

Bacteriophage replication modules

Christoph Weigel & Harald Seitz

Max-Planck-Institut für molekulare Genetik, Berlin, Germany

Correspondence: Christoph Weigel,
Technical University Berlin, Faculty III, Inst. for
Biotechnology, Sekr. MA 5-11, Straße des 17.
Juni 136, D-10623 Berlin, Germany.
Tel.: +49 30 84131614;
e-mail: weigel@molgen.mpg.de

Received 22 March 2005; revised 6 September
2005; accepted 8 November 2005.
First published online 21 March 2006.

doi:10.1111/j.1574-6976.2006.00015.x

Editor: Ramon Diaz Orejas

Keywords

DNA polymerase; helicase; helicase loader;
initiator; primase; replication origin; replication
fork.

Contents

Introduction

Replication mechanisms

- Initiation by nicking: 'rolling circle'-type DNA replication
- Initiation by melting: theta (θ)-type DNA replication
- Initiation at the ends of linear DNA: protein-primed DNA replication
- Initiation of DNA replication by transcription
- Recombination-dependent DNA replication
- Replication restart

Bacteriophage replication modules

- Phages encoding initiator proteins
- 'Initiator-solo' replication modules
- 'Initiator-helicase loader' replication modules
- 'Initiator-helicase' replication modules
- 'Initiator-helicase loader-helicase' replication modules
- Conclusions for 'Phages encoding initiator proteins'
- Replication module exchange among bacteriophages

Abstract

Bacteriophages (prokaryotic viruses) are favourite model systems to study DNA replication in prokaryotes, and provide examples for every theoretically possible replication mechanism. In addition, the elucidation of the intricate interplay of phage-encoded replication factors with 'host' factors has always advanced the understanding of DNA replication in general. Here we review bacteriophage replication based on the long-standing observation that in most known phage genomes the replication genes are arranged as modules. This allows us to discuss established model systems – ϕ 1/fd, ϕ X174, P2, P4, λ , SPP1, N15, ϕ 29, T7 and T4 – along with those numerous phages that have been sequenced but not studied experimentally. The review of bacteriophage replication mechanisms and modules is accompanied by a compendium of replication origins and replication/recombination proteins (available as supplementary material online).

- ϕ P4 α -type helicase-primase encoding replication modules
- Phages-encoding DNA polymerases
- The phage T4-type replication module
- The phage T7-type replication module
- The phage D29-type replication module
- The replication modules of the phages K, Bx21 and T5
- The phage ϕ 29-type replication module
- Replication modules of phages replicating by RCR
- Phage replicons lacking replication protein genes

Evolutionary considerations

- The different types of phage-encoded helicases
- Phage-encoded homologues of the *E.coli* DnaB helicase
- Chromosomally encoded homologues of phage helicase loaders

Perspectives

Acknowledgements, Supplementary material, References

Introduction

Chromosomes, plasmids and bacteriophages (bacterial viruses) represent three of the four genetic elements that are, permanently or transiently, present in a prokaryotic cell. They are entities – *replicons* in the terminology of the 'replicon model' – whose key regulatory elements for

propagation are a replication origin – the *replicator* – and an *initiator*, in most cases a protein (Jacob *et al.*, 1963). Transposable elements, the fourth type, are covalently linked to one of the other genetic elements and therefore not considered as replicons.

In spite of this clear-cut definition of a replicon, a satisfactory definition of 'bacteriophages' remains problematic. Mu, to give just one example, is a typical temperate phage of its host *Escherichia coli* for all but one stage of its 'life-cycle': its genome propagates as transposon (Nakai *et al.*, 2001; Morgan *et al.*, 2002). Taxonomy-oriented biology tends to classify phages, plasmids and transposons as 'mobile genetic elements' to account for such overlaps (Toussaint & Merlin, 2002) and resorts to chromosomal genes for the definition of bacterial species. However, bacterial chromosomes contain a variety of intact, defective, and degraded prophages, i.e. phages in the integrated state and their remnants. Hence the taxonomy-oriented approach cannot cut the Gordian knot: prophage genes account for a notable portion of the genes in bacterial chromosomes (Casjens, 2003; Canchaya *et al.*, 2003) and are largely responsible for phenotypic variations among the strains of bacterial species, including such important traits as pathogenicity (Banks *et al.*, 2002; Bruessow *et al.*, 2004). Replication research has traditionally focused on chromosomes, plasmids and phages as experimental systems without caring much about their exact classification, and we follow this route here.

Bacterial chromosomes are fully autonomous genetic elements, i.e. they carry all genes whose products are required for their replication. Also bacteriophage and plasmid replicons multiply as intact entities but their autonomy is limited due to their partial or complete dependence on factors encoded by chromosomal genes for reproduction. The elucidation of the intricate interplay of phage- or plasmid-encoded replication factors with chromosomally encoded 'host' factors has always advanced the understanding of both systems. Chromosome replication has been studied for decades in *E. coli* and *Bacillus subtilis* – more recently also in *Streptomyces lividans* (reviewed in Mariani, 1996; Messer & Weigel, 1996; Kogoma, 1997; Moriya *et al.*, 1999; Messer, 2002; Messer & Zakrewska-Czerwinska, 2002). A comprehensive review of plasmid replication has been presented by del Solar and colleagues (del Solar *et al.*, 1998). Reviews of the replication of phages with DNA genomes focus on favourite model systems: λ (Campbell, 1994; Taylor & Wegrzyn, 1995), T4 (Mosig, 1998), ϕ 29 (Meijer *et al.*, 2001), f1/fd (Horiuchi, 1997), SPP1 (Alonso *et al.*, 2005) and T7 (Richardson, 1983). Several recent research papers and reviews cover the field of RNA phages, which will not be discussed here (Bollback & Huelsenbeck, 2001; Chetverin, 2004; Makeyev & Grimes, 2004; Mindich, 2004; Poranen & Tuma, 2004).

We will discuss bacteriophage replication based on a long-standing observation: genes encoding replication functions tend to be located close to each other in many phage genomes, resulting in what has been termed a 'replication module'. The replication module of a phage can be determined experimentally by dissecting cognate (phage-encoded)

proteins from host proteins recruited for replication. This approach leads to the elucidation of its replication mechanism. Alternatively, one can determine the replication module by subjecting a phage genome to a thorough homology search in the available databases. A replication mechanism cannot be reliably predicted by this approach unless the replication genes of the phage genome under study are similar to those of one of the established model systems.

In addition to discussing the replication modules of established model systems we will explore whether the concept of 'replication modules' can lead to a better understanding of the replication of those numerous phages that have been sequenced but not studied experimentally. We will thus evaluate which of the well-studied phages are valid model systems and which should be regarded as unique cases.

Taking into account that most readers prefer the printed version of a paper for studying a topic from a broader viewpoint and the online version for selective searches, we have decided to present the different aspects of phage replication in two parts:

- (1) Bacteriophage replication mechanisms and replication modules are discussed in this part of the review.
- (2) A compendium of phage replication origins and phage replication/recombination proteins is presented in the supplementary material available online.

Note: all parts of the 'compendium of origins and proteins' referred to in the following are marked with 'COM' to encourage and facilitate navigation between the two parts. In addition, all numbers of Sections, Tables and Figures of the compendium are 'tagged' with the prefix 'C'.

Replication mechanisms

The structure of double-stranded B-DNA 'immediately suggests a possible copying mechanism for the genetic material' (Watson & Crick, 1953), and three possible molecular mechanisms for the initiation of this copying process, DNA replication: (1) 'nicking', i.e. the breakage of the covalent phosphodiester bond between two neighbouring bases on one strand; (2) 'melting', i.e. the localised disruption of the hydrogen bonds that tether together the two complementary DNA strands; and (3) melting of the terminal hydrogen bonds of linear double-stranded DNA (dsDNA) molecules. All three initiation mechanisms generate single-stranded regions as templates for the synthesis of complementary daughter strands, resulting in what is known as semi-conservative DNA replication since the hallmark experiments of Meselson & Stahl (1958).

These three possible initiation mechanisms have been studied in detail for circular and linear dsDNA phage replicons. The replication of linear dsDNA molecules seems straightforward, irrespective of the initiation mechanism.

However, all known (replicative) DNA polymerases synthesise DNA exclusively in 5'→3' direction and require a primer, mostly a short oligo-ribonucleotide. Thus, the 5'-ends of both strands cannot be replicated, resulting in a loss of genetic information during successive rounds of replication. Research on phage replication has revealed four different molecular mechanisms to overcome this problem:

(1) *Bacillus subtilis* phage ϕ 29 uses a specialised protein as 'portable primer' that remains covalently attached to both genome ends after completion of replication.

(2) *Escherichia coli* phage T7 uses direct terminal repeats that are regenerated by processing of genome concatemers during the packaging of monomeric genomes into phage heads.

(3) *Escherichia coli* phage N15 employs a specialised enzyme, protelomerase, to (re)generate the covalently closed ends of the linear double-stranded prophage genome after replication of a circular intermediate.

(4) Many phage genomes that enter the cell in a linear form are converted to a circular form prior to replication or, alternatively, integration into the host chromosome as prophage.

The genetic information is faithfully conserved during replication of covalently closed circular dsDNA molecules but the helical nature of DNA creates a topological problem: the progeny molecules are intertwined and require a recombination step for resolution. A comparable problem arises from cutting a Möbius ribbon with $2n$ twists along the middle. *Escherichia coli* phage λ replication is initiated by the 'melting' mechanism early after infection, and proceeds by simultaneous synthesis of both daughter strands, thus creating catenated progeny molecules that are resolved by host topoisomerases. Later during infection, the circular progeny molecules are converted, probably by recombination proteins, to structures that allow the continuous synthesis of (linear) concatemeric phage DNA. These concatemers are finally processed by the phage packaging apparatus to yield monomeric linear genomes. A complete understanding of the replication mechanisms of circular replicons includes knowledge of cognate recombination processes, therefore. *Escherichia coli* phage P2 avoids the topological problem by replicating each parent DNA strand of its circular(ised) genome separately, involving a single-stranded replication intermediate not catenated with the dsDNA molecule.

The genomes of phages P2 (33.6 kb), T7 (39.9 kb), N15 (46.4 kb) and λ (48.5 kb) are fairly similar in size but their replication follows different routes, as outlined briefly above. Only the experimental and/or computational search for replication origin structures and genes encoding replication proteins, i.e. the elucidation of the 'replication module', together with a comparison with the known mechanisms discussed below can lead to a prediction of the likely

replication mechanism operating in a phage replicon under study.

Initiation by nicking: 'rolling circle'-type DNA replication

Replicons that propagate by the 'rolling circle' mode of DNA replication (RCR) include bacterial phages and plasmids with circular dsDNA genomes (Khan, 1997, 2000). In all cases that have been studied experimentally, a DNA-bound initiator protein nicks one strand of the dsDNA molecules. The 5'-end of the disrupted DNA strand becomes covalently linked to a specific tyrosine residue of the initiator, while the free 3'-OH end is elongated by a replisome. After one round of replication, the strand-transfer reaction is reversed, liberating a single-stranded from a double-stranded molecule in a reaction that does not require recombination. DNA replication is completed by the synthesis of the complementary strand of the single-stranded molecule. For the phages with single-stranded circular DNA genomes, e.g. ϕ X174, M13 and fd, this 'completion-step' is the conversion of the viral or (+)-strand to the double-stranded 'replicative form' (RF).

It was shown by Horiuchi and co-workers for the filamentous *E. coli* phage fd that the nicking reaction is preceded by a localised DNA unwinding around the nick-site (Higashitani *et al.*, 1994) (COM section C2.1.1.). This points to the reason why RCR has so far only been found for circular dsDNA replicons: protein-induced DNA unwinding is apparently only possible with negatively supercoiled, i.e. undertwisted, DNA. Small linear dsDNA molecules cannot be undertwisted because both strands rotate freely around each other, with one exception: linear dsDNA with covalently closed ends as in the ϕ N15 prophage (see next section). The situation is more complex with large linear DNA molecules, e.g. bacterial chromosomes, where an intricate interplay of topoisomerases, gyrases and a number of nucleoid-associated proteins (among others: HU, H-NS) creates transient 'topological domains' of undertwisted DNA that are anchored to cell structures (Worcel & Burgi, 1972; Postow *et al.*, 2004).

Phage fd: Replication in the 'rolling circle' mode was for some time considered specific for small plasmid and phage replicons. As we know today, large conjugative plasmids use RCR coupled to a specific secretion system for DNA transfer to recipient cells (Llosa *et al.*, 2002) and also a number of phages with mid-sized genomes (~30 kb) replicate via RCR (see below). We first discuss the successive steps of RCR of the filamentous *E. coli* phage fd as an example (Fig. 1), which largely resembles RCR of the isometric phage ϕ X174: Step 1. The single-stranded circular (+)-strand DNA of ϕ fd that enters the host cell is covered by host single-strand binding protein (SSB) except for the single-strand origin

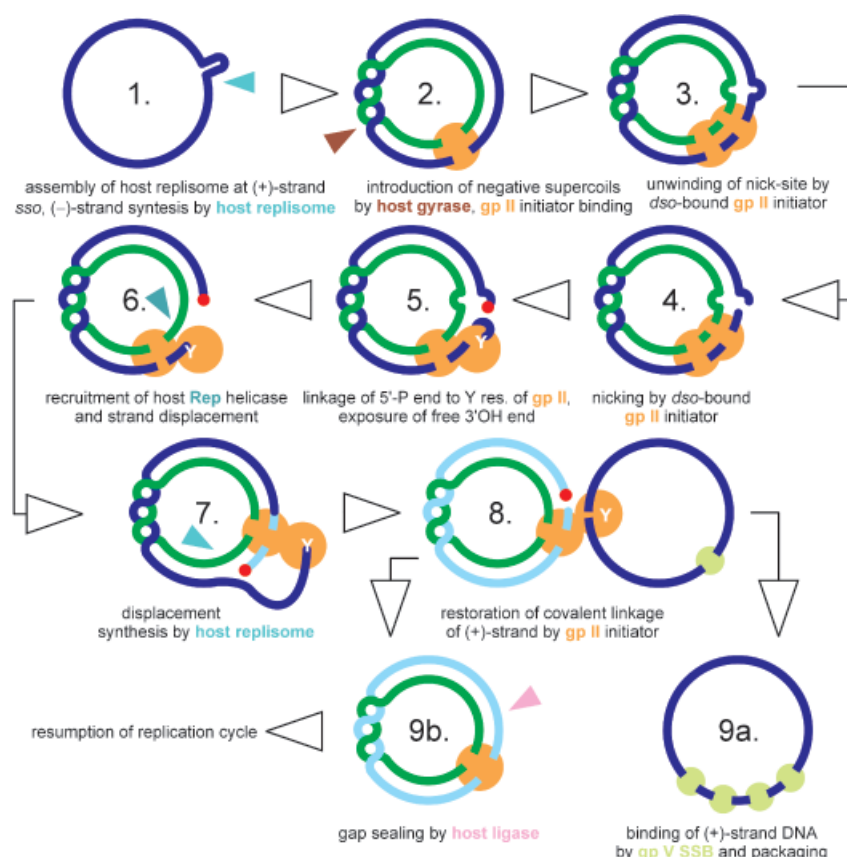


Fig. 1. The mechanism of 'rolling circle' replication: phage fd. Successive steps are indicated by open arrows/arrowheads plus numbering (see text for details). Complementary DNA strands are shown as parallel lines; twisted lines indicate (negative) supercoiling of the DNA molecule. Dark blue/green: parental DNA strands; light blue/green: daughter DNA strands; yellow: RNA primers. Red dots indicate free 3'-hydroxyls used for DNA synthesis. Phage-encoded proteins are shown as coloured circles, host factors as coloured triangles. The colour code used for individual proteins corresponds to the coloured protein names in the text.

(*ss*). The *ss* is recognised by the host (σ^{70})RNA polymerase although its structure does not resemble a typical promoter (Kaguni & Kornberg, 1982) (COM section C2.1.). RNA polymerase synthesises a short untranslated transcript (20 nt; Higashitani *et al.*, 1993) that is elongated for (–)-strand synthesis by host DNA Pol III holoenzyme. At this step, ϕ X174 requires assembly of the restart primosome (formerly PriA primosome, see 'Replication restart' section) at the *ss*, which directs primer synthesis by DnaG primase and (–)-strand synthesis by DNA Pol III holoenzyme. A third variation exists for complementary-strand synthesis of phage G4: DnaG primase alone was found to be required for efficient primer synthesis *in vitro* (Bouché *et al.*, 1978; Stayton & Kornberg, 1983; Stayton *et al.*, 1983; Hiasa *et al.*, 1989). In all three systems, primer removal is performed by PolA and gap sealing by DNA ligase.

Step 2. The conversion of the viral (+)-strand DNA into the covalently closed circular RF is completed upon introduction of negative supercoils by the host gyrase. The phage initiator gpII binds to the double-strand origin (*dso*) in the linear or (relaxed) circular forms ('complex I') but origin unwinding and nicking by 'complex II' requires the negatively supercoiled form (COM section C2.1.1.).

Step 3. The *dso* of ϕ fd is located close to the *ss*, and consists of two structural elements: the so-called nick-site and, adjacent, the binding sites for gpII. gpII bound to its binding sites on a supercoiled substrate induces a conformation of the nucleoprotein complex that results in the localised unwinding of 7 bp encompassing the nick-site as a prerequisite for the nicking reaction (COM section C2.1.+C3.1.1.). Steps 4 and 5. The nicking reaction is performed by an appropriately positioned gpII protomer within the oligomeric complex II. Nicking occurs simultaneously with the (transient) covalent linkage of the (+)-strand 5'-end to a specific tyrosine residue of gpII. Only for clarity, this reaction is shown as two separate steps in Fig. 1.

Step 6. The unwound region serves as entry site for the host Rep helicase, which dimerises upon DNA-binding and unwinds the duplex in 3' \rightarrow 5' direction; it is not known whether gpII attracts Rep by direct physical interaction (Hours & Denhardt, 1979; Takahashi *et al.*, 1979; Meyer & Geider, 1982; Chao & Lohman, 1991).

