate GTA donor strains in RCV broth and grow aerobically overnight at 35°C (200-250 RPM).
    2. 2 days prior, measure optical densities (OD) of overnight cultures of GTA donor strains and normalize them by dilution with RCV broth. It is simplest to dilute all ODs to match the lowest, as the actual OD is irrelevant so long as all cultures are at the same final density. Use a 1% v/v inoculum of normalized donor to inoculate YPS broth without antibiotics to grow anaerobically and photosynthetically over two days at 35°C without shaking, mixing occasionally by inverting the culture tubes. These culture tubes should be filled to the brim with YPS broth and sealed tightly, to create anaerobic conditions, and placed equidistant from the light source. Incandescent bulbs are better than fluorescent bulbs, but the heat generated from the bulbs must be dissipated (e.g. By having the culture tubes in a water tank) in order to maintain culturing temperature.
    3. 1 day prior, inoculate recipient strain in RCV broth and grow aerobically overnight at 35°C (200-250 RPM).
2. Assay:
    1. Pass donor strain cultures through a 0.45 μm filter, collecting the filtrates in polypropylene tubes.
    2. Centrifuge 1 ml of recipient cultures, decant the supernatant and re-suspend in an equal volume of GTA buffer.
    3. Mix the following in a polypropylene tube:
    -Filtrate Controls (for each Donor filtrate): 0.5 ml GTA buffer, 0.1 ml filtrate
    -Recipient Controls (for each Recipient Strain): 0.5 ml GTA buffer, 0.1 ml recipient cells
    -Experimental samples: 0.4 ml GTA buffer, 0.1 ml filtrate, 0.1 ml recipient cells
    4. Incubate tubes at 35°C for 1 hr with no shaking.
    5. Add 0.9 ml RCV broth to each tube.
    6. Incubate tubes at 35°C for 3 hrs with shaking at ~200 rpm.
    7. Transfer the 1.5 ml mixtures from the polypropylene tubes to individual 1.5 ml microcentrifuge tubes.
    8. Plate 150 μl of each filtrate + recipient mix on plates to represent 10% of the total. This is not necessary for the recipient and filtrate controls.
    9. Centrifuge all tubes, decant supernatant, and resuspend pellets in the small (~100 μl) volume that remains. Spread these resuspensions on plates to represent 90% and 100% for experimental and control assays, respectively. These plates should be selective for the transfer of the marker, or grown in conditions that select for a transferred marker.
    10. Grow for 2-3 days at 30-35°C.
    11. Count colonies. Determine ratios of colonies on experimental plates over the number of colonies found on a positive control (e.g. a wild type strain).

Notes:

1. One common bioassay employs monitoring the transfer of the puhA gene to puhA deletion mutant DW5 (Wong et al., 1996), with selection for transfer of the puhA gene being the ability to grow photosynthetically. This bioassay shows no spontaneous mutation background. Another bioassay is the

transfer of rifampicin resistance, which is a property of some strains of R. capsulatus, to rifampicin-sensitive strains such as the natural isolate strain B10 (Weaver et al., 1975). This bioassay has a low but detectable rate of spontaneous mutation background that must be accounted for in the data analysis.
2. It is essential to compare transfer of the same marker, as marker sizes can affect bioassay results (Hynes et al., 2012), presumably affecting both packaging and recombination rates.
3. Bioassay absolute numbers can vary greatly depending on growth state of donor and recipient cells, batch of media (e.g. the batch of Yeast Extract was used for YPS), so it is essential to compare ratios within one bioassay, and perform multiple independent replicates. When performing bioassays for or into strains/mutants with impaired growth or viability, it can help to normalize transfer rate to the number of viable cells (either donors or recipients, whichever is impaired) by performing viable cell counts alongside the bioassay.

Recipe

1. RCV Medium
   7.5 mM $(NH_4)_2SO_4$
   30 mM DL-malate, pH 6.8 with NaOH
   54 μM EDTA
   0.8 mM $MgSO_4$
   0.51 mM $CaCl_2$
   43 μM $FeSO_4$
   3 μM Thiamine-HCl
   9.5 μM $MnSO_4$
   45 μM H_3BO_3
   0.2 μM $Cu(NO_3)_2$
   0.83 μM $ZnSO_4$
   3 μM $NaMoO_4$
   4.5 mM KH_2PO_4
   5.1 mM K_2HPO_4
   -pH to 6.8
2. YPS Medium
   3.0 g/L Yeast Extract
   3.0 g/L Peptone
   2 mM $MgSO_4$
   2 mM $CaCl_2$
   -pH to 7.0 with NaOH or HCl, as needed
   -Add 1.5% Agar for solid media
3. GTA Buffer
   10mM Tris-HCl (pH 7.8) with NaOH
   1.0 mM $MgCl_2$
   1.0 mM $CaCl_2$
   1.0 mM NaCl
   500 μg/mL BSA (Fraction V)
   -Filter sterilize with 0.22 μm filter.

References

1. Beatty, J. T. and H. Gest (1981). Generation of succinyl-coenzyme A in photosynthetic bacteria. *Archives of Microbiology* 129(5): 335-340.
2. Wall, J. D., P. F. Weaver and H. Gest (1975). Genetic transfer of nitrogenase-hydrogenase activity in Rhodopseudomonas capsulata. *Nature* 258(5536): 630-631.
3. Solioz, M. and B. Marrs (1977). The gene transfer agent of Rhodopseudomonas capsulata. Purification and characterization of its nucleic acid. *Arch Biochem Biophys* 181(1): 300-307.
4. Wong, D. K., W. J. Collins, A. Harmer, T. G. Lilburn and J. T. Beatty (1996). Directed mutagenesis of the Rhodobacter capsulatus puhA gene and orf 214: pleiotropic effects on photosynthetic reaction center and light-harvesting 1 complexes. *J Bacteriol* 178(8): 2334-2342.
5. Weaver, P. F., J. D. Wall and H. Gest (1975). Characterization of Rhodopseudomonas capsulata. *Arch Microbiol* 105(3): 207-216.
6. *Hynes, A. P., R. G. Mercer, D. E. Watton, C. B. Buckley and A. S. Lang (2012). DNA packaging bias and differential expression of gene transfer agent genes within a population during production and release of the Rhodobacter capsulatus gene transfer agent, RcGTA. *Mol Microbiol* 85(2): 314-3255.

* How to cite this protocol: please cite Reference 6.

Questions and Comments:

We highly appreciate your feedback. Please login to post your questions/comments. Your questions will be directed to the authors of the protocol. The authors will be requested to answer your questions at their earliest convenience. Once your questions are answered, you will be informed using the email address that you register with bio-protocol.

Login | Register

Other protocols by Alexander

P. Hynes
Other
protocols
by
Andrew
S. Lang