Step 7. Strand-displacement synthesis starting from the free 3'-OH end is performed by host DNA Pol III holoenzyme (Meyer & Geider, 1982). Replication intermediates at this stage appear in the electron microscope as dsDNA circles with attached single-stranded loops, thus the term 'rolling

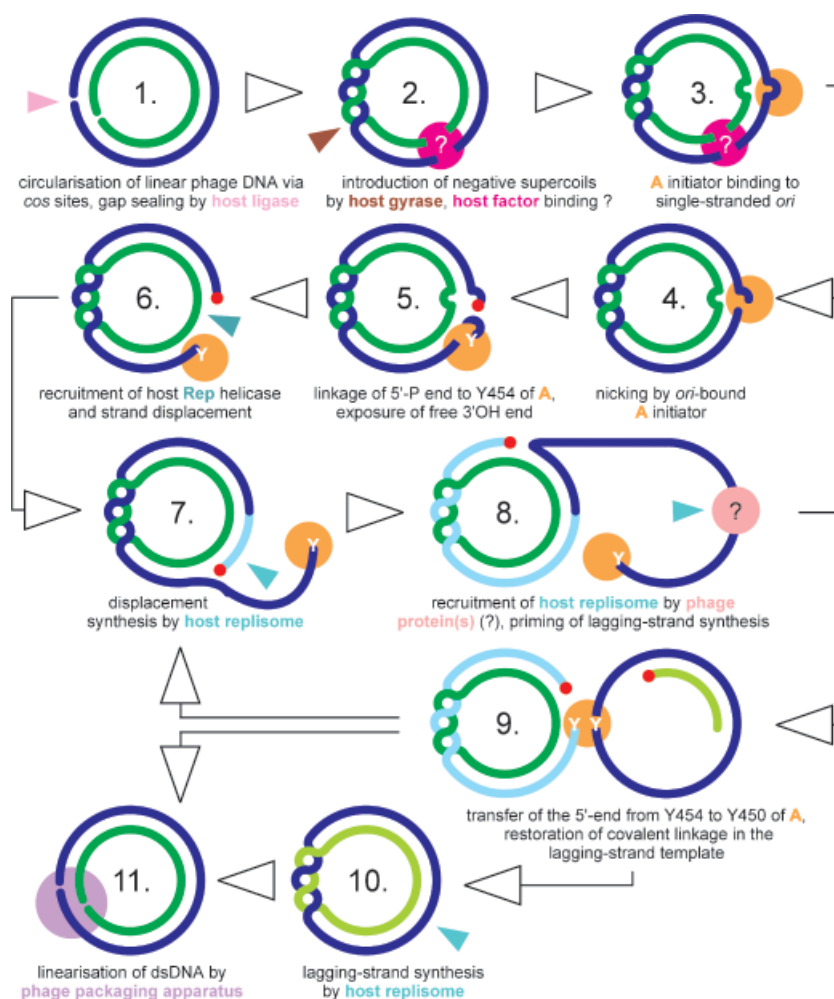


Fig. 2. The mechanism of 'rolling circle' replication (RCR): phage P2. Successive steps are indicated by open arrows/arrowheads plus numbering (see text for details). Complementary DNA strands are shown as parallel lines; twisted lines indicate (negative) supercoiling of the DNA molecule. Dark blue/green: parental DNA strands; light blue/green: daughter DNA strands; yellow: RNA primers. Red dots indicate free 3'-hydroxyls used for DNA synthesis. Phage-encoded proteins are shown as coloured circles, host factors as coloured triangles. The colour code used for individual proteins corresponds to the coloured protein names in the text.

circle' DNA replication (Gilbert & Dressler, 1968). Depending on the conditions used for sample preparation, the single-stranded part can appear as a single-stranded tail rather than as circle, giving the molecules a shape resembling a Greek sigma (σ) (Allison *et al.*, 1977). RCR and σ -type DNA replication (σ DR) are frequently used synonymously in the literature (see also Kornberg & Baker, 1992). However, the lengths of the 'tails' derived from RCR never exceed the contour lengths of the circular parent molecules. In this review, we reserve the term σ DR for a replication mechanism that also produces 'tailed' molecules but involves a recombination step for initiation; an example is the switch from θ DR to σ DR at later stages during λ replication (see below). 'Tails' derived from σ DR are always (partially) double-stranded owing to coupled leading- and lagging-strand DNA synthesis and are, as genome concatemers, usually much longer than the circular parent molecules to which they are attached.

Step 8. When the replisome reaches the initial nick-site, the displaced single strand is liberated from the double-stranded circle by a reversal of step 5: the 5'-end transiently bound

to the tyrosine residue of gpII is transferred back to the free 3'-OH end.

Step 9a. The single-stranded replication intermediate of step 8 represents the (+)-strand, i.e. the phage genome, which associates with gpV SSB already during liberation from the double-stranded replication intermediate. The rapid association of the (+)-strand DNA with gpV SSB and, subsequently, with other coat proteins (gpVII, gpIX) (Feng *et al.*, 1997) prevents strand-switching or coupled leading- and lagging-strand DNA synthesis by the host replisome.

Step 9b. The RF is restored from the double-stranded replication intermediate of step 8 by the action of host DNA ligase and gyrase, allowing the resumption of the replication cycle with step 2. It has not been firmly proven that the gpII initiator remains bound to the *dso* throughout all steps of the replication cycle, as tentatively shown in Fig. 1. Horiuchi and co-workers have shown that a small stretch upstream of the nick-site is important for the termination reaction (step 8), suggesting that gpII may remain bound to the *dso* throughout DNA synthesis (Dotto *et al.*, 1984).

Similar to other initiation systems, however, one or more binding sites for gpII could be occupied throughout the replication cycle without allowing the formation of complex II. This would localise the displaced single strand, via its linkage to gpII, close to the position required for the reversal of the strand-transfer reaction (step 5).

The molecular mechanism responsible for the balanced synthesis of ϕ fd RF and (+)-strand is not known precisely but gpX may be involved. During *in vitro* replication of ϕ X174, the C protein is involved in DNA packaging and, by binding to the initiation complex formed by A protein, promotes multiple rounds of (+)-strand synthesis while preventing the accumulation of RF (Aoyama & Hayashi, 1986).

Phage P2: As a second example for RCR, we discuss the replication of *E. coli* phage P2 (Fig. 2):

Step 1. ϕ P2 DNA enters the host cell as linear dsDNA with 19 bp complementary 5'-overhangs (*cos*). Following intramolecular circularisation, the gaps are sealed by the host DNA ligase.

Step 2. The conversion of the circular ϕ P2 DNA into the replication-proficient form is completed upon introduction of negative supercoils by the host gyrase. Binding of the phage initiator A to dsDNA *in vitro* has not yet been demonstrated, and although it is clear that no other ϕ P2 protein is involved, additional host factor(s) that could support unwinding have not yet been identified (Liu & Haggård-Liungquist, 1994).

Step 3. The initiator protein A binds to the partially single-stranded nick-site, located within the ϕ P2 *ori* (COM section 3.1.).

Steps 4 and 5. The nicking reaction is performed by an appropriately positioned A monomer (COM section C3.1.1.). Nicking occurs simultaneously with the (transient) covalent linkage of the (+)-strand 5'-end to a specific tyrosine residue of A, most likely Y454 (Odegrip & Haggård-Liungquist, 2001). Only for clarity, this reaction is shown as two separate steps in Fig. 2.

Step 6. The unwound region serves as entry site for the host Rep helicase and, subsequently, host replisomal proteins.

Step 7. Strand-displacement synthesis starting from the free 3'-OH end is performed by host DNA Pol III holoenzyme.

Step 8. The displaced single strand serves as lagging-strand template already during ongoing strand-displacement synthesis on the double-stranded phage DNA. Haggård-Liungquist and co-workers were able to show that single-strand DNA (ssDNA) replication intermediates of ϕ P2 minichromosomes are converted to dsDNA solely by host proteins, i.e. DnaB helicase and its helicase loader DnaC, DnaG primase and DNA Pol III holoenzyme. By contrast, P2 phage ssDNA intermediates require the ϕ P2 B protein as helicase loader, and probably other phage-encoded proteins

in addition to host replisomal proteins (COM section 3.2.) (Liu *et al.*, 1993). It is not known which segment of the ϕ P2 genome serves as 'single-strand origin', and also the molecular mechanism of primosome recruitment is presently not known.

Step 9. When the replisome reaches the initial nick-site, the completely displaced single strand is liberated from the double-stranded circle by a reversal of step 5: the 5'-end transiently bound to Y454 of A is transferred back to the free 3'-OH end; the second 'active tyrosine' Y450 is apparently instrumental in binding of the free 5'-end (Odegrip & Haggård-Liungquist, 2001). It is not known how A protein is kept 'in place' to perform the strand-transfer reaction. After gap sealing by the host DNA ligase and adjustment of negative superhelicity by gyrase, the closed circular dsDNA may undergo a new round of replication starting with step 2 or serve as substrate for DNA packaging into phage heads. It is not known precisely which molecular mechanism triggers the choice between ongoing replication and packaging, but it may be the availability of packaging proteins.

Step 10. After completion of 'lagging-strand' synthesis by the host replisome, the resulting double-stranded progeny molecule is processed by PolA, DNA ligase and gyrase. The closed circular dsDNA may undergo a new round of replication starting with step 2 or serve as substrate for DNA packaging into phage heads.

Step 11. The ϕ P2 terminase consists of the P and M subunits. M was proposed to contribute the endonuclease activity required for the linearisation of the circular replication intermediates at the *cos*-sites during packaging of the phage DNA (Linderoth *et al.*, 1991).

Although not all steps are yet known in necessary detail, ϕ P2 replication demonstrates that (1) RCR is not confined to replicons with small genomes (< 10 kb), and (2) RCR is easily adopted for the replication of ss and dsDNA genomes. A highly specific feature of phage replication in the 'rolling circle' mode is the involvement of the Rep helicase during strand-displacement synthesis. Also plasmid propagation by RCR depends on Rep helicase – or PcrA, its homologue in Gram-positive bacteria (Petit *et al.*, 1998). Rep and PcrA belong to the superfamily I helicases and are involved in recombination processes rather than in chromosome replication of their hosts (COM section C3.3.) (Petit & Ehrlich, 2002). In fact, the inability of plasmids or phages to replicate in a *rep/pcrA* mutant host may be taken as an indication that these replicons propagate via RCR.

Initiation by melting: theta (θ)-type DNA replication

Replicons that propagate by the theta (θ)-mode of DNA replication (θ DR) include bacterial chromosomes, plasmids

and phages. In the known cases, a specialised protein (the initiator) binds to its recognition site(s) adjacent to an AT-rich region within a replication origin. The nucleoprotein complex formed by the initiator protein and the origin DNA results in 'melting', i.e. partial unwinding of the AT-rich region. The unwound region serves as entry site for the primosomal proteins (helicase loader/helicase+primase) and, subsequently, the replisomal proteins (DNA polymerase+accessory proteins). The initiator performs the function of a primosomal protein by recruiting the helicase to the unwound region. Because the primosomal proteins, helicase+primase in particular, promote the assembly of the replisome, the initiator protein is also often called 'replisome organiser'. We use the term ' θ DR' for initiation by a specialised initiator protein, and use the term 'tDR' for initiation by transcription, which also involves 'DNA melting' for primosome and replisome assembly on the (locally) single-stranded template (see 'Initiation of DNA replication by transcription' section).

The chromosome of *E. coli* replicates in the θ -mode (Cairns, 1963). It contains a unique replication origin, *oriC* (Meijer *et al.*, 1979; Sugimoto *et al.*, 1979), and DnaA is the initiator (Kohiyama *et al.*, 1966; Chakraborty *et al.*, 1982; Fuller *et al.*, 1984). DnaA is responsible for origin 'melting' (Fuller *et al.*, 1984; Roth & Messer, 1995; Krause *et al.*, 1997) and directs the replicative helicase DnaB to the unwound region (Marszalek & Kaguni, 1994; Weigel & Seitz, 2002). Subsequently, DnaB recruits the DnaG primase, and both proteins together promote the assembly of the DNA Pol III holoenzyme for bidirectional coupled leading- and lagging-strand DNA synthesis (Fuller *et al.*, 1981; Kaguni *et al.*, 1982; Kaguni & Kornberg, 1984). Present research efforts concentrate on a better understanding of the regulation of DnaA activity in the cell cycle (Speck & Messer, 2001; Katayama, 2001; Su'etsugu *et al.*, 2004) and of the molecular details of the multiple protein-protein (Zechner *et al.*, 1992; Weigel *et al.*, 1999; Chang & Mariani, 2000; Seitz *et al.*, 2000) and protein-DNA interactions (Fujikawa *et al.*, 2003; McGarry *et al.*, 2004). Most sequenced bacterial chromosomes contain detectable *oriC* structures (Mackiewicz *et al.*, 2004) and encode *dnaA* gene(s), and we may assume that DnaA/*oriC*-dependent θ DR is the 'normal' route for chromosome replication in bacteria (Messer, 2002). Initiation of chromosome replication has not been studied in detail in the (very) few bacterial species that either lack a *dnaA* homologue, e.g. *Wigglesworthia glossinidia*, or where disruption of the *dnaA* gene does not produce a phenotype (Richter *et al.*, 1998).

Studies of plasmid and phage replication revealed that the 'ABC-pathway' of *E. coli* is just one possibility for θ DR. For example, the broad host-range plasmid RSF1010 (IncQ) encodes a set of replication proteins that are entirely unrelated to the *E. coli* proteins but perform analogous

functions during θ DR: the initiator RepC is responsible for *oriV* unwinding, and RepA is the cognate helicase whose action is followed by the RepB' primase. Following these initiation steps, the host DNA Pol III holoenzyme synthesises the progeny molecules, probably by strand displacement (Rawlings & Tietze, 2001).

A number of θ -replicating plasmids use dual initiators, and their replication origins contain DnaA binding sites in addition to binding sites for the cognate initiator. In these systems, DnaA is either used in support for the unwinding step (pSC101 (Datta *et al.*, 1999), F plasmid (Kawasaki *et al.*, 1996), ϕ P1 prophage plasmid (Park & Chattoraj, 2001), R6K Lu *et al.*, 1998), for the recruitment of the replicative helicase DnaB, or for both functions (RK2/RP4, Konieczny *et al.*, 1997; reviewed in Messer, 2002). The intricate host-dependent interplay of the TrfA initiator, DnaA and DnaB for replication of RK2 is discussed in more detail in the 'Evolutionary considerations' section. A particularly intriguing finding was the observation that a mutation in the *repA* initiator gene of the *Pseudomonas* sp. plasmid pPS10 resulted in a protein that extended the host range of the plasmid, allowing its replication in *E. coli* through interaction of RepA with DnaA (Giraldo & Fernandez-Tresguerres, 2004).

Contrary to the expectation of Campbell & Botstein (1983), phages that encode DnaA homologues have not yet been found, and also, contrary to the mentioned plasmid systems, phage replicons that propagate via θ DR and engage the DnaA protein of their host for replication are not known. An exception is the regulation of p_R -mediated 'transcriptional activation' of λ replication by DnaA, but in this case DnaA acts as transcription factor rather than as primosomal protein (Glinkowska *et al.*, 2003).

Phage λ : *Escherichia coli* phage λ was the first phage replicon for which replication in the θ -mode could be demonstrated in all details (reviewed in Taylor & Wegrzyn, 1995) (Fig. 3):

Step 1. λ DNA enters the host cell as linear dsDNA with 12 bp complementary 5'-overhangs (*cos*). Following intramolecular circularisation, the gaps are sealed by host DNA ligase.

Step 2. The conversion of the linear λ DNA into the replication-proficient form is completed upon introduction of negative supercoils by the host gyrase. The phage initiator O binds to the replication origin (*ori λ*) in the linear or (relaxed) circular forms but origin unwinding requires the negatively supercoiled form (Schnos *et al.*, 1988).

Step 3. *ori λ* is located in the middle of the O gene (COM section C2.2.). O protein bound to its binding sites on a supercoiled substrate *in vitro* induces a conformation of the nucleoprotein complex that results in origin unwinding. Origin unwinding *in vivo*, however, requires 'transcriptional activation', i.e. transcription driven by the p_R promoter

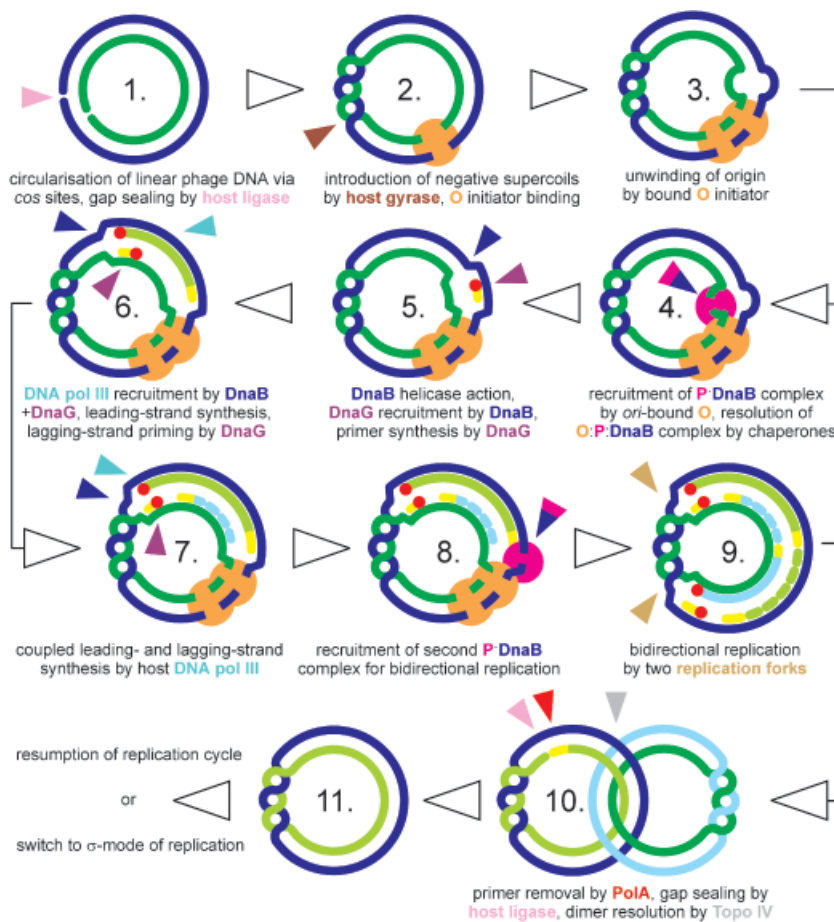


Fig. 3. The mechanism of replication in the theta (θ)-mode for circular DNA: phage λ . Successive steps are indicated by open arrows/arrowheads plus numbering (see text for details). Complementary DNA strands are shown as parallel lines; twisted lines indicate (negative) supercoiling of the DNA molecule. Dark blue/green, parental DNA strands; light blue/green, daughter DNA strands; yellow, RNA primers. Red dots indicate free 3'-hydroxyls used for DNA synthesis. Phage-encoded proteins are shown as coloured circles, host factors as coloured triangles. The colour code used for individual proteins corresponds to the coloured protein names in the text.

upstream of the *O* gene and extending to a region downstream of *ori λ* (COM sections C2.2. + C3.1.2.) (Hase *et al.*, 1989).

Step 4. λ P forms a tight 3:6 complex with the host replicative helicase DnaB. The P·DnaB complex is recruited to the unwound region by direct interaction of P with *ori λ* -bound O. The initial binding of P·DnaB to the unwound region involves the cryptic ssDNA-binding property of P. DnaB is liberated from the P·DnaB complex by the action of the host chaperones DnaJ and DnaK (COM section C3.2.).

Step 5. DnaB helicase action widens the single-stranded region. DnaB recruits the host DnaG primase for priming of leading-strand synthesis.

Step 6. The DnaB·DnaG primosome recruits host DNA Pol III holoenzyme for leading-strand synthesis. Strand-switching of DnaG results in priming for lagging-strand synthesis.

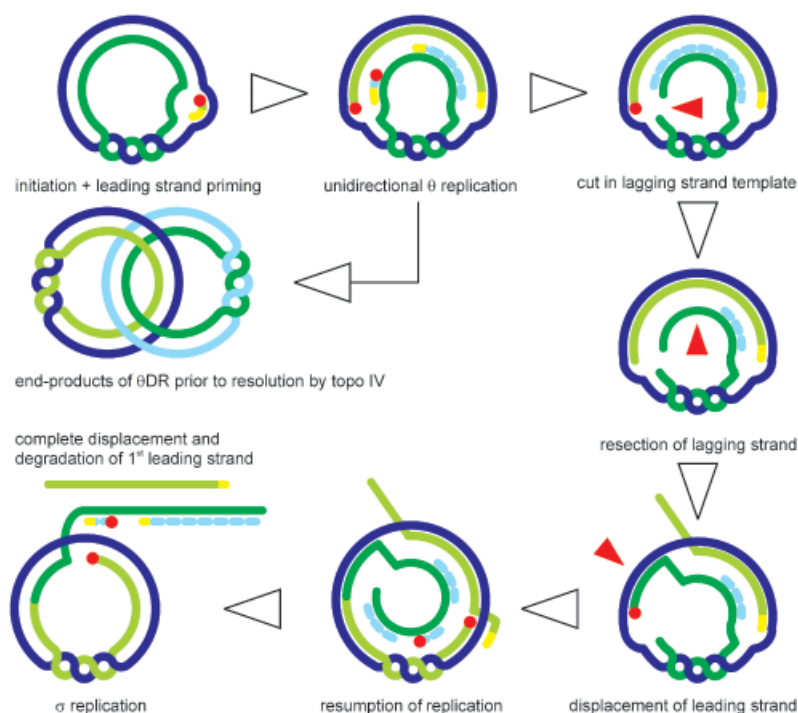
Step 7. Unidirectional coupled leading- and lagging-strand synthesis by the host replisome. Early after infection, λ replication proceeds with step 8. At later stages, prior to the switch to σ DR (see below), unidirectional replication becomes prevalent, probably by a combination of (1) cessation of transcriptional activation from the *p_R* promoter, and (2) depletion of host DnaB.

Step 8. *ori λ* -bound O protein recruits a second P·DnaB complex to the replication bubble, but now with the opposite orientation. As before, DnaB helicase is released from the O·P·DnaB complex by the action of chaperones, and DnaB recruits DnaG primase for priming of DNA synthesis.

Step 9. Two replication forks are engaged in coupled leading- and lagging-strand synthesis, resulting in bidirectional replication away from the origin. Replication intermediates at this step appear as structures resembling the Greek letter theta (θ) with (mostly) double-stranded loops in the electron microscope, thus the name. It has been shown that O monomers are removed from *ori λ* by the action of ClpX/ClpP protease (Zylicz *et al.*, 1998), but some may remain bound to binding sites in *ori λ* throughout the replication cycle (for clarity, binding of O to sites in *ori λ* is not shown for this step in Fig. 3).

Step 10. After completion of DNA synthesis the RNA primers are removed by the 5' \rightarrow 3'-exonuclease activity of PolA, the gaps simultaneously filled by the DNA polymerase activity of PolA, the gaps sealed by DNA ligase and negative superhelicity introduced by gyrase. Dimer resolution is performed by host topoisomerase IV (Espeli & Marians, 2004).

Fig. 4. Changing the replication mode: switch from θ type to σ -type DNA replication. Dark colour, parental strands; light colour, daughter strands. Dark blue, leading-strand template; dark green, lagging-strand template; light green, leading strand; light blue, lagging strand with Okazaki fragments indicated by dashed line. RNA primers synthesised by DnaG primase are shown in yellow. For clarity, single-strand binding protein, helicase, primase, polymerase and single-strand annealing protein (SAP) are not shown. Instead, filled red arrows indicate site(s) of enzymatic action(s). Red dots indicate free 3'-hydroxyls used for DNA synthesis. Subsequent steps are indicated by open arrows.



Step 11. Both progeny molecules may resume θ DR starting from step 2 or switch to the σ -mode of replication (see below). The packaging of λ DNA into phage (pro)heads requires genome concatemers for processing, at the *cos*-sites, by the small subunit of λ terminase; the closed-circular monomeric progeny molecules from step 11 are not appropriate substrates (Collins & Hohn, 1978; Feiss *et al.*, 1985; Smith & Feiss, 1993; Sippy & Feiss, 2004).

The O (initiator) and P (helicase loader) proteins are the only phage-encoded proteins required for λ replication in the θ -mode; all other replication factors and accessory factors, gyrase, DNA ligase, PolA and Topo IV, are recruited from the host. The general outline of the 'replication scheme' for λ is virtually identical for the large family 'lambdoid' phages (discussed in detail in 'Bacteriophage replication modules' section). Note that here we use the fuzzy term 'lambdoid' exclusively to characterise phages with replication modules showing similarity to the λ replication module.

Phage SPP1: Recombination is a prerequisite for the propagation of ϕ SPP1, and could play a role for λ and other phages that replicate via θ DR (see above). Hints come from the following observations: (1) the linear dsDNA genomes of lambdoid phage circularise upon entry into the host cell, (2) linear head-to-tail genome concatemers are required for the packaging of replicated dsDNA into phage capsids (Taylor & Wegrzyn, 1995), and (3) the lambdoid phages replicate via the θ -mode, which leads to circular progeny

molecules. Takahashi found σ -like structures, i.e. head-to-tail genome concatemers protruding from a circular parent molecule, by electron microscopy of λ replication intermediates (Takahashi, 1975). Interestingly, σ -structures but no replication intermediates typical for θ DR have been observed in ϕ SPP1-infected *B. subtilis* cells, although initiation of ϕ SPP1 replication occurs by the unidirectional θ -mode (Missich *et al.*, 1997). In addition, mutations in the ϕ SPP1 genes 34.1 (exonuclease) and gene 35 (SAP) have been shown to result in a replication arrest phenotype (Burger & Trautner, 1978; Weise *et al.*, 1994). For propagation of ϕ SPP1, the switch from θ DR to σ DR seems therefore to depend on cognate recombination functions.

Recombination steps that could lead to a switch from θ DR to σ DR are shown schematically for a circular model replicon in Fig. 4. The switch is initiated by an interruption of replication fork progression. Theoretically, it does not matter whether a progressing fork encounters a strand break, or whether a fork is halted by a 'road-block', i.e. a nucleoprotein complex. Next, a scission is set in the fork region by an endonuclease, thus creating a double-strand break (DSB). As shown in Fig. 4, the lagging strand is partially degraded by the action of a 5' \rightarrow 3' exonuclease. The exposed 3'-OH end of the lagging-strand template is covered by a single-strand annealing protein (SAP) and annealed to the leading-strand template, thus displacing parts of the leading strand. Depending on the size of the single-stranded gap in the leading-strand template, the annealing reaction can be performed either by host RecA as

strand invasion, or by the phage-encoded SAP as strand annealing (Stahl *et al.*, 1997). The resumption of replication results in the complete displacement of the first leading strand. The successive displacement of the linearised lagging-strand template from the circular leading-strand template circle gives the resulting structure, the typical σ shape observed in the electron microscope. Ongoing coupled leading- and lagging-strand DNA synthesis produces the genome concatemers required for DNA packaging into virion capsids.

Only parts of this model are supported by experimental evidence. Proteins with 5' \rightarrow 3' exonuclease activity and with single-strand annealing activity are encoded by λ , ϕ SPP1 and by the Rac prophage in the *E. coli* K12 genome, and have been studied to detail (COM section C3.6.2.). Experimental evidence for a phage-encoded endonuclease responsible for the initial scission in one of the parental strands of the circular replicon is elusive (Taylor & Wegrzyn, 1995). More recent results indicate, however, that replication fork arrest may lead to fork regression and thus the formation of a Holliday junction by annealing of the newly synthesised strands (Seigneur *et al.*, 1998; McGlynn & Lloyd, 2000). Accordingly, (pro)phage-encoded Holliday junction resolvases like RusA or λ Rap could be the long-sought candidate nucleases responsible for the initial scission (COM section C3.6.2.).

In Fig. 4, we show the cut in the parental lagging-strand template because a cut in the leading-strand template would create an unusually orientated structure: leading-strand synthesis would use the linear 'tail' template, and the synthesis of Okazaki fragments would be directed by the circular template. To our knowledge, such structures have never been found in experiments. It is completely unknown whether and how the scission in one of the parental strands is specifically directed to the parental lagging-strand template, in order to create the known σ -structure. We show the displacement of the leading strand by the lagging-strand template after exonucleolytic resection of the lagging strand. Whether this kind of 'self-invasion' *in cis* occurs during phage replication *in vivo* is uncertain, but it demonstrates that a second molecule providing the invading strand is not necessary for a switch from θ DR to σ DR – at least in theory.

Although λ encodes a pair of exonuclease/SAP recombination proteins (Red α /Red β) like ϕ SPP1, their contribution to the switch from θ DR to σ DR remains controversial. Bidirectional replication is altered to unidirectional prior to the switch, probably as response to (1) a cessation of DnaA-dependent transcriptional activation (Baranska *et al.*, 2001), and (2) a decrease in available DnaB helicase. Zylicz *et al.* (1998) showed that a decrease in availability of the host ClpX/ClpP protease promotes unidirectional replication. Echols and co-workers proposed that, following one round

of unidirectional replication, the 5'-end of the leading strand is displaced by the arriving replication fork (Dodson *et al.*, 1986). However, also a DSB in one circular genome could be trimmed to obtain the linear substrate required for invasion of another circular genome *in trans* mediated by host recombination proteins. Although λ Rap may be instrumental for creating a DSB, the major source of progeny molecules with a DSB are probably those that were cut at the λ cos site but failed to be packaged into virion capsids because they were not part of concatemers (Stahl *et al.*, 1985).

In comparison with ϕ SPP1, the dependence on the cognate recombination proteins seems to be less strict for λ replication. As a consequence, the presence of recombination genes in the genomes of other lambdoid phages can be taken as an indication but not as proof of their (essential) role during replication. In each case, only experiments can help to find decisive answers.

A noteworthy difference among the 'lambdoid' phages exists with respect to the form in which the (linear) phage genomes are delivered into the host cell and packaged to phage heads, following replication. All λ monomers end with identical 12 bp 5'-overhangs (*cos*-sites) generated by λ terminase, and are thus packaged from identical sites along a concatemer. By contrast, the 'headful packaging' mechanism of ϕ SPP1 driven by the hetero-oligomeric G1P · G2P (1 : 10) terminase generates, using *pac*-site(s), a heterogeneous population of terminally redundant and partially circularly permuted DNA molecules with 2 bp overhangs (Chai *et al.*, 1992). The individual steps of replication of both phages are virtually identical and not influenced by this 'logistic' difference, however.

Phage N15: As mentioned above, the 'melting' step during θ DR and RCR requires: (1) interaction among origin-bound initiator protomers and (2) negative superhelicity of the origin DNA. All established *in vitro* assays for the 'melting' step use, besides the purified initiator, closed-circular DNA carrying the replication origin under study (COM section C3.1.). For practical reasons, closed-circular DNA is purified from cells: it is negatively supercoiled, and the degree of superhelicity is easily controlled. Hence it has become a general notion that RCR and θ DR are initiation mechanisms for circular DNA replicons. The replication of the linear prophage of *E. coli* phage N15 demonstrates that it is negative superhelicity and not circularity that is important for initiation in the θ -mode (reviewed in Ravin, 2003; Fig. 5):

Steps 1 and 2. Like λ DNA, ϕ N15 DNA enters the host cell as linear dsDNA with 5'-overhangs (*cos*). Following intramolecular circularisation, the gaps are sealed by host DNA ligase. Host gyrase action provides the negative superhelicity required for origin unwinding.

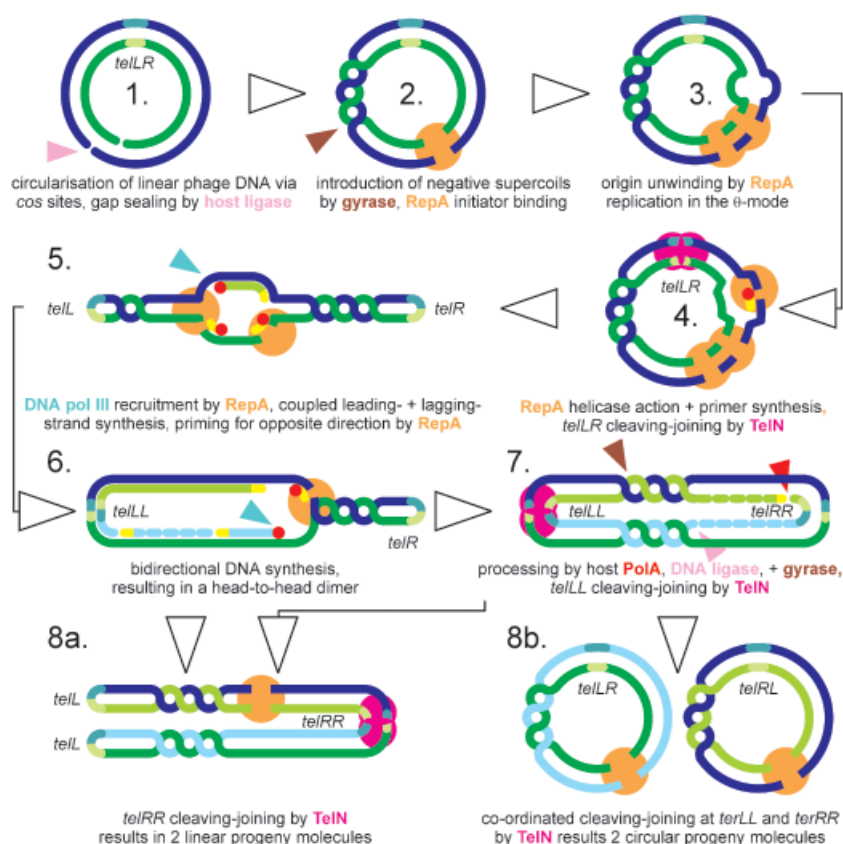


Fig. 5. The mechanism of replication in the theta (θ) mode for linear DNA: phage N15. Successive steps are indicated by open arrows/arrowheads plus numbering (see text for details). Complementary DNA strands are shown as parallel lines; twisted lines indicate (negative) supercoiling of the DNA molecule. Dark blue/green, parental DNA strands; light blue/green, daughter DNA strands; yellow, RNA primers. Red dots indicate free 3'-hydroxyls used for DNA synthesis. Phage-encoded proteins are shown as coloured circles, host factors as coloured triangles. The colour code used for individual proteins corresponds to the coloured protein names in the text. The recognition sites for the *TelN* protomerase are indicated as *telLR*, *telL*, *telLL*, *telR*, *telRR* and *telRL* (see text for details).

Step 3. Initiation is performed by the multifunctional *RepA* protein at the replication origin residing in the 3' part of the *repA* gene.

Step 4. Origin unwinding and priming is performed by *RepA*, in analogy to the action of α protein during initiation of replication of ϕ P4 (COM section C3.1.2.) (Mardanov *et al.*, 2004). In addition, the homologous *RepA* initiator of ϕ PY54 has very recently been shown to possess primase and helicase activity (Ziegelin *et al.*, 2005). The phage-encoded protomerase *TelN* recognises the palindromic *telLR* sequence as a dimer and, by a cleaving-joining reaction, generates the *telL* and *telR* 'hairpin' ends (Deneke *et al.*, 2000, 2002). The covalently closed ends preserve the negative superhelicity in the linearised molecule. If *TelN* activity is suppressed at this step, ϕ N15 replication follows the ' λ -scheme' for θ DR and the subsequent switch to σ DR, which leads to the formation of genome concatemers that are processed by the ϕ N15 packaging apparatus (Ravin, 2003).

Step 5. Replication of ϕ N15 proceeds bidirectionally and with coupled leading- and lagging-strand synthesis (Ravin *et al.*, 2003).

Step 6. Replication encompassing the *telL* site (re)generates a palindromic *telLL* site that is substrate for cleaving-joining by *TelN*. If *TelN* already acts on the replication intermediate

at this step, a Y-structure with *telL* hairpin ends is formed (step 8a). Otherwise, replication ends with the formation of a circular head-to-head dimer.

Step 7. The head-to-head dimer is processed by *PolA*, *DNA ligase* and *gyrase* for completion of the DNA synthesis, just as in the case of λ replication (Fig. 3; step 10).

Step 8a. Successive cleaving-joining by *TelN* at *telLL* and *telRR* generates two linear progeny molecules with hairpin ends. This form of the ϕ N15 prophage is fully competent to enter new rounds of replication beginning with step 4.

Step 8b. Co-ordinated cleaving-joining by *TelN* at *telLL* and *telRR* generates two circular progeny molecules with *telRL* and *telRL* sites, respectively. This form of ϕ N15 follows the ' λ -scheme' for θ DR and the subsequent switch to σ DR, which leads to the formation of genome concatemers that are processed by the ϕ N15 packaging apparatus.

Efficient propagation of linear minichromosomes carrying the *repA* gene together with the *telN* gene and a *telLR* site has been demonstrated (Ravin *et al.*, 2001). It is presently not known, however, by which mechanism the activity of *TelN* is suppressed to allow the linear prophage to enter θ DR with circular molecules to provide – after the switch to σ DR – the genome concatemers for DNA packaging during the lytic cycle.

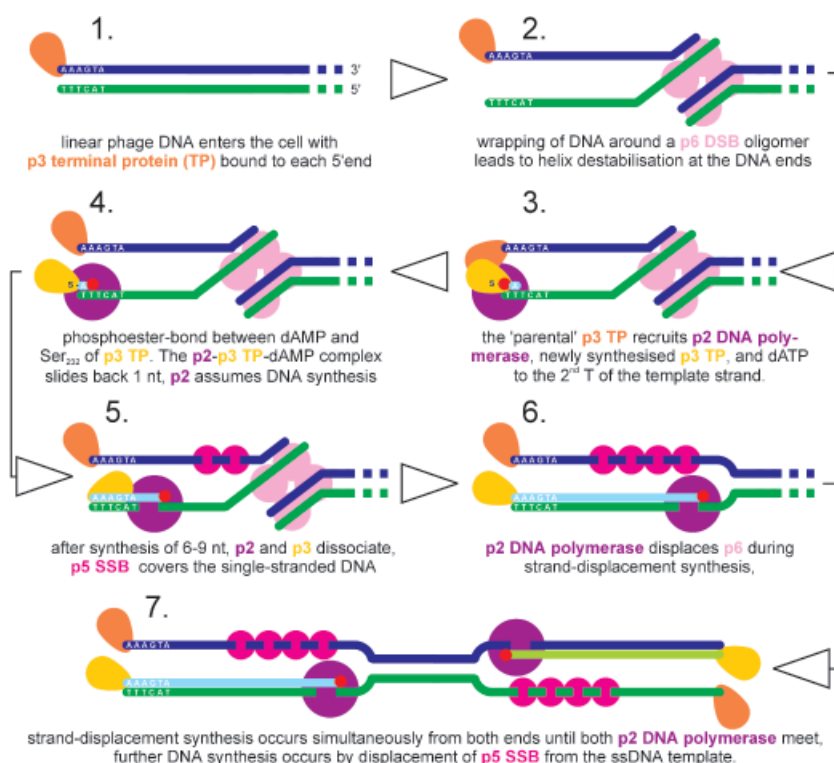


Fig. 6. The mechanism of protein-primed DNA replication (ppDR): ϕ 29. Successive steps are indicated by open arrows/arrowheads plus numbering (see text for details). Complementary DNA strands are shown as parallel lines. Dark blue/green, parental DNA strands; light blue/green, daughter DNA strands. Red dots indicate free 3'-hydroxyls used for DNA synthesis. The colour code used for individual proteins (coloured circles) corresponds to the coloured protein names in the text.

Initiation at the ends of linear DNA: protein-primed DNA replication

All known (replicative) DNA polymerases require a 'primer' – a free 3'-hydroxyl group provided by the strand complementary to the template strand – because they cannot start *de novo* DNA synthesis. DNA polymerases can elongate either the 3'-OH end of nicked dsDNA as in RCR (see 'Initiation by nicking: 'rolling circle'-type DNA replication' section) or the 3'-OH end of a short RNA primer synthesised by a specialised RNA polymerase, a primase, as in θ DR (see 'Initiation by melting: theta (θ)-type DNA replication' section). In the forthcoming section, we will discuss 'priming' by transcripts that remain bound to their templates, and in the subsequent section, 'priming' by the 3'-end of ssDNA invading a duplex. In addition, the CCA-3' stem of (uncharged) tRNA can serve as 'primer', as was found for retrovirus replication and has been proposed for the replication of Cauliflower Mosaic Virus (CaMV) (Pfeiffer & Hohn, 1983). Finally, a sterically favourably positioned hydroxyl-group of Ser, Thr, or Tyr residues in a protein can serve as 'primer', which led to the terminus 'protein-primed DNA replication' (ppDR). In all known cases, this 'portable primer' protein remains attached to the 5'-end of the newly synthesised DNA strand, and is called 'terminal protein' (TP) by convention.

Examples for linear, dsDNA replicons with proteins covalently attached to both ends include pro- and eukaryotic viruses, e.g. Adenoviruses, and eukaryotic plasmids, e.g. the Kalilo plasmid(s) of *Neurospora*. The linear chromosomes of *Streptomyces* species and of several linear *Streptomyces* plasmids contain terminal proteins; these replicons initiate DNA synthesis at internal origins, and the terminal proteins are part of a special mechanism allowing the full replication of the partially single-stranded telomers (Bao & Cohen, 2003). Note that the linear chromosome and several linear plasmids of *Borrelia burgdorferi* do not contain terminal proteins but possess hairpin ends like *E. coli* phage ϕ N15 (see above), and also a cognate protelomerase has been identified (Deneke *et al.*, 2004; Huang *et al.*, 2004).

Phage 29: The initiation of replication of *B. subtilis* phage ϕ 29 occurs exclusively at the phage ends, and depends on the terminally attached protein. ϕ 29 is therefore the 'model of choice' for studying replication by 'melting' of the terminal hydrogen bonds of linear dsDNA molecules (mechanism 3; see above). More importantly, ppDR has been studied in great detail for ϕ 29 (reviewed in (Meijer *et al.*, 2001), and the individual steps are discussed in the following (Fig. 6):

Step 1. The linear, double-stranded phage DNA enters the host cell with p3 TP covalently bound to the 5'-terminal bases. To discriminate between TP bound to genomes that

have already been replicated and free, newly synthesised TP, the former is called 'parental' TP.

Step 2. The $\phi 29$ p6 double-strand binding protein has been shown to form oligomers (Abril *et al.*, 1997, 1999), and to bind preferentially to sites with intrinsic DNA curvature; two such sites are located at a distance of 46–62 bp from the left end of the phage genome, and 68–125 bp from the right end, respectively (Serrano *et al.*, 1989). Nucleoprotein complexes formed by p6 heavily distort the bound DNA, and this distortion has been suggested to be responsible for helix destabilisation at the phage genome ends (Serrano *et al.*, 1993).

Step 3. The penultimate T residue of $\phi 29$ DNA serves as template for dAMP incorporation to TP catalysed by p2 DNA polymerase (Méndez *et al.*, 1992) (sequence shown in Fig. 6). $\phi 29$ p2 DNA polymerase binds newly synthesised TP in solution and performs the deoxyadenylation reaction *in vitro* without a DNA template, albeit with higher selectivity for dATP and more efficiently in the presence of parental p3 TP bound to $\phi 29$ DNA (Blanco *et al.*, 1992). Binding of the p2 · p3 complex to $\phi 29$ ends is activated, in addition, by the correctly positioned p6 nucleoprotein complex (Freire *et al.*, 1996).

Step 4. p2 catalyses the deoxyadenylation of Ser₂₃₂ in p3 TP (Hermoso *et al.*, 1985). Following this reaction, the p2 · p3 · dAMP complex slides back 1 nt, and p2 starts DNA synthesis at position +2 (Méndez *et al.*, 1992).

Step 5. The 'elongation phase' of $\phi 29$ replication starts with the dissociation of p3 TP and p2 DNA polymerase after template-directed synthesis of 6–9 nt (Méndez *et al.*, 1997). Covering of the ssDNA by p5 SSB protects the displaced strand from nuclease digestion. In addition, p5 prevents hairpin formation that might slow down DNA synthesis by p2, prevents template-switching of p2 and supports p2 processivity by helix destabilisation (Martín *et al.*, 1989; Soengas *et al.*, 1995; Esteban *et al.*, 1997).

Steps 6 and 7. During the elongation phase, p2 displaces the p6 nucleoprotein complex. DNA synthesis is initiated at both ends of the $\phi 29$ genome, and the two parental strands become separated when the replication forks pass each other. Both p2 DNA polymerases continue with DNA synthesis until the end of the single-stranded template. Each progeny molecule contains p3 TP bound to its 5'-ends.

Salas and colleagues point to the intriguing observation that two additional phage-encoded proteins participate in $\phi 29$ replication *in vivo*: p1 and p16.7 (reviewed in Bravo *et al.*, 2005). The $\phi 29$ replisome could be targeted to a membrane-associated p1 multimeric structure by interaction between p1 and primed TP. The integral membrane protein p16.7 is thought to recruit the $\phi 29$ DNA replisome through interaction with both the parental TP and the ssDNA. Both proteins can thus be envisaged as parts of

' $\phi 29$ replication factories', i.e. stationary protein complexes through which the DNA is threaded during replication.

Initiation of DNA replication by transcription

The mechanism of initiation of DNA replication by transcription (tDR) has been studied in four experimental systems representative for all three types of prokaryotic replicons: constitutively stable DNA replication (cSDR) of the *E. coli* chromosome, replication of ColE1-type plasmids, 'early' replication of *E. coli* phage T4 and replication of *E. coli* phage T7. The basic features of tDR are remarkably similar in all four systems. Briefly, RNA polymerase binds to a promoter on dsDNA and synthesises a short (untranslated) transcript that remains attached to its template. The transcript provides the 3'-OH end used by DNA polymerase for displacement synthesis of a leading strand. Formally, RNA polymerase performs the triple function of: (1) an initiator (DNA melting), (2) a helicase (DNA unwinding) and (3) a primase (providing the 3'-OH end of an RNA primer). Strand-displacement synthesis switches to unidirectional, coupled leading- and lagging-strand synthesis upon assembly of a primosome (primase/helicase) on the displaced strand (R-loop). Lagging-strand synthesis on the displaced strand further widens the loop allowing primosome assembly on the opposite strand. The net result of these reactions is bidirectional replication of the template.

In ColE1-type plasmids the RNAPII transcript assumes a complex tertiary structure upon trimming by RNase H, and its elongation is performed by PolA. Accordingly, ColE1-type plasmids (e.g. pBR322) cannot be propagated in *polA* mutants of *E. coli*. A primosome-assembly site (PAS) that becomes single-stranded during PolA-driven displacement synthesis serves as entry site for the restart primosome (PriA, PriB, PriC, DnaT) (see 'Replication restart' section). Subsequently, the primosome recruits the replicative helicase, DnaB, and the primase, DnaG, and bidirectional replication is performed by DNA Pol III holoenzyme (reviewed in del Solar *et al.*, 1998). To be operative in *E. coli*, the mechanism of cSDR requires inactivation of the *rnhA* gene encoding RNase H, resulting in a longer half-life of various transcripts (reviewed in Kogoma, 1997). cSDR is not possible in *priA* or *polA* mutant strains, indicating that both proteins perform essential functions, similar to their functions for ColE1 replication. cSDR can sustain chromosome replication in *E. coli dnaA* or *oriC*-deletion mutants, showing that cSDR bypasses the 'normal' pathway of initiation of chromosome replication (Messer, 2002).

Phage T7: The replication pathways of phage T4 will be addressed in the following section. Here we discuss tDR of $\phi T7$, which can be divided into the following steps (Fig. 7):

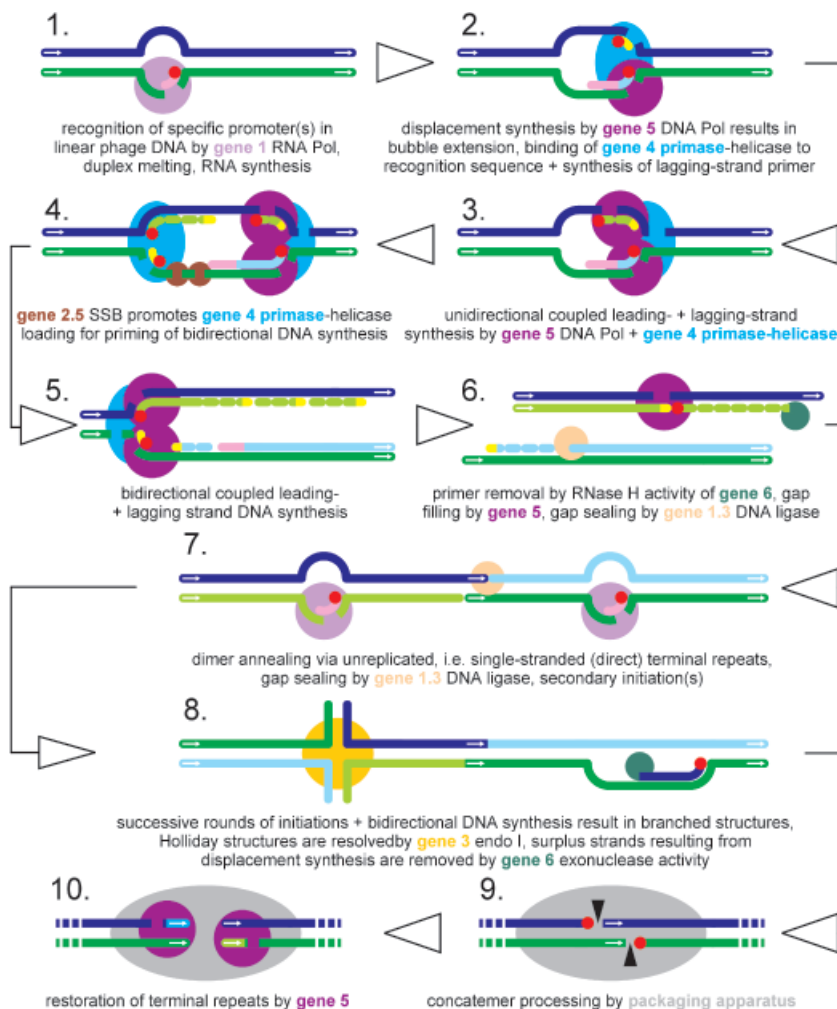


Fig. 7. The mechanisms of replication initiated by transcription (tDR): phage T7. Successive steps are indicated by open arrows/arrowheads plus numbering (see text for details). Complementary DNA strands are shown as parallel lines. Dark blue/green, parental DNA strands; light blue/green, daughter DNA strands; yellow, RNA primers. Red dots indicate free 3'-hydroxyls used for DNA synthesis. Phage-encoded proteins are shown as coloured circles, host factors as coloured triangles. The colour code used for individual proteins corresponds to the coloured protein names in the text. The 160 bp-long direct terminal repeats in ϕ T7 DNA are indicated by white arrows.

Step 1. Entry of the linear dsDNA of T7 into the host cell starts with the 'left end', and host RNA polymerase transcribes the 'early' genes, including gene 1 encoding T7 RNA polymerase. Gene 1 RNA polymerase is responsible for transcription of the 'primary origin' (Saito *et al.*, 1980; Fuller & Richardson, 1985b). Deletion of the primary origin results in initiation of replication from other T7 RNA polymerase promoters in the genome (Tamanai *et al.*, 1980; Wever *et al.*, 1980).

Step 2. The transcript is elongated through displacement synthesis by gene 5 DNA polymerase; priming of lagging-strand synthesis is performed by gene 4 primase-helicase upon binding to a gene 4 recognition site (5'-GGGTC) that becomes single stranded during displacement synthesis (Fuller & Richardson, 1985b).

Step 3. Coupled leading- and lagging-strand synthesis widens the replication 'bubble'. Such bubble structures were instrumental in defining the replication origin by early electron microscopic studies (Dressler *et al.*, 1972).

Step 4. Although it would be reasonable to assume that gene 2.5 SSB could cover the replication 'bubble' also during earlier steps, Fuller & Richardson (1985a) found a measurable positive influence of gene 2.5 protein only for the priming of bidirectional DNA synthesis *in vitro*.

Step 5. Bidirectional coupled leading- and lagging-strand synthesis results in Y-shaped replication intermediates observed in the electron microscope at some time after initiation (Dressler *et al.*, 1972).

Step 6. Enzymes involved in processing of the progeny molecules include gene 6 protein with RNase H activity, gene 5 and gene 1.3 DNA ligase.

Step 7. The replication of the linear ϕ T7 DNA is inherently incomplete. Owing to the presence of ~160 bp long direct terminal repeats, replication intermediates with 3'-overhangs can hybridise to each other, forming head-to-tail concatemers. Covalent linkage of the concatemers is achieved by the action of gene 1.3 DNA ligase. Concatemers

are subject to secondary initiation events, resulting in the cyclical growth of concatemer lengths.

Step 8. Secondary initiation events are instrumental for the overall growth of the concatemeric phage DNA, but in some cases collapsing or aborted forks result in Holliday structures. In addition, partially synthesised strands resulting from displacement synthesis create branched structures. Because branched DNA molecules are not appropriate substrates for packaging into phage heads, gene 3 Holliday-junction resolvase, gene 6 5' → 3' exonuclease activity and gene 1.3 DNA ligase are required for proper trimming of the concatemeric DNA.

Step 9. ϕ T7 DNA packaging into phage heads starts with the 'right end' (Son *et al.*, 1993). During packaging, the concatemeric DNA is cut by a site-specific nuclease in order to liberate genome monomers with single-stranded 5'-overhangs. The nuclease activity is associated with capsid proteins, and the recognition sites are known, but the responsible phage-encoded protein has not yet been unequivocally identified (White & Richardson, 1987; Chung & Hinkle, 1990).

Step 10. Processing of phage concatemers occurs in a way that allows the restoration of the terminal repeats by a fill-in reaction. This reaction requires, in addition to gene 5 DNA polymerase, gene 6 exonuclease to prevent strand-displacement synthesis (White & Richardson, 1987; Serwer *et al.*, 1990). Successful packaging also requires gene 2 protein as inhibitor of host RNA polymerase (LeClerc & Richardson, 1979).

When the replication schemes for ϕ T7 (Fig. 7) and λ (Fig. 3) are compared, it is apparent that they are virtually identical for the steps following priming of lagging-strand synthesis up to the end of the first round of DNA synthesis. The similarities and differences among the proteins responsible for performing the successive enzymatic steps are discussed in detail in COM section C3. tDR and θ DR are replication mechanisms that rely on duplex melting: either by RNA polymerase or on an initiator. We have discussed above that θ DR requires a negatively supercoiled substrate. By contrast, tDR can be initiated by RNA polymerase on a relaxed linear substrate. However, this difference may not be as significant as it appears at first sight: RNA polymerases are known to modulate the local superhelicity of their templates during transcription, and this (local) superhelicity has been proposed to be important for R-loop stability (Liu & Wang, 1987; Rahmouni & Wells, 1992; Drolet *et al.*, 1994).

Recombination-dependent DNA replication

The basic reaction in recombination-dependent DNA replication (RDR) is the annealing of a single-stranded stretch of

DNA to a homologous strand in dsDNA in such a way that the free 3'-hydroxyl end of the 'invading' strand can serve as primer for DNA polymerase. The annealing of the two complementary strands during this recombination process is promoted by proteins with strand-annealing property (SAPs), helicase(s) and SSBs. Homologous recombination resulting in displacement loops (D-loops) may proceed with further annealing – including also the complementary strand of the invading 3' end – and branch migration. The recombination intermediates are finally resolved by structure-specific endonucleases, e.g. Holliday junction resolvases. Depending on the pathway, the resulting structures are 'splice' or 'patch' variants of 'join-break' recombination. If the D-loop created by homologous recombination serves for primosome and replisome assembly, we may talk of 'join-copy' recombination (Mosig, 1994), or RDR. The formation of a D-loop as the first step to initiate DNA synthesis classifies RDR as a 'melting' mechanism, by formal criteria.

RDR is not suitable for complete *de novo* replication of a replicon, unless it contains tandemly repeated sequences. However, if (partially) duplicated replicons are present in a cell, RDR can serve as an efficient bypass mechanism for replication initiation, e.g. when the 'normal' initiation pathway is disabled. RDR of the *E. coli* chromosome was first detected by Lark and Kogoma (Kogoma & Lark, 1975), and studied in great detail by Kogoma (1997) as 'induced stable DNA replication' (iSDR). iSDR can sustain chromosome replication for several hours in the absence of protein synthesis upon induction of the SOS-response. Because the induction of the SOS-response inhibits cell division, iSDR is, unlike cSDR, not a replication bypass mechanism allowing cell proliferation. iSDR does not require DnaA, but depends crucially on intact recombination functions (RecA, RecBC) and PriA to form restart primosomes (see 'Replication restart' section). The importance of RDR for chromosome replication in *E. coli* under normal growth conditions is still a matter of debate; that RDR serves to rescue broken chromosomes and stalled replication forks is, however, generally accepted (Kuzminov, 1999; Cox *et al.*, 2000; Maisnier-Patin *et al.*, 2001) (see 'Replication restart' section).

Escherichia coli phage Mu depends entirely on host enzymes for the replication of its genome. The phage-encoded, oligomeric MuA transposase complex transfers ϕ Mu ends to (nonhomologous) target DNA. The MuA 'transpososome' creates a fork at each end, and remains tightly bound to both forks. The host ClpX chaperone is required for a 'loosening' of the DNA interaction(s) of the transpososome. An as yet unidentified host factor further displaces the transpososome and promotes the assembly of a restart primosome, which subsequently recruits the DnaBC

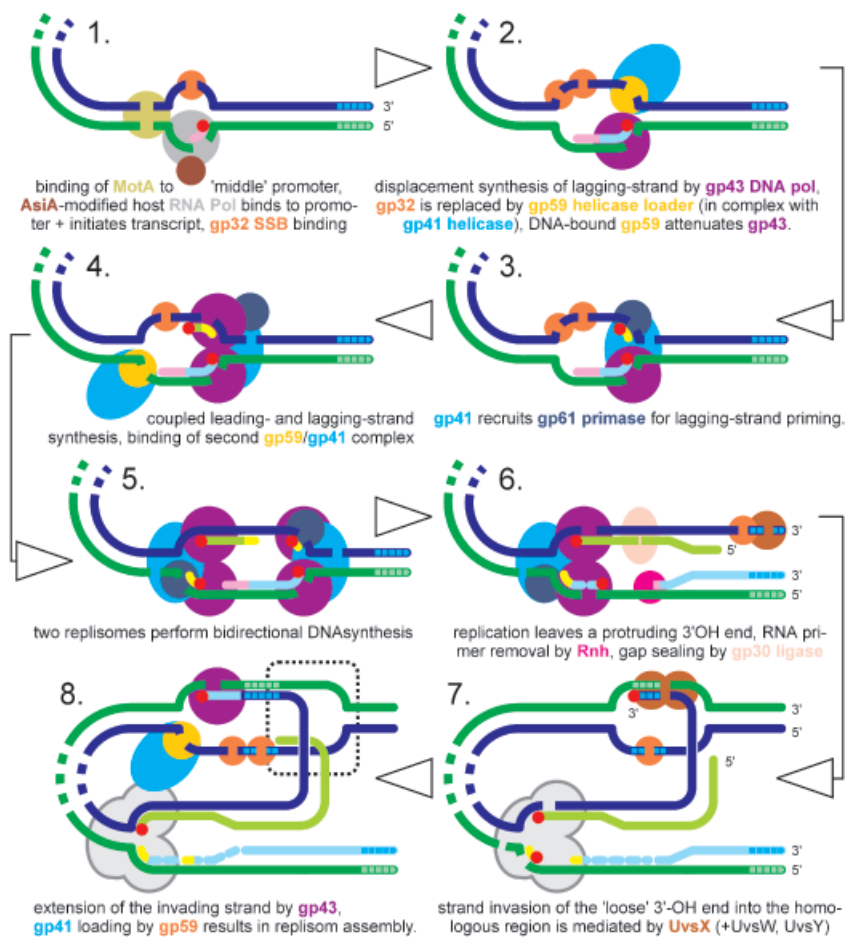


Fig. 8. The mechanism of recombination-dependent DNA replication (RDR): phage T4. Successive steps are indicated by open arrows/arrowheads plus numbering (see text for details). Complementary DNA strands are shown as parallel lines. Dark blue/green, parental DNA strands; light blue/green, daughter DNA strands; yellow, RNA primers. Red dots indicate free 3'-hydroxyls used for DNA synthesis. The colour code used for individual proteins (shown as coloured circles) corresponds to the coloured protein names in the text. The terminally redundant sequences of the ϕ T4 genome are indicated by small blocks within the DNA strands. For steps 7 and 8 the replisome formed during tDR is shown as a grey silhouette to avoid the diagram becoming overly complicated.

helicase complex (see 'Replication restart' section). 'Replicative transposition' of ϕ Mu thus ends up in the 'normal' pathway for RDR (reviewed in Nakai *et al.*, 2001).

Phage T4: Among phage replicons that (unlike Mu) encode cognate replication proteins, RDR is best understood for *E. coli* phage T4. The interdependence of recombination and replication was already the subject of a review by Broker & Doermann (1975), at that time mostly based on results of genetic and electron microscopic studies. Later it was established that ϕ T4 replication proceeds in two stages: the initial, rifampicin-sensitive stage that depends on (host) RNA polymerase, and the second, 'burst' stage that is suppressed in recombination mutants (Luder & Mosig, 1982). The genome of ϕ T4 has a size of 168.8 kb (Miller *et al.*, 2003). Linear ϕ T4 DNA entering the host cell has a size of ~ 173 kb and is circularly permuted, i.e. it contains ~ 3 – 5 kb terminal redundancy. As Mosig pointed out, the terminal redundancy is 'sufficiently large to allow homologous recombination between the terminal regions of a single chromosome, allowing successful infection of a host cell by a single T4 particle' (Mosig *et al.*, 1995, p. 86). We

present in Fig. 8 a scheme for ϕ T4 replication, which combines the initial phase of tDR with the subsequent RDR phase. This scheme should be regarded as 'minimal' in the sense that the essential steps are included but not the amazing number of known bypass mechanisms, which have always made ϕ T4 replication a topic suitable for monographs rather than for reviewing articles. As above, we discuss individual steps (Fig. 8):

Step 1. 'Early' replication of ϕ T4 is initiated by transcription from one of several origins, *oriA*, *F*, *G* and *E*. These origins are promoters that are specifically recognised by host RNA polymerase after replacement of the $\sigma 70$ subunit by the phage-encoded *AsiA* σ -factor. In addition, promoter-recognition by the modified RNA polymerase requires phage-encoded transcriptional activators: *MotA* in the case of *oriA*, *oriF* and *oriG*, and *DbpC* in the case of *oriE* (Mosig *et al.*, 1995). The transcripts synthesised by RNA polymerase remain attached to their template strands, thus forming an R-loop structure. A structure downstream of the promoter has been shown to possess properties of a 'DNA-unwinding element' (DUE) and might be required for the stability of the RNA·DNA heteroduplex (Carles-Kinch & Kreuzer,

1997). In addition, transcript stability may require formation of complex secondary structures including partial hybridisation to the displaced DNA strand (Mosig *et al.*, 1995).

Step 2. Nossal *et al.* (2001) were able to show that the ϕ T4 replication proteins perform DNA synthesis *in vitro* on an artificial substrate mimicking the R-loop structure at *ori-F(uvsY)* using the 3'-OH end of the RNA transcript as primer. Binding of gp43 DNA polymerase requires the gp45 sliding clamp for processive DNA synthesis. The homo-trimeric gp45 sliding clamp is assembled around the ssDNA downstream of the free 3'-OH end of the RNA · DNA heteroduplex by the gp44/gp46 clamp loader complex (to avoid 'molecular crowding', gps 45, 44 and 46 are not shown in Fig. 8; for details see COM section C3.5.2.). It appears that efficient DNA synthesis *in vitro* is synchronised by the versatile gp59 helicase loader: (1) gp59 removes gp32 SSB from the displaced DNA strand in the R-loop upon interaction (Ishmael *et al.*, 2001), (2) it loads the gp41 helicase to the displaced strand (Venkatesan *et al.*, 1982; Barry & Alberts, 1994b) and (3) it can slow down polymerase activity of gp43 until the helicase is completely loaded (Nossal *et al.*, 2001). Barry & Alberts (1994a) identified an alternative pathway for gp41 loading *in vitro*: in the absence of gp59, the Dda helicase can remove the RNA polymerase ahead of the synthesizing gp43 DNA polymerase, which in turn allows recruitment of gp41 directly by gp43.

Step 3. The gp41 helicase recruits the gp61 primase for synthesis of the first lagging-strand primer (Burke *et al.*, 1985). Steps 4 and 5. Elongation of the lagging-strand primer is performed by a second gp43 DNA polymerase recruited to the forming replisome. The complete ϕ T4 replisome is now composed of gp41 helicase, gp61 primase for cyclical priming of lagging-strand synthesis (Okazaki fragments), two gp45 sliding clamps plus their gp44/gp62 clamp loaders, and two gp43 DNA polymerases for coupled leading- and lagging-strand synthesis (Salinas & Benkovic, 2000; Kadyrov & Drake, 2001). The first lagging strand is converted to the leading strand by a second replisome assembled for DNA synthesis in the opposite direction (step 5).

Step 6. Replication of the linear ϕ T4 DNA is inherently incomplete. The 3'-end of the lagging-strand template is covered by gp32 SSB and UvsX SAP. UvsX requires UvsY as accessory protein, but the function of UvsY is not known exactly. Recent results suggest the UvsY supports UvsX loading by weakening the association of gp32 with ssDNA (Bleuit *et al.*, 2004).

Step 7. Analyses by electron microscopy revealed that the unreplicated ends of newly replicated ϕ T4 DNA molecules preferentially 'invade' the terminally redundant region at the other end of the same molecule, or in other chromosomes in the case of multiple infections (small coloured blocks at the chromosome ends in Fig. 8) (Dannenbergh & Mosig, 1983).

H-type structures indicative for (double) branch migration could only be found under conditions when replication was repressed (Broker, 1973). Therefore, the switch from tDR to RDR during ϕ T4 replication seems to be highly efficient under normal conditions. Alberts and Formosa simulated the initial step of RDR *in vitro*: replication could be obtained with the appropriate DNA substrates and purified proteins DNA polymerase 'holoenzyme' (gps 43, 44, 45 and 62), gp32 SSB, Dda helicase and UvsX SAP (Formosa & Alberts, 1986). Step 8. We have followed ϕ T4 replication up to the point where strand invasion by the unreplicated 'loose' 3'-OH end into the terminal redundancy of the same molecules solves the problem to replicate fully the linear phage DNA. Gp43 DNA polymerase elongates the annealed 3'-OH end, and replisome formation occurs in the D-loop as above in step 2. Thus, the switch from tDR to RDR is completed. The complex topological structure (Fig. 8, boxed area) created by the recombination step can be resolved through endonucleolytic cleavage by gp49 endonuclease VII. Alternatively, additional priming of DNA synthesis may occur in the opposite direction because the gp59 helicase loader shows preferential binding to branched fork structure (three-way or four-way junctions) (Jones *et al.*, 2000). During the 'burst' phase of ϕ T4 replication, steps 6–8 are repeated until the exhaustion of the dNTP pools, but secondary origin-dependent initiations for tDR occur rarely (Mosig *et al.*, 1995). UvsW helicase, expressed later during infection, may participate in suppression of origin-dependent initiations by removing the RNA from the R-loop (Dudas & Kreuzer, 2001).

Packaging of ϕ T4 DNA into phage heads requires genome concatemers. Therefore, the 'network' of interwoven recombination structures created by RDR has to be 'trimmed', i.e. Holliday junctions resolved, branches created by dismissed replication forks eliminated and all gaps sealed. The phage-encoded proteins mentioned above can perform all the required functions, and make ϕ T4 replication independent of host functions up to this last step, DNA packaging.

We wish to emphasise again that this scheme for ϕ T4 replication presents a 'minimal version' and only includes the recombination 'pathway II' believed to play the major role for ϕ T4 replication under normal growth conditions (Mosig, 1998). The multiple replication and recombination pathways encoded by ϕ T4 are probably the result of consecutive adaptations of the phage to a great variety of growth conditions, preserving its (almost complete) independence of host functions.

Replication restart

Replication research has always been greatly influenced by the 'replicon model' (Jacob *et al.*, 1963) and the

physiological studies of Maaloe and his collaborators, summarised in their statement: 'We are therefore led to believe that the overall production of DNA, RNA, and protein is regulated by mechanisms that control the frequencies with which the synthesis of individual nucleotide and amino acid chains are initiated' (Maaloe & Kjeldgaard, 1966, p. 163). Translation, transcription and replication have indeed been shown to be regulated primarily at the initiation step (see the previous subsections). However, replication research has for many years neglected the fact that any premature abortion of DNA synthesis is not just a waste of energy but a challenge to replicon integrity, so severe that the existence of a salvage pathway(s) could have been anticipated. For about 10 years, the elucidation of molecular pathway(s) promoting restart of DNA synthesis at stalled or dismissed replication forks has developed into a new field bringing together recombination and replication research (Cox *et al.*, 2000; Lusetti & Cox, 2002). Replication restart is mostly studied for chromosome replication, but two bacteriophage replication systems have greatly influenced the present models: (1) the conversion of (+)-strand DNA of ϕ X174 into the replicative form by *E. coli* enzymes (see 'Initiation by nicking: 'rolling circle'-type DNA replication' section) and (2) the recombination-dependent DNA replication of ϕ T4 (see 'Recombination-dependent DNA replication' section).

The conversion of (+)-strand DNA of ϕ X174 into the replicative form *in vitro* depends on the *E. coli* proteins N (PriB), N' (PriA or Y), N'' (PriC), I (DnaT), DnaB, DnaC, DnaG and DNA polymerase III holoenzyme; the conversion is completed upon removal of the RNA primer by PolA and gap sealing by DNA ligase (Schekman *et al.*, 1975). PriA, PriB, PriC and DnaT are required to load the replicative helicase as DnaB₆C₆ hetero-hexamers to the SSB-coated template. DnaC dissociates from the complex after helicase loading, and the remaining proteins are collectively called the preprimosome. The recruitment of the DnaG primase by DnaB converts the preprimosome into the primosome (Tougu & Marians, 1996; Chang & Marians, 2000). DnaB and DnaG recruit DNA polymerase III through multiple protein-protein interactions, thus forming the replisome (Zechner *et al.*, 1992; reviewed in Kornberg & Baker, 1992; Marians, 1996). Results from *in vitro* studies led to the suggestion that the primosomal proteins PriA, PriB, PriC and DnaT remain in physical contact with the replisome during DNA synthesis (Ng & Marians, 1996). During *E. coli* chromosome replication from *oriC*, primosome formation requires DnaA, DnaB₆C₆ and DnaG (Messer & Weigel, 1996; Hiasa & Marians, 1999). The terms 'DnaA primosome' (or 'ABC primosome'; Masai *et al.*, 1990) and 'PriA primosome' (or ' ϕ X primosome') reflect the differences of both primosomes with respect to protein composition. Despite its established function for ϕ X174 replication, a role for the

PriA primosome in chromosome replication remained elusive until Kogoma discovered that *E. coli priA*(null) strains are defective in recombination and DSB repair and proposed that PriA is responsible for replisome assembly at recombination intermediates, leading to RDR (Kogoma *et al.*, 1993; Masai *et al.*, 1994).

Escherichia coli PriA has detectable homologues in the genomes of most bacteria across all phyla and is genetically and biochemically well characterised (reviewed in Sandler & Marians, 2000). PriA is an SF2-type helicase (COM section C3.3.) that can unwind DNA in 5' → 3' and 3' → 5' direction, but the helicase activity is dispensable for its function as primosomal protein (Zavitz & Marians, 1992). PriA has a marked preference for binding to branched DNA structures *in vitro*, and binding to D-loops occurs downstream of the invading strand, albeit without strand preference in the absence of SSB (Jones & Nakai, 1999; Liu & Marians, 1999; Cadman & McGlynn, 2004). The results of recent genetic and biochemical experiments suggest that two (partially overlapping) pathways exist for restart primosome assembly: the PriA-dependent pathway involving PriA, PriB and DnaT, and a second, PriA-independent pathway mediated by PriC in conjunction with the *E. coli* Rep helicase (Sandler *et al.*, 1999; Sandler & Marians, 2000; Heller & Marians, 2005). Earlier, Seufert and Messer described yet another pathway for replication restart: if a replisome encounters a block shortly downstream from *oriC*, it re-initiates at PAS sites ~2 kb away. The interpretation was that the helicase remains attached and unwinds until a PAS site is exposed as a single-strand and used for (PriA-dependent) replisome re-assembly (Seufert & Messer, 1986). Homologues of PriB and PriC are present in the sequenced genomes of various Gram-negative bacteria, but could not be detected by BLAST searches in the genomes of Gram-positive bacteria (in addition, we could not detect phage-encoded homologues of PriA, PriB or PriC). Homologues of DnaT, however, are only present in the genomes of those species that also encode homologues of DnaC, and partial homologues are present in several phage genomes that code for a DnaC-type helicase loader. We will discuss in the 'Evolutionary considerations' our hypothesis that the *dnaTC* gene pair was acquired by *E. coli* from a replication module of an ancient lambdoid phage. From the above, we anticipate that further variants of the protein composition of PriA primosomes will be revealed upon analyses of replication systems in Gram-positive bacteria.

During the elongation phase, replisomes may encounter two types of nonprogrammed stops that result in replication fork stalling or collapse, and disassembly of the replisome: (1) chemically modified bases in one of the template strands, or (2) nicks in one of the two template strands or DSBs. In the first case, replication fork stalling may lead to

fork regression and annealing of the newly synthesised strands, i.e. formation of a Holliday junction or 'chicken-foot' structure. Fork regression regenerates templates for repair of the damage by either the nucleotide excision or base excision repair pathways. Repair can result in reversal of the fork regression and the fork structure itself may serve as substrate for primosome assembly. Alternatively, supported by recent experimental results, the resolution of the Holliday junction by RuvC or RusA could trigger replication restart in a reaction similar to that found for restart triggered by DSB (Seigneur *et al.*, 1998; McGlynn & Lloyd, 2000). In the second case, a nick in either of the template strands will lead to fork collapse and results in a DSB (Michel *et al.*, 1997). Resection of the linearised arm by the 5' → 3' exonuclease activity of RecBCD leaves a 3'-OH tail that is covered by RecA. Subsequently, RecA-mediated strand invasion in the 'intact' arm creates a D-loop (Kuzminov, 1999) that can be used as substrate for replication restart (Liu *et al.*, 1999). Because in all these cases replication restart could be shown to depend on the PriA primosome, Sandler & Marians (2000) proposed its re-naming as '(replication) restart primosome'. By genetic analysis of an *E. coli gyr* mutant strain, Ehrlich and co-workers showed that a requirement for the restart primosome also exists under conditions where replication restart did not involve recombination (Grompone *et al.*, 2003). It seems possible, therefore, that the 'backup' properties of the restart primosome are also required to face less severe impairments of replication fork progression. Estimates vary but it is likely that under most growth conditions a replisome starting from *oriC* has a 15–50% chance of being inactivated before reaching a *Ter* site (Maisnier-Patin *et al.*, 2001). This again emphasises the importance of the restart primosome.

During ϕ T4 replication, the origin-dependent initial phase (tDR) is followed by the 'burst' phase that (almost) entirely occurs in the RDR-mode (see previous section). It was shown by Kreuzer and co-workers that *in vitro* not only the invading 3'-ends of ϕ T4 can efficiently trigger RDR but also artificially introduced DSBs (George *et al.*, 2001). There is thus a convincing similarity between RDR in ϕ T4 and recombination-dependent replication restart of chromosome replication: with respect to the mechanism, but also with respect to the enzymatic functions involved (see Table 1 in Cox, 2001). However, both systems differ with respect to (1) the timely order of primosome and replisome assembly and (2) the properties (of some) of the primosomal proteins. In *E. coli*, the assembly of the restart primosome is a prerequisite for replisome assembly. By contrast, the ϕ T4 gp59 helicase loader promotes the loading of the gp41 helicase to the D-loop and slows down simultaneously ongoing DNA synthesis by gp43, probably for efficient 'coupling' of gps 41 and 43 in the replisome (Barry & Alberts, 1994a).

Marians and colleagues have proposed to use the term 'co-ordinated processing of damaged replication forks' (CPR) to account for the observation that recombination-dependent replication restart of chromosome replication in *E. coli* is an (essentially) error-free process, in contrast to error-prone DNA repair-synthesis during SOS induction (Sandler & Marians, 2000). To avoid 'abbreviation overload', we prefer to use the term RDR, at least as long as no fundamental mechanistic differences between 'recombination-dependent DNA replication' (RDR) in ϕ T4 and 'recombination-dependent restart' (RDR) of *E. coli* chromosome replication are revealed.

Bacteriophage replication modules

The term 'replication module' is often used in recent papers dealing with the architecture of bacteriophage genomes to account for the recurrent observation that replication genes co-localise in a distinct segment of phage genomes. In some cases, the detection of similarities of one or more predicted ORFs to particularly well-conserved proteins (e.g. helicases, DNA polymerases) were thought sufficient to pinpoint the 'replication module' of a particular phage genome. We do not reject this somewhat sloppy use of the term 'module' because it results in positive 'hits' in most cases. However, only a more precise definition of the replication module can prevent the misleading impression that the replication of a given phage is understood by pinpointing its 'replication module' the sloppy way.

Following accepted practice in molecular biology, a definition of bacteriophage replication modules should rely largely on the results of genetic and biochemical studies. A straightforward approach would start with phage DNA fragments ligated to a selectable marker, searching for autonomous replicating plasmids after transformation of an appropriate host. Comparable strategies led to the detection of λ dv plasmids (Matsubara & Kaiser, 1968), of the *E. coli* prophage Rac replication module (Díaz & Pritchard, 1978), of the ϕ adh replication module (Altmann *et al.*, 1999) and of the replication module of ϕ c2-type phages (Rakonjac *et al.*, 2003). However, this 'functional approach' is unsatisfactory at present, mainly for three reasons. One trivial reason is the lack of functional studies for the vast majority of known phage replicons. Another trivial reason is the implicit assumption that replication genes occur tightly packed in a single cluster, which is the case in most but not all known phage groups. The third reason becomes apparent when one looks more closely at the long record of research on the λ dv plasmids, which were discovered in 1968 (!) by Matsubara & Kaiser (1968). The initially studied plasmids contained the replication origin *ori λ* located within *O*, and the *O* (initiator) and *P* (helicase loader) genes transcribed from the *p_R* promoter together

with the *cII* and *cro* genes in an ill-defined context. It was shown in numerous subsequent studies that the *cII* and the *cro* regulatory loops are not essential for λ plasmid replication; transcription from the p_o promoter seems important but not the transcript, *oop* RNA (for details, see Taylor & Wegrzyn, 1995). Finally it was shown that p_R can be replaced by a different (inducible) promoter (Herman-Antosiewicz *et al.*, 2001), which relieves λ plasmid replication from the intricate control by host DnaA (Glinkowska *et al.*, 2003). These results led to the functional definition of the λ -type replication module being composed of the *O* (*ori λ*) and *P* genes (Wrobel & Wegrzyn, 2002). λ plasmids could thus serve as excellent model systems for the initiation – and initiation control – of bidirectional λ replication in the θ mode. In addition, unidirectional replication of λ plasmids, which precedes the switch from θ DR to σ DR during λ phage replication, could be shown (Baranska *et al.*, 2002). However, the switch from θ DR to σ DR – characteristic for λ phage replication – was never observed with λ plasmids. It is not clear at present whether this is due to the lack of the required recombination functions Red α /Red β (Exo/Bet), and RapA (NinG) in λ plasmids, or due to the lacking Gam function (inhibitor of host RecBCD). This demonstrates that the straightforward ‘functional approach’ to define replication modules can eventually fail to reveal auxiliary components.

Historically, the first useful definition of a prokaryotic replication module was given in the ‘replicon model’ by Jacob, Brenner and Cuzin: ‘The replicon is assumed to be a circular structure carrying two specific genetic determinants. A structural gene determines the synthesis of a diffusible active element, the initiator. The initiator acts on a replicator, allowing the beginning of the replication which proceeds along the circular structure’ (Jacob *et al.*, 1963, p. 331). A particularly startling aspect of the ‘replicon model’ was the hypothesis that the initiation of replication is positively regulated, which is indeed the case for all known bacteriophage replicons (Nordström, 2003). However, this clear-cut definition can only be applied to phage replicons with several important modifications. (1) The replicon may be circular or linear DNA. Many linear phage genomes recircularise prior to replication, but others initiate replication on the linear substrate. (2) The replicator (in modern terms: replication origin) is a unique structure in most phage replicons, but multiple origins are known for those phages where replication is initiated at D- or R-loops. (3) Many phages encode cognate initiators. However, phage replicons using R-loops for replisome assembly do not encode a cognate initiator in the strict sense.

With the notable exception of the ϕ T4-type phages, bacteriophages are semiautonomous replicons and have evolved various strategies to recruit components of the host

replication machinery. Therefore, we must include all phage-encoded replication functions in order to obtain a useful definition of ‘bacteriophage replication modules’. As we will show in the following, the close linkage of replication genes in most phage genomes justifies this expansion of the ‘replicon model’, and even suggests possible functions for experimentally uncharacterised proteins in some cases.

For a precise definition of phage replication modules, the emphasis on the initiation step in the ‘replicon model’ appears as a weak point. The replication of many phage genomes requires recombination steps that are, in most cases, performed by cognate recombination proteins in order to provide the relinearised form that is the substrate for packaging into phage capsids (COM section C3.6.2.). As we will show in the following, there is a striking co-localisation of replication and recombination genes in many phage genomes. Therefore, we include known and putative recombination genes in our definition of phage replication modules.

The discussion in this section will focus on four major types of replication modules: (1) modules containing initiator genes, (2) modules containing DNA polymerase genes, (3) modules containing ϕ P4 α -type helicase-primase genes and (4) the replication modules of filamentous phages. Although this formal division seems somewhat eclectic, it reflects the present knowledge – but not phage systematics, *nota bene*. Where possible, the definition of the individual types of replication modules is based on experimental results. We will include, in addition, the results of similarity searches discussed in COM section C3. Furthermore, the definitions will be based on the gene arrangements of fully functional phages as represented in the completely sequenced phage genomes. We include in the discussion several prophage genomes but because their replication/recombination genes might have undergone rearrangements and/or inactivation in the prophage state they cannot serve as a basis for the definition. We do not discuss in depth the important point of the transcription, and its regulation, of the bacteriophage replication/recombination genes because experimental results are too scarce and predictions doubtful. We expect, nevertheless, that a formal classification of phage replication modules will help to improve the assignment of putative (pro)phage gene functions in future genomic sequencing projects.

Phages encoding initiator proteins

We have discussed in COM section C3.1.2. the phage-encoded initiator proteins for θ DR, with λ O and ϕ SPP1 G38P as the best understood examples. Both initiator genes contain the phage replication origin, a common feature of

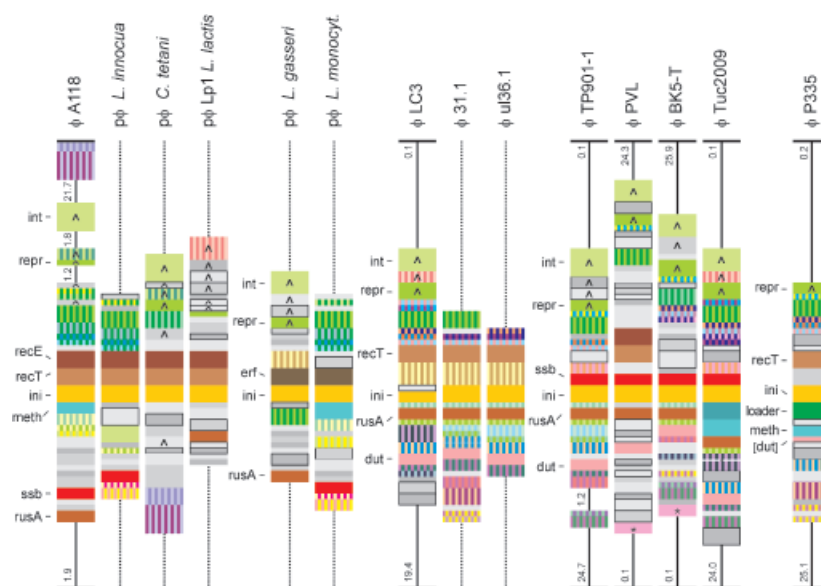


Fig. 9. 'Initiator-solo' replication modules of Gram(+) specific phages: part A. Genes encoding replication and recombination functions are shown in their genomic context. ϕ P335 (IL-type module) was included for better comparison. The alignment is shown with homologues of ϕ A118 gp49 (initiator) at fixed positions. Blocks with solid colours indicate (putative) gene functions detected by BLAST comparison: int, integrase; repr, λ cI-type phage repressor; recE, recT, rusA, ruvC, erf, putative recombinases; ini, initiator; loader, DnaC-type helicase loader; meth, methylase; ssb, single-strand DNA binding protein; dut, dUTPase. Apparently truncated genes are shown in square brackets. ORFs with significant similarity ($> 30\%$ identical residues) are indicated by striped colouring. Homologues of ϕ PVL orf63 and ϕ BK5-T orf63 (pink colour marked with an asterisk) were found in ϕ E125 (gp70) (see Fig. 12) and in ϕ 3626 (orf50) (see Fig. 15). Dark and light grey colouring indicates ORFs lacking homologues in any other phage genome. Dark and light grey colouring with black outline indicates ORFs with homologues in (completely sequenced) phages other than compared here. The ORF size is indicated by block height: ≤ 100 residues = 1 U, ≤ 200 residues = 2 U, ≤ 300 residues = 3 U, etc. The relative positions of the ORFs in the phage genomes are indicated by distances in kilobases (not to scale). Except where indicated by ' $>$ ' the direction of transcription is from up to down. The sequences were taken from the genome entries for ϕ A118 (40.8 kb) [NC_003216], *L. innocua* ϕ (3.0 Mb) [NC_003212], *C. tetani* ϕ (2.8 Mb) [NC_004557], *L. lactis* ϕ Lp1 (3.3 Mb) [NC_004567], *L. gasseri* ϕ (10.7 kb, partial sequence) [NZ_AAA02000006], *L. monocytogenes* ϕ (2.94 Mb) [NC_003210], ϕ LC3 (32 kb) [NC_005822], ϕ 31.1 (9.9 kb, partial sequence) [AF208055], ϕ ul36.1 (8.1 kb, partial sequence) [AF212846], ϕ TP901-1 (37.7 kb) [NC_002747], ϕ PVL (41.4 kb) [NC_002321], ϕ BK5-T (40 kb) [NC_002796], ϕ Tuc2009 (38.3 kb) [NC_002703] and ϕ P335 (36.6 kb) [NC_004746].

this type of initiator gene (COM section C2.2.). In their respective genomes, both initiator genes are directly followed by other replication genes: the λ P gene encoding the helicase loader and the ϕ SPP1 genes 39 and 40 encoding the helicase loader and the helicase, respectively (COM sections C3.2. + C3.3.). The initiation of replication of both phages differs with respect to the entry point of the host replication machinery: the origin-bound λ O recruits the host replicative helicase as λ P₃·DnaB₆ complex, while the origin-bound ϕ SPP1 G40P helicase recruits the host primase after dissociation of the unstable G38P·G39P·G40P·ATP intermediate complex (COM section C3.2.). We therefore discuss 'initiator-helicase loader' (IL-type) and 'initiator-helicase loader-helicase' (ILH-type) replication modules separately. A third type, the 'initiator-helicase' (IH-type) replication module, is represented by *Salmonella* sp. phage P22: the initiator gene 18, containing the ϕ P22 replication origin, is directly followed by the helicase gene 12. We will start the discussion with yet a fourth type, the 'initiator-solo' (I-solo) type of replication modules.

'Initiator-solo' replication modules

A number of Gram(+) specific phages possess a (putative) initiator gene – containing the phage replication origin – in addition to recombination genes, but lack detectable helicase loader or helicase genes. These phage genomes share a common architecture: genes encoding integrases and phage repressors are found upstream of the initiator gene, and transcribed in the opposite direction. Genes encoding exonuclease/SAP gene pairs (COM section C3.6.2.) are located between the initiator and repressor gene, and the direction of their transcription is the same as for the repressor gene. Genes encoding Holliday junction resolvases are, when present, invariably found downstream of the initiator gene. Genes encoding SSBs are present in several phage genomes, but at varying positions. Among the genes that are also invariably found downstream of the initiator gene are the genes encoding (putative) dUTPases, most frequently found in the genomes of *Lactococcus/Lactobacillus* phages (Fig. 9).

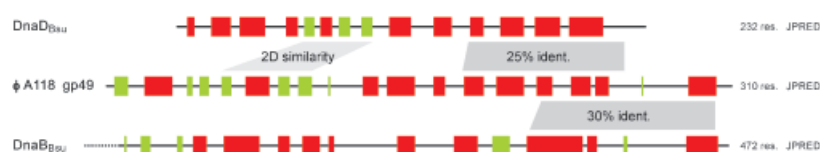


Fig. 10. Secondary structure prediction for the ϕ A118 gp49 initiator, and the *Bacillus subtilis* DnaD and DnaB helicase loaders. Secondary structure predictions for ϕ A118 gp49 [NP_463514], *B. subtilis* DnaB [NP_390777] and *B. subtilis* DnaD [NP_390116] were obtained by the Jpred method (Cuff *et al.*, 1998). Colour code: red, α -helical region; green, β -strand; black line, unstructured. Regions with significant similarity are indicated by grey blocks. '2D similarity' indicates a region showing a comparable secondary structure prediction but lacking protein sequence similarity. '% ident.' indicates regions with significant protein sequence similarity in addition to a similar secondary structure prediction.

The initiator, recombination and *ssb* genes are embedded in a highly variable context of mostly small ORFs with unknown function. A number of these small ORFs are found at corresponding or at different positions in other phage genomes. But there are also ORFs lacking known homologues. We can exclude the possibility that these small ORFs encode essential replication functions because none of them is found conserved in the entire set of phages compared in Fig. 9. The I-solo type of replication modules can thus be defined as: an initiator gene containing the phage replication origin, exonuclease/SAP genes either of the *recE/recT*- or the *erf*-type, and resolvase genes of the *rusA*-type. Because recombination genes are, like *ssb* genes, not found in all phage genomes we have to define them as accessory functions.

Whether it is appropriate to include ϕ Tuc2009 in the I-solo type group cannot be answered satisfactory at present: orf17 (241 res.) downstream of the orf18 initiator occupies the position where the helicase loader gene is found in ϕ P335 (Fig. 9). Homologous genes of ϕ Tuc2009 orf17 are found at corresponding positions in phages ϕ bIL285, ϕ ul36, ϕ ul36.2 ϕ Q33, ϕ Q30 and *L. lactis* p ϕ pi2. BLAST searches suggest a distant relationship of ϕ Tuc2009 orf17 to DnaI proteins of *Firmicutes* but the characteristic Walker A-type NTP-binding motif is not detectable. It seems possible that ϕ Tuc2009 orf17 represents a yet unknown type of helicase loader.

The initiation of replication in these phages has not been studied experimentally. There is good reason, however, to hypothesise that the mechanism resembles the intricate mechanism for helicase loading in *B. subtilis*, which involves the concerted action of the DnaB, DnaD and DnaI helicase loaders to recruit the replicative helicase DnaC (see COM section C3.2. for details). The (putative) initiator gp49 of *Listeria* sp. phage ϕ A118, and the almost identical (putative) initiators of *Staphylococcus aureus* phages ϕ PVL (orf46) and ϕ N315 (sa1791) contain in their C-terminal domain a region of similarity with the DnaD helicase loader of their hosts and with DnaD_{Bsu} [pfam04271] (Fig. 10), directly followed by – and partially overlapping with – a stretch of ~ 50 residues that shows significant similarity ($\geq 40\%$ ident. residues) with the C-termini of the DnaB helicase loaders of

their hosts, but with DnaB_{Bsu} only for ϕ A118 gp49 (Table 1, lanes 1–3; Fig. 10).

Several putative initiators encoded by prophages of bacillales genomes also show this particular arrangement of their C-terminal domains (Table 1, lanes 4–6). Interestingly, the replication initiators of various *Staphylococcus* sp. plasmids, which are structurally not related to phage initiators, also contain at their C-termini a stretch of ~ 50 residues showing significant similarity with the DnaB helicase loaders of their hosts (Table 1, lanes 7 and 8) and with the C-terminal ~ 50 res. of the phage initiators, respectively (not shown). 'MultAlin' analysis of the protein sequences from Table 1 did not produce a reasonable consensus sequence (Corpet, 1988). 'JPred' and 'PHD' secondary structure prediction analysis suggests, however, that the C-terminal half of the ~ 50 -residue-long 'DnaB-tail' assumes α -helical conformation in most proteins preceded by an unstructured loop with a conserved tryptophan residue (Fig. 10). The unstructured loop containing the conserved tryptophan is also detectable at the extreme C-terminus of most known DnaD_{Bsu} orthologues, but which all lack the predicted α -helix. The relatively small size of this 'DnaB-tail' suggests that it is of functional rather than of structural importance.

Phages ϕ TP901-1, ϕ BK5-T, ϕ 31.1 and ϕ ul36.1 of *Lactobacillus* sp. and related prophage genomes encode putative initiators with a detectable 'DnaB-tail' (Table 1, lanes 13–16). The initiators of ϕ BK5-T, ϕ 31.1, and ϕ ul36.1 show an overall similarity of $\sim 30\%$ among each other, with the similarity rising to $\sim 90\%$ (identical residues) within the C-terminal 50 residues. In contrast to the ϕ TP901-1 REP gene and the p ϕ Lp1 gene20 the putative initiators of phages ϕ BK5-T, ϕ 31.1 and ϕ ul36.1 and the *Lactobacillus gasseri* p ϕ gene lga0588 lack detectable similarity with the cognate DnaD. The putative initiators encoded by gene20 of the *Lactobacillus plantarum* prophages Lp1 and Lp2 lack detectable similarity in their N-termini but have virtually identical C-termini; only the former is therefore included in Table 1. For p ϕ Lp1 we found no similarity with DnaD_{Bsu} but with DnaD of *Bacillus halodurans* instead (Table 1, lane 15).

Orthologues of DnaB_{Bsu} are only detectable among species of the bacillales and lactobacillales subgroups of the

Table 1. Similarity of the initiators of Gram(+) specific phages to the DnaD and DnaB helicase loaders of their hosts

ϕ /pϕ/p/chr.	Accession no.	Gene	Residue	Host DnaD		DnaD _{Bsu}		Host DnaB (%)		DnaB _{Bsu} (%)	
				(entire length)		(entire length)		n – 100	n – 50	n – 100	n – 50
1. ϕ A118 (<i>Listeria monocytogenes</i>)	NP_463514	gp49	310	29%	172–275	25%	177–257	34	43	30	–
2. ϕ PVL (<i>Staphylococcus aureus</i>)	NP_058485	orf46	297	24%	87–273	23%	152–265	21	42	–	–
3. ϕ N315 (<i>S. aureus</i>)	NP_835538	sa1791	297	24%	87–273	–	–	21	42	–	–
4. <i>Listeria innocua</i> pϕ	NP_471742	lin2412	310	27%	152–273	25%	177–257	34	43	29	–
5. <i>L. innocua</i> pϕ	NP_469432	lin0086	303	22%	37–266	28%	171–249	30	41	33	–
6. <i>L. monocytogenes</i> pϕ	NP_465841	lmo2317	324	27%	161–287	25%	191–271	30	43	32	–
7. pIP1629 (<i>Streptococcus epidermidis</i>)	AAD02381	Rep1	285	–	–	–	–	40	41	–	–
8. pN315 (<i>S. aureus</i>)	NP_395563	sap027	286	–	–	–	–	–	45	–	–
9. <i>S. aureus</i>	NP_374796	DnaB	466	24%	268–394	–	–	nd	nd	29	–
10. <i>L. innocua</i>	NP_470932	DnaB	458	–	–	–	–	nd	nd	43	35
11. <i>L. monocytogenes</i>	NP_465086	DnaB	458	–	–	–	–	nd	nd	37	44
12. <i>Bacillus subtilis</i>	NP_390777	DnaB	472	nd	–	–	–	nd	nd	nd	nd
13. ϕ TP901-1 (<i>Lactococcus lactis</i>)	NP_112676	REP	272	27%	142–229	29%	145–245	–	48*	–	–
14. ϕ BK5-T (<i>L. lactis</i>)	NP_116541	orf49	269	–	–	–	–	–	48*	–	–
15. <i>Lactobacillus plantarum</i> pϕ Lp1	NP_784408	gene 20	310	29%	184–257	31% [†]	194–266	26	32*	21	–
16. <i>Lactobacillus gasseri</i> pϕ	ZP_00046421	lgas0588	307	–	–	–	–	30	43	–	–

Values are percentage identical residues; a dash indicates no significant homology detectable by BLAST (bl2seq); nd, not done (self-comparison). BLAST (bl2seq; Tatusova & Madden, 1999) similarity searches were performed for 'host DnaD' and 'DnaD_{Bsu}' [NP_390116] with the complete sequence as query; 'host DnaD' were: *L. monocytogenes* DnaD [NP_465419], *S. aureus* DnaD [NP_374567], *L. innocua* DnaD [NP_471343], *S. epidermidis* DnaD [NP_764696], *L. lactis* DnaD [NP_267226], *L. plantarum* DnaD [NP_785314] and *L. gasseri* DnaD [ZP_00045943]. For the columns showing the BLAST results with host DnaD and DnaD_{Bsu} as queries, the percentage identity value is given together with the position of the matching region in the subject sequence. BLAST (Bl2seq) similarity searches were performed for 'host DnaB' and DnaB_{Bsu} sequentially (1) for the C-terminal 100 residues (n–100) and (2) for the C-terminal 50 residues (n–50); 'host DnaB' were: *L. lactis* DnaB [NP_266907], *S. epidermidis* DnaB [NP_764914], *L. gasseri* DnaB [ZP_00046732] and *L. plantarum* DnaB [NP_785118].

*Similarity detected in a stretch of ~25 residues by genome BLAST with the C-terminal 50 residues of REP and orf49.

[†]Similarity with *B. halodurans* DnaD [NP_242563].

firmicutes (COM section C3.2.). This observation corresponds to the finding that initiators containing a 'DnaB-tail' could only be detected in the genomes of phages that infect species from these two phylogenetic groups. Several of the (putative) phage initiators analysed here contain in addition to the 'DnaB-tail' a region of similarity with DnaD_{Bsu} preceding the 'DnaB-tail' in their C-terminal domains. The direct comparison of the DnaB and DnaD proteins of *S. aureus* reveals a region of ~100 residues of limited similarity (Table 1, lane 9). This region corresponds to the region of similarity with DnaD found in the initiators of phages ϕA118, ϕPVL, ϕN315 and ϕTP901-1. Given the related function of the DnaD and DnaB proteins, this finding may point to a common evolutionary origin of both proteins and might help to unravel the origin of the phage initiators containing a 'DnaB-tail'. By contrast, the DnaD and DnaB proteins of *B. subtilis*, *L. gasseri*, *Listeria* sp. (Table 1, lanes 10–12), and *Lactococcus/Lactobacillus* sp. seem unrelated. We note, in addition, the rather low conservation of the DnaB proteins (C-termini) among the closely related genera *Bacillus*, *Listeria* and *Staphylococcus*. By contrast, other replication proteins are highly conserved, e.g. DnaA:

> 70% identical residues *Bacillus/Listeria*, ~60% identical residues *Bacillus/Staphylococcus*.

We deduce from the above that the initiators in the 'I-solo' type replication modules of Gram(+) specific phages contain in their C-terminal domains subdomains that interact with the DnaC replicative helicases of their hosts, and/or with the third helicase loader, DnaI, in addition. A more precise hypothesis would require a more detailed knowledge about helicase loading in firmicutes. It is safe, however, to assume that the initiators are sufficient to direct the host replication machinery to the phage replication origin at the step of helicase loading.

With a length of only 137 and 88 residues, the (putative) initiators of phages ϕNIH1.1 and ϕMM1, respectively, are unusually short (Fig. 11). The ϕNIH1.1 orf08 initiator protein shows similarity (27% identical residues) to the C-terminus of the ϕTP901-1 REP initiator. orf07 and orf08 are separated by an untranslated stretch of 473 bp (NCBI entry NC_003157; position 5719–6192). A 111-residue-long ATG-less ORF could be readily identified within this stretch that shows 42% identity with the N-terminus of the ϕTP901-1 REP initiator, and which may represent the 'missing' orf08

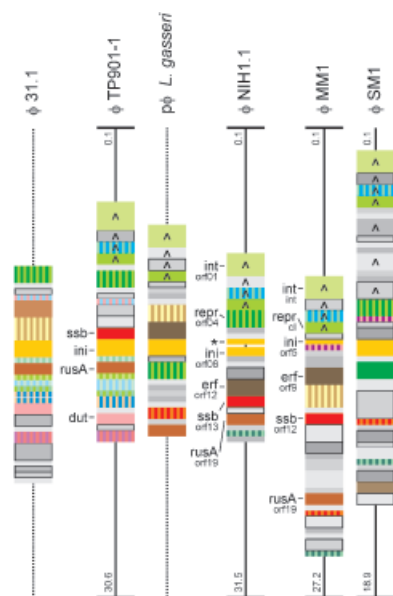


Fig. 11. 'Initiator-solo' replication modules of Gram(+)-specific phages: part B. Genes encoding replication and recombination functions are shown in their genomic context. ϕ 31.1, ϕ TP901-1 and *Lactobacillus gasserii* p ϕ (I-solo type module), as well as ϕ SM1 (IL-type module) were included for better comparison. The alignment is shown with homologues of ϕ TP901-1 REP (initiator) at fixed positions. Blocks with solid colours indicate (putative) gene functions detected by BLAST comparison: int, integrase; repr, λ cI-type phage repressor; erf, rnaA, putative recombinases; ini, initiator; ssb, single-strand DNA binding protein; dut, dUTPase. In ϕ NIH1.1, an asterisk marks the ATG-less ORF representing the putative initiator N-terminus (see text for details) upstream of the orf08 initiator (C-terminus). ORFs with significant similarity (> 30% identical residues) are indicated by striped colouring. Dark and light grey colouring indicates ORFs lacking homologues in any other phage genome. Dark and light grey colouring with black outline indicates ORFs with homologues in (completely sequenced) phages other than compared here. The ORF size is indicated by block height: ≤ 100 residues = 1 U, ≤ 200 residues = 2 U, ≤ 300 residues = 3 U, etc. The relative positions of the ORFs in the phage genomes are indicated by distances in kilobases (not to scale). Except where indicated by ' > ' the direction of transcription is from up to down. The sequences were taken from the genome entries for ϕ TP901-1 (37.7 kb) [NC_002747], ϕ 31.1 (9.9 kb; partial sequence) [AF208055], *L. gasserii* p ϕ (10.7 kb; partial sequence) [NZ_AAAO02000006], ϕ NIH1.1 (41.8 kb) [NC_003157], ϕ MM1 (40.2 kb) [NC_003050] and ϕ SM1 (34.7 kb) [NC_004996].

N-terminus (marked with an asterisk in Fig. 11). Although we have to await a revision of the ϕ NIH1.1 DNA sequence, we tentatively assume that ϕ NIH1.1 does not possess an 'unusual' I-solo type replication module. The situation is more complex for ϕ MM1. The (putative) initiator encoded by orf5 shows significant similarity to the N-terminal DNA-binding domains of several related phages, including the 111-residue-long (putative) ORF upstream of ϕ NIH1.1 orf08 (37% identical residues), but lacks a C-terminal oligomerisation domain. The distance to the downstream

gene orf6 is too short to accommodate a gene encoding a putative initiator C-terminus, which was also not found elsewhere in the ϕ MM1 genome, assuming the possibility of gene splitting. López and co-workers could identify a putative replication origin downstream of orf5, but a phage replication mechanism involving a (putative) initiator devoid of an oligomerisation/interaction domain has so far not been studied experimentally (Obregon *et al.*, 2003). The architecture of the ϕ MM1 genome resembles most closely that of ϕ NIH1.1: in both genomes several genes usually found upstream of the initiator gene are located downstream instead. It seems possible that the (putative) initiator gene ϕ MM1 orf5 suffered a deletion during this rearrangement (see Fig. 11).

Gram(−)-specific phages with I-solo type replication modules include ϕ SfV, ϕ ST64B and phage ϵ 15 of enterobacteria, and the *Burkholderia mallei* (Betaproteobacteria) phage ϕ E125. All four phages show some gross similarity in genome architecture among each other (Fig. 12) and in comparison to the Gram(+)-specific phages (Fig. 9). ϕ SfV and ϕ ST64B are closely related with long stretches of significant DNA sequence similarity along their entire genomes. Not surprisingly, the (putative) initiator genes ϕ SfV orf39 and ϕ ST64B sb42 are homologous (87% identical residues), and the arrangement of the flanking genes is well conserved. By contrast, the (putative) initiators ϵ 15 p42 and ϕ E125 gp60 share no protein sequence similarity, only the latter shows some weak similarity to the N-terminal DNA-binding domains of ϕ SfV orf39 and ϕ ST64B sb42. All four initiator genes contain the (putative) phage replication origins (COM section C2.2.). The apparent differences in the regions flanking the initiator genes of the four phages allows for a straightforward description of the replication module: an initiator gene containing the replication origin, and – as accessory functions – genes encoding Holliday junction resolvases of the RusA- (ϕ SfV, ϕ ST64B) or RuvC-type (ϵ 15). As found for the Gram(+)-specific phages with I-solo-type replication modules, the resolvase genes are located downstream of the initiator genes. A pair of *recE/recT*-type recombination genes is only encoded by ϵ 15, and located between the integrase and the phage repressor genes. This localisation seems to be conserved in Gram(−)-specific phage genomes. The function of the ParB-like protein gp58 of ϕ E125 remains to be studied. Despite the different arrangement of the recombination genes, the replication module of these four Gram(−)-specific phages is identical to that of the 'I-solo'-type replication module of the Gram(+)-specific phages.

The initiation of replication of these four phages has not been studied, but we can assume that the phage initiators recruit the host replicative helicase directly, i.e. without involving a specific helicase loader, in order to gain access to the host replication machinery. It should be kept in mind

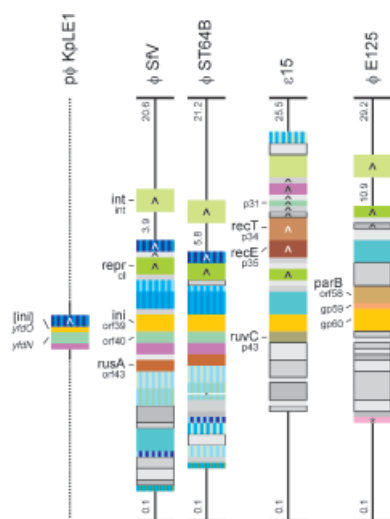


Fig. 12. 'Initiator-solo' replication modules of Gram(–)-specific phages. Genes encoding replication and recombination functions are shown in their genomic context. The alignment is shown with homologues of ϕSfV orf39 (initiator) at fixed positions. Blocks with solid colours indicate (putative) gene functions detected by BLAST comparison: int, integrase; repr, λ *cI*-type phage repressor; recE, recT, rnaA, parB, putative recombinases; ini, initiator. Apparently truncated genes are shown in square brackets. ORFs with significant similarity (> 30% identical residues) are indicated by striped colouring. Homologues of ϕE125 gp70 (pink colour marked with an asterisk) were found in ϕPVL (orf63), ϕBK5-T (orf63) (see Fig. 9) and ϕ3626 (orf50) (see Fig. 15). Dark and light grey colouring indicates ORFs lacking homologues in any other phage genome. Dark and light grey colouring with black outline indicates ORFs with homologues in (completely sequenced) phages other than compared here. The ORF size is indicated by block height: ≤ 100 residues = 1 U, ≤ 200 residues = 2 U, ≤ 300 residues = 3 U, etc. The relative positions of the ORFs in the phage genomes are indicated by distances in kilobases (not to scale). Except where indicated by '>' the direction of transcription is from up to down. The sequences were taken from the genome entries for *Escherichia coli* K12 ϕKpLE1 (4.6 Mb) [NC_000913], ϕSfV (37.1 kb) [NC_003444], ϕST64B (40.1 kb) [NC_004313], ϕε15 (39.7 kb) [NC_004775] and ϕE125 (53.4 kb) [NC_003309].

that the mechanism of helicase loading is fundamentally different in *E. coli* and *B. subtilis*: the *E. coli* DnaB helicase is recruited to the site of loading as a stable hetero-hexameric complex (DnaB₆C₆). By contrast, the *B. subtilis* DnaC helicase hexamer is assembled at the site of loading by the concerted action of the DnaB, DnaD and DnaI helicase loaders (see above). Despite their identical I-solo-type replication modules the initiator proteins of Gram(+) and Gram(–)-specific phages have (had) to adapt to these specific requirements of the host proteins. An interesting feature of the phage ε15 p42 initiator is its significant C-terminal similarity to the *E. coli* primosomal protein DnaT (residues 157–217; 40% identical residues). DnaT directs the DnaB₆C₆ double-hexamer to the restart primosome (DNA-bound PriA, PriB and PriC) during replication restart (see 'Replication restart' section). The region of similarity does

not include the ~20 C-terminal residues of DnaT, which may be important for the interaction with DnaC (see 'Evolutionary considerations' section). It is possible, therefore, that the 'DnaT-like' region in ε15 p42 includes a site for interaction with DnaB. Such a 'DnaT-like' region could not be detected in the other three initiators, though. It thus remains unresolved how they attract the host helicase. A hint may be the observation that the ϕSfV orf40 protein (162 residues) shares a region of significant similarity with the *E. coli* primosomal protein PriA (residues 34–75; 38% identical residues). This region is missing in the truncated orf40 homologue p31 of phage ε15. The molecular architecture of *E. coli* PriA is not well understood, and therefore an easily testable hypothesis cannot be derived from this observation. In the *E. coli* K12 genome, gene *yfdN* encodes a homologue of ϕSfV orf40 (95% identical residues). The preceding *yfdO* gene encodes a protein with significant similarity to the C-terminus of the ϕSfV orf39 initiator. YfdO and YfdN are thus the remnants of the replication module of the highly rearranged and truncated KpLE1 prophage (Fig. 12).

The small orf59 (86 residues) upstream of the orf60 initiator in the ϕE125 genome shows (BLAST) similarity to the N-terminus of the ϕV orf39 initiator, but no similarity to orf60. Although orf59 encodes a (putative) DNA-binding domain, we believe that the protein is a recombination relic – a partial duplication – rather than a functional important protein. The (putative) initiator encoded by the recently sequenced *Burkholderia pseudomallei* phage ϕ1026b is virtually identical with ϕE125 gp60 (97% identical residues), and the initiator gene is also preceded by the partial initiator duplication found in ϕE125. Both phage genomes are therefore very closely related, and ϕ1026b is not discussed separately here.

'Initiator-helicase loader' replication modules

In addition to the O initiator and the P helicase loader as essential factors, replication of λ phage may require the action of the Redα, Redβ and RapA recombination proteins (see 'Initiation by melting: theta (θ)-type DNA replication' section). Genes encoding all these functions are found at corresponding positions in the genomes of ϕ933W and ϕ4795, and define the components of the IL-type replication module. There is a considerable degree of similarity in arrangement and type of ORFs downstream of the initiator and helicase loader genes in λ, ϕ933W and ϕ4795, but this similarity is not found in the region upstream of the phage repressor gene (Fig. 13). Thus, the three phages are clearly distinct though related. ϕH-19B encodes an Erf-type instead of a Redβ-type SAP at the corresponding position in its (partially sequenced) genome, which adds further support to the notion that these proteins are functionally equivalent (COM section C3.6.2.). We found homologues of the λ *ren*

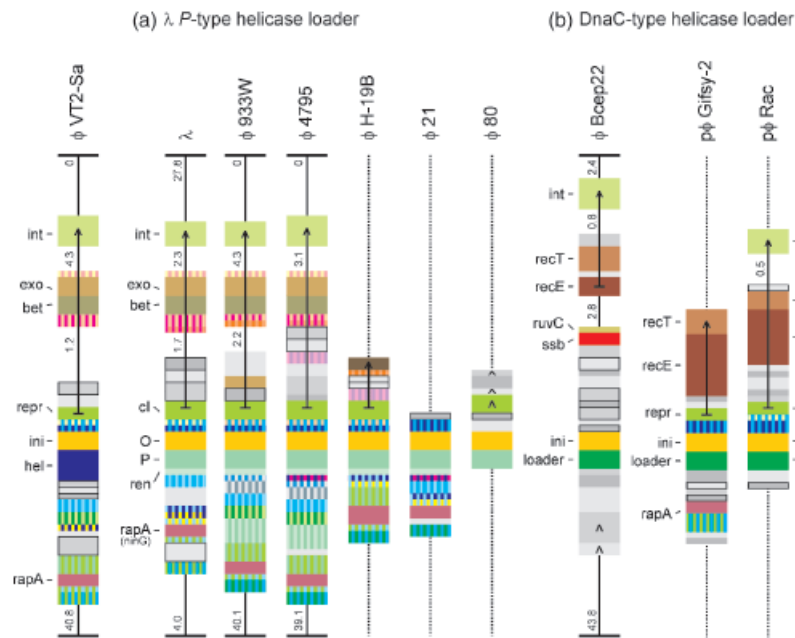


Fig. 13. 'Initiator-helicase loader' replication modules of Gram(–)-specific phages. Genes encoding replication and recombination functions are shown in their genomic context. ϕ VT2-Sa (IH-type module) was included for better comparison. The alignment is shown with homologues/functional analogues of λ O (initiator) at fixed positions. Blocks with solid colours indicate (putative) gene functions detected by BLAST comparison: int, integrase; repr, λ C1-type phage repressor; exo, λ red α -type exonuclease; bet, λ red β -type single-strand annealing protein (SAP); erf, ϕ P22 erf-type SAP; recE, recT, RecE-, RecT-type recombinases; ini, O, initiator; loader, DnaC-type helicase loader; P, λ P-type helicase loader; hel, F4-type helicase; rapA, RapA-type resolvase; ruvC, RuvC-type resolvase; ssb, single-strand DNA binding protein. For the Rac prophage, the gene names of the *Escherichia coli* genome are indicated. ORFs with significant similarity (> 30% identical residues) are indicated by striped colouring. Dark and light grey colouring indicates ORFs lacking homologues in any other phage genome. Dark and light grey colouring with black outline indicates ORFs with homologues in (completely sequenced) phages other than compared here. The ORF size is indicated by block height: ≤ 100 residues = 1 U, ≤ 200 residues = 2 U, ≤ 300 residues = 3 U, etc. The relative positions of the ORFs in the phage genomes are indicated by distances in kilobases (not to scale). Except where indicated by ' > ' or by an arrow spanning several ORFs, the direction of transcription is from up to down. The sequences were taken from the genome entries for ϕ VT2-Sa (60.9 kb) [NC_000902], λ (48.5 kb) [NC_001416], ϕ 933W (61.7 kb) [NC_000924], ϕ 4795 (57.9 kb) [NC_004813], ϕ H-19B (18.4 kb; partial sequence) [AF034975], ϕ 21 (4.9 kb; partial sequence) [B21237660], ϕ 80 (6 kb; partial sequence) [BP80ER], ϕ Bcep22 (63.9 kb) [NC_005262], *Salmonella enterica* (typhimurium) LT2 p ϕ Gifsy-2 (4.95 Mb) [NC_003197], *E. coli* K12 p ϕ Rac (4.6 Mb) [U00096]. (a) λ P-type helicase loader; (b) DnaC-type helicase loader.

gene in all phages encoding a P-type helicase loader, and invariably downstream of the P gene homologue, but not outside this phage group. *ren* provides resistance to Rex exclusion, and has therefore not to be considered to have a replication function (Campbell, 1994).

The architectures of λ and ϕ VT2-Sa are almost identical (Fig. 13). In fact, there is a stretch of ~ 5 kb of almost complete DNA identity in both phage genomes, encompassing the *red $\alpha\beta$* genes. However, both phages possess different replication modules, an IH-type module in the case of ϕ VT2-Sa and the IL-type in the case of λ . Although both replication modules are characterised by a different mechanism and point of attraction of the host replication machinery, this observation gives strong support to the notion that both modules are functionally equivalent and may be exchanged between phages by recombination, a topic discussed to more detail in the 'Replication module exchange among bacteriophages' section.

We found only few examples for IL-type replication modules with a DnaC-type helicase loader among Gram(–)-specific (pro)phages (Fig. 13b). The *Burkholderia* sp. (*Betaproteobacteria*) phage ϕ Bcep22 has a gross architecture resembling that of λ with three major differences: (1) it lacks a detectable phage repressor gene, (2) the exonuclease/SAP gene pair is of the *recE/recT*-type and (3) a *ruvC*-type resolvase gene is located upstream of the (putative) initiator-helicase loader gene pair. The direction of transcription of the ϕ Bcep22 *recET* gene pair corresponds to that of λ *red $\alpha\beta$* , and therefore ϕ Bcep22 is probably no exception from the 'rule' that recombination genes are located between the integrase and repressor genes in Gram-negative-specific phages. The small size (77 residues) of the *ruvC*-type resolvase ϕ Bcep22 gp15 suggests that the protein is inactive and represents a recombination relic.

The architectures of the *Salmonella* sp. prophage Gifsy-2 and the *E. coli* K12 Rac prophage closely resemble that of

ϕ Bcep22, and also that of λ . Given the highly similar architecture of Gifsy-2 and Rac, they may represent derivatives of a common ancestral phage. Both prophages encode fully functional replication genes, and are interesting examples for phage replicons where a phage-encoded helicase loader competes with a host orthologue, DnaC, for the host replicative helicase. The extremely long *recE* genes are uniquely found in these two prophages, with the exonuclease activity residing in the C-terminal ~ 300 residues (COM section C3.6.2.). Although RecE and RecT are functionally analogous to the λ Red α and Red β proteins, we note that the gene order is usually reversed in the *recET* gene pairs (Fig. 13).

As mentioned in COM section C3.2., the protein encoded by the p ϕ Rac *ydaV* gene (b1360) shows significant homology to *E. coli* DnaC (50% identity) (Wrobel & Wegrzyn, 2002; Casjens, 2003). Wróbel & Wegrzyn (2002) found, in addition, that the protein encoded by the p ϕ Rac *ydaU* gene upstream of *ydaV* shows significant homology (40% identity) to the C-terminal 80 residues of *E. coli* DnaT (179 residues). In the genomes of *E. coli* K12 and the other completely sequenced *E. coli* strains CFT073, O157:H7 and O157:H7 EDL933, *dnaT* is located directly upstream of *dnaC* in a small operon (Masai & Arai, 1988). *dnaTC* gene pairs are also present in the chromosomes of *Shigella flexneri* 2a (DnaT: 100% identical residues/DnaC: 100% identical residues), *Salmonella enterica* (typhimurium) LT2 (81%/93% identical residues), *Klebsiella pneumoniae* (74%/93% identical residues), *Buchnera aphidicola* APS (42%/65% identical residues) and *B. aphidicola* Sg (36%/65% identical residues). As mentioned in COM section C3.2, the *dnaTC* gene pair is not present in the chromosomes of sequenced *Yersinia* sp. strains (Thomson *et al.*, 2002), and also not in those of other species outside the enterobacterial branch of the *Gammaproteobacteria*. We discuss in the 'Evolutionary considerations' the evolutionary relationship of the *E. coli* *dnaTC* and the p ϕ Rac *ydaUV* initiator-helicase loader gene pair.

The Gram(+) -specific phages with IL-type replication modules encode initiators of the λ O-type but exclusively helicase loaders of the DnaC_{Eco}-type (COM section C3.2.). Also in this group, the phage replication origins reside within the initiator genes (COM section C2.2.). ϕ SM1 provides the rare exception from the 'rule' that the (putative) initiator gene is located directly upstream of the helicase loader gene: both genes are separated by a small intervening ORF (52 residues) of unknown function (Fig. 14). The conserved linkage of initiator/helicase loader genes allows us to assign the function of a helicase loader, e.g. to ϕ 7201 orf5 protein, which gives equally significant hits with the structurally similar IstB-like transposase small subunits in BLAST searches. The initiator/helicase loader gene pairs are embedded in a comparable context as already described

for the I-solo-type replication modules of Gram(+) -specific phages: integrase and phage repressor genes are located upstream of the initiator gene, and transcribed in the opposite direction. Exonuclease/SAP recombination genes of the *recE/recT*- or *erf*-type are in all cases located downstream of the phage repressor genes, and either upstream (ϕ SLT, ϕ PV83, ϕ r1t, ϕ P335, ϕ LL-H, ϕ mv4) or, less frequently, downstream (ϕ EJ-1, ϕ 7201) of the initiator/helicase loader gene pairs. Resolvase genes, mostly *rusA*-type, are invariably found downstream of the initiator/helicase loader gene pairs (Fig. 14). The position of the *ssb* genes is, as observed before, highly variable. An *ssb* gene could not be detected in the ϕ P335 genome, but we note that orf10 (306 residues) upstream of the initiator orf11 has a highly acidic SSB-like C-terminus. Although the IL-type replication modules of the Gram(+) -specific phages is identical to that of the Gram(-) -specific phages with respect to gene composition, there is definitively a higher degree of similarity within each group with respect to the positioning of the exonuclease/SAP recombination genes. Within the group of Gram(+) -specific phages with IL-type modules, a comparison of the smaller ORFs surrounding the replication/recombination genes allows a further differentiation at the level of host species subgroups: one can clearly distinguish between *Staphylococcus*, *Streptococcus* and *Lactobacillus/Lactococcus* phages. In Fig. 14 we have chosen red-striped colouring for ORFs exclusively present in *Staphylococcus* phages (a), green-striped colours for ORFs specific for *Streptococcus* phages (b), and violet-striped colours for ORFs specific for *Lactobacillus/Lactococcus* phages (c). ORFs with homologues in more than one subgroup are indicated by blue-striped colours, and these ORFs are in many cases located distantly to the initiator/helicase loader gene pairs. This indicates that recombination events that lead to viable progeny occur frequently and preferentially among phages with the same host range. Although this remark seems somewhat trivial, it points to a possible reason why the gross architecture of many phages with different replication modules is so strikingly similar: the replication modules are apparently functionally equivalent.

'Initiator-helicase' replication modules

Salmonella phage P22 encodes two essential replication proteins: gene 18, the λ O-type initiator, and gene 12, the DnaB-type replicative helicase. In addition, the ϕ P22 *erf* gene encodes an essential recombination function (Botstein & Matz, 1970). Wickner (1984b) showed that purified ϕ P22 gene 12 protein can replace *E. coli* DnaB helicase in the ϕ X174 DNA *in vitro* DNA synthesis assay, and bypasses the requirement for DnaC. This suggests that during initiation of ϕ P22 replication the gene 18 initiator recruits the gene 12 helicase to the unwound origin by direct interaction.

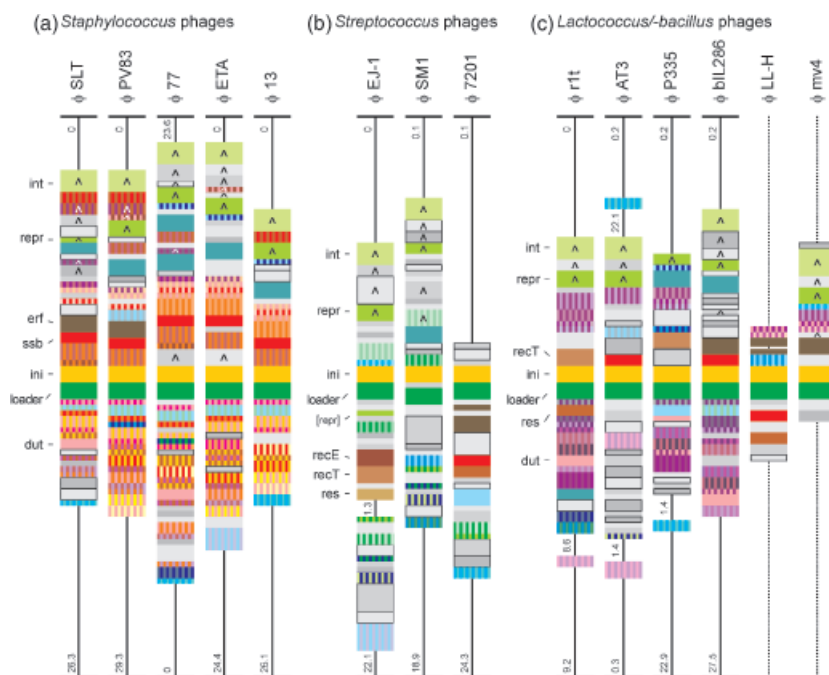


Fig. 14. 'Initiator-helicase loader' replication modules of Gram(+) specific phages. Genes encoding replication and recombination functions are shown in their genomic context. The alignment is shown with homologues/functional analogues of ϕ SLT orf256 (initiator) at fixed positions. Blocks with solid colours indicate (putative) gene functions detected by BLAST comparison: int, integrase; repr, λ *cl*-type phage repressor; erf, recE, recT, single-strand annealing proteins; ini, initiator; loader, DnaC-type helicase loader; res, RusA- and RuvC-type resolvases; ssb, single-strand DNA binding protein; dut, dUTPase. Apparently truncated genes are shown in square brackets. ORFs with significant similarity ($> 30\%$ identical residues) are indicated by striped colouring. Red-striped colouring indicates ORFs conserved within the group of *Staphylococcus* phages (a), green-striped colours indicate ORFs conserved within the group of *Streptococcus* phages (b) and violet-striped colours indicate ORFs conserved within the group of *Lactobacillus/Lactococcus* phages (c). ORFs with homologues in more than one subgroup are indicated by blue-striped colours. Dark and light grey colouring indicates ORFs lacking homologues in any other phage genome. Dark and light grey colouring with black outline indicates ORFs with homologues in (completely sequenced) phages other than compared here. Diagonal stripes indicate chromosomal genes flanking prophage genes. The ORF size is indicated by block height: ≤ 100 residues = 1 U, ≤ 200 residues = 2 U, ≤ 300 residues = 3 U, etc. The relative positions of the ORFs in the phage genomes are indicated by distances in kilobases (not to scale). Except where indicated by ' > ' the direction of transcription is from up to down. The sequences were taken from the genome entries for ϕ SLT (42.9 kb) [NC_002661], ϕ PV83 (45.6 kb) [NC_002486], ϕ 77 (41.7 kb) [NC_005356], ϕ ETA (43 kb) [NC_003288], ϕ 13 (42.7 kb) [NC_004617], ϕ EJ-1 (42.9 kb) [NC_005294], ϕ SM1 (34.7 kb) [NC_004996], ϕ 7201 (35.5 kb) [NC_002185], ϕ r1t (33.4 kb) [NC_004302], ϕ AT3 (39.2 kb) [NC_005893], ϕ P335 (36.6 kb) [NC_004746], ϕ bil286 (41.8 kb) [NC_002667], ϕ LL-H (7 kb, partial sequence) [LLHSSB] and ϕ mv4 (8.7 kb, partial sequence) [AF182207].

Subsequently, origin-bound gene 12 protein attracts the host DnaG primase. ϕ P22 can thus serve as a model for the IH-type replication module that is found in a number of Gram(−)- and Gram(+)-specific (pro)phages.

For the initiator proteins of this group, a specific signature for initiators of IH-type modules could not be detected (COM section C3.1.2.). The helicases in the IH-type modules are invariably DnaB_{Eco}-type helicases (F4 superfamily; see COM section C3.3.). All phage genomes compared in Fig. 15 show a striking similarity in their gross architecture, whether their replication module is located in the left (e.g. ϕ P22, ϕ 11) or right half (e.g. ϕ HK97, ϕ 3626). The architecture closely resembles that of λ (included for comparison in Fig. 15). In all cases, the integrase and phage repressor genes are located upstream of the initiator and helicase genes, and transcribed in the opposite direction.

The replication genes are surrounded by a dazzling diversity of small ORFs of mostly unknown function. Many of these small ORFs are well conserved among several closely related phages – ϕ P22 and its 'cousins' serve as example here – but their positions vary considerably. Other small ORFs lack known homologues, and there are in fact no two phage genomes where the initiator and helicase genes are embedded in an identical context. Remarkably, however, the replication genes are always found directly adjacent to each other, the initiator gene upstream of the helicase gene. This probably reflects a selective advantage of coupled genes encoding interacting proteins in a genetic context that is subject to rearrangements by recombination.

This conserved arrangement of the (putative) initiator and helicase genes in mind, it should not be forgotten that there is considerable variation among the replication genes

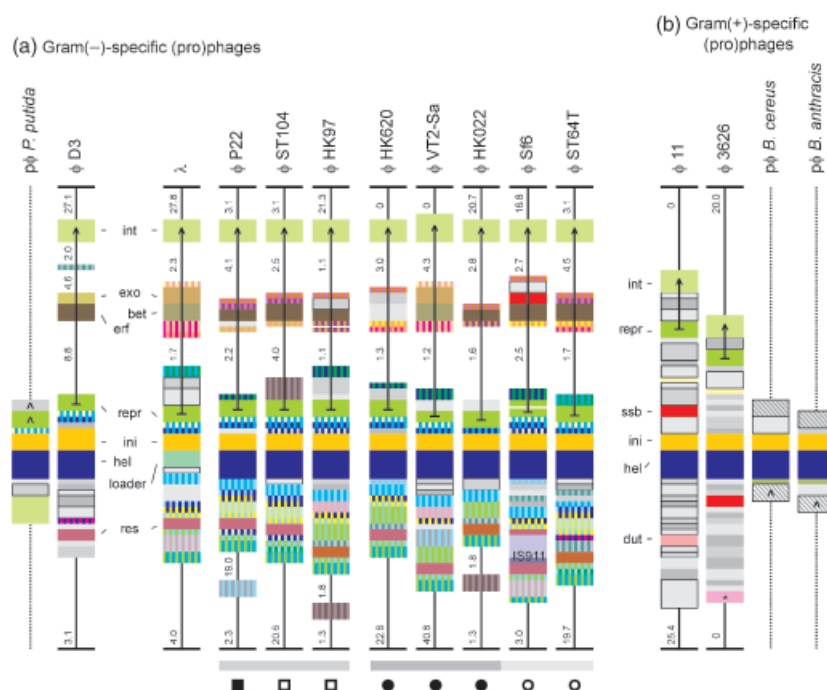


Fig. 15. 'Initiator-helicase' replication modules. Genes encoding replication and recombination functions are shown in their genomic context. λ (IL-type module) was included for better comparison. The alignment is shown with homologues/functional analogues of λ O (initiator) at fixed positions. Blocks with solid colours indicate (putative) gene functions detected by BLAST comparison: int, integrase; repr, λ cI-type phage repressor; exo, *red α* -type exonuclease; bet, *red β* -type single-strand annealing protein (SAP); erf, ϕ P22 *erf*-type SAP; ini, initiator; loader, λ P-type helicase loader; hel, F4-type helicase; res, RsaA- and RapA-type resolvases; ssb, single-strand DNA binding protein; dut, dUTPase. ORFs with significant similarity (> 30% identical residues) are indicated by striped colouring. Dark and light grey colouring indicates ORFs lacking homologues in any other phage genome. Dark and light grey colouring with black outline indicates ORFs with homologues in (completely sequenced) phages other than compared here. Diagonal stripes indicate chromosomal genes flanking prophage genes. Homologues of ϕ 3626 orf50 (pink colour marked with an asterisk) were found in ϕ PVL (orf63), ϕ BK5-T (orf63) (see Fig. 9) and ϕ E125 (gp70) (see Fig. 12). The ORF size is indicated by block height: ≤ 100 residues = 1 U, ≤ 200 residues = 2 U, ≤ 300 residues = 3 U, etc. The relative positions of the ORFs in the phage genomes are indicated by distances in kilobases (not to scale). Except where indicated by ' > ' the direction of transcription is from up to down. The sequences were taken from the genome entries for λ (48.5 kb) [NC_001416], *P. putida* p ϕ (6.18 Mb) [NC_002947], ϕ D3 (56.4 kb) [NC_002484], ϕ P22 (41.7 kb) [NC_002371], ϕ ST104 (41.4 kb) [NC_005841], ϕ VT2-Sa (60.9 kb) [NC_000902], ϕ HK620 (38.3 kb) [NC_002730], ϕ HK97 (39.8 kb) [NC_002167], ϕ HK022 (40.8 kb) [NC_002166], ϕ Sf6 (39 kb) [NC_005344], ϕ ST64 λ T (40.7 kb) [NC_004348], ϕ 11 (43.6 kb) [NC_004615], ϕ 3626 (33.5 kb) [NC_003524], *B. cereus* p ϕ (5.4 Mb) [NC_004722], *B. anthracis* p ϕ (5.2 Mb) [NC_003997]. (a) Gram(-)-specific (pro)phages; (b) Gram(+)-specific (pro)phages. In the lower part, grey blocks indicate different degrees of similarity of the initiators and helicases as discussed in the text (see also Table 2); filled plus open boxes indicate the two different dIR-type iterons and filled plus open circles the two different rep-type iterons in the (putative) phage replication origins, respectively, as discussed in the text.

themselves, including variations of the (putative) replication origins. The almost identical helicases of ϕ P22, ϕ HK97 and ϕ ST104 are only distantly related to the helicases of the other phages from the 'P22+cousins' group, which also form a subgroup of almost identical proteins (Table 2). ϕ P22 on one side and ϕ HK97+ ϕ ST104 on the other side have initiators that are distinct enough from each other to result in different replication origins (Table 2; see also Fig. 15). In addition, the initiators (and replication origins) of phages ϕ Sf6+ ϕ ST64T are clearly different from the initiators of ϕ HK620, ϕ HK022 and ϕ VT2-Sa. In this case, the similarity of the initiators is largely confined to the C-

termini, and the replication origins of both subgroups are different (Table 2, Fig. 15).

Exonuclease/SAP recombination genes in addition to the initiator/helicase gene pair are only found in the genomes of the Gram(-)-specific phages, and should be included in the definition of the IH-type module. The exonuclease/SAP gene pairs are invariably located between the integrase and phage repressor genes. ϕ VT2-Sa encodes homologues of the *red α* /*red β* gene pair, while the other phages encode only an Erf-like SAP. As an exception, ϕ D3 encodes an exonuclease together with an Erf-like SAP. The remarkable conservation of the position of these exonuclease/SAP gene pairs suggests

Table 2. IH-type replication modules of Gram(–)-specific phages: similarities of initiators and helicases.

φ/pφ	Gene	Res.	Accession	φST104	φP22	φHK97	φSf6	φST64T	φHK620	φVT2 – Sa	φHK022	Gene	Res.	Accession	φST104	φP22	φHK97	φSf6	φST64T	φHK620	φVT2 – Sa	φHK022	
Initiator												helicase											
φ ST104	gene 18	273	YP_006383	x								gene 12	458	YP_006384	x								
φ P22	p30	271	NP_059610	51	x							p31	458	NP_059611	98	x							
φ HK97	O	273	NP_037739	98	50	x						P	458	NP_037740	91	90	x						
φ Sf6	gene 43	296	NP_958218	22		22		x				gene 44	478	NP_958219	31	31	31		x				
φ ST64T	gp18	299	NP_720302				82	x				gp12	478	NP_720303	31	31	32	98	x				
φ HK620	O	299	NP_112057				66	65	x			P	478	NP_112058	31	31	31	99	98	x			
φ VT2-Sa	O	299	NP_050525				66	65	98	x		P	483	NP_050526	31	31	31	99	98	99	x		
φ HK022	O	299	NP_037692		21	23	67	66	98	98	x	P	478	NP_037693	31	31	31	99	98	99	99	x	

The values given are percentage identical residues obtained by BLAST bl2seq; Tatusova & Madden (1999) similarity searches. A dash indicates no similarity detected. Boxed values indicate almost complete DNA sequence identity. Stippled boxes indicate significant DNA sequence similarity, particularly in the 3'-part of the gene.

that the two ORFs at the corresponding position in the φHK620 genome also encode recombination functions.

The genes encoding (putative) Holliday junction resolvases are invariably located downstream of the initiator and helicase genes in the genomes of the Gram(–)-specific phages, and at roughly corresponding positions. The resolvases are of two types: RapA-like (e.g. φP22, φVT2-Sa, φD3) and RusA-like proteins (e.g. φST64T, φHK97). This suggests that both resolvase types are functionally equivalent. The *rapA*-like gene in the φSf6 genome is interrupted by insertion element IS911, and does probably not allow the synthesis of a functional protein. This observation suggests that a Holliday junction resolvase is not an essential protein for phage propagation, probably because backup functions exist in the host cell.

φSf6, φ11 and φ3626 encode SSBs but there is no preferential position of the *ssb* gene, as observed also for the other module types. A *ssb* gene is not an essential component of the IH-type replication module but should be considered as an accessory replication gene.

The 'IH-type' modules of prophages in the genomes of *B. cereus* and *B. anthracis* are highly homologous, located at corresponding positions in the genomes of both species, and flanked by orthologous host genes (Fig. 15). This suggests that a prophage was already present in the genome of the ancestor of both species. However, the direct neighbourhood of both initiator and helicase genes shows that prophage degradation proceeded differently after divergence.

Relatively few phages are known from *Streptomyces* sp., and the best studied case, phage C31, belongs to the group of phages that encode DNA polymerases (see 'Phages encoding DNA-polymerases' section). Interestingly, the *Streptomyces* plasmids pSLA2 and pSCL have replication modules, which suggest that they were originally derived from a bacteriophage(s) with IH-type replication module (Chang *et al.*,

1996). The related Rep1 genes of both plasmids contain the replication origins, and the Rep1 proteins show the same domain structure as λ O-type initiators, unusual for plasmid initiators. The Rep2 gene downstream of Rep1 encodes a (putative) F4-type helicase, with Mycobacteriophage CJW1 gp82 as closest homologue (29% identical residues, full length).

'Initiator-helicase loader-helicase' replication modules

The replication of *B. subtilis* phage SPP1 has been studied in detail (see above). The essential replication functions include the G38P initiator, the replication origin *oriL* located within gene 38, the unique helicase loader G39P and the G40P helicase (DnaB_{Eco}-type). The replication of φSPP1 DNA requires, in addition, as essential recombination functions the G34.1P exonuclease and the G35P SAP proteins, and the origin-like structure *oriR* (Ayora *et al.*, 2002). Gene 45 may be a truncated version of a *rusA*-like gene, and a Holliday junction resolvase has not yet been determined as essential for φSPP1 replication. In addition, the G36P SSB is not essential under laboratory conditions. φSPP1 is most closely related to *Listeria* sp. φA118 (see Fig. 16). The apparent differences are due to the fact that the lytic phage SPP1 lacks a detectable repressor gene, and for the integrase gene, only a putative remnant can be found at an unusual position: gene 37 downstream of gene 36 (SSB). φSPP1 is the prototype Gram(+)-specific phage with an ILH-type replication module.

A unique variant of the IHL-type replication module is present in φ315.1 in the *Streptococcus pyogenes* MGAS315 genome: the helicase gene (locus tag SpyM3_0690) is located directly upstream of the initiator-helicase loader gene pair. Because the gene context of this prophage closely resembles the organisation of other phages, recombinatorial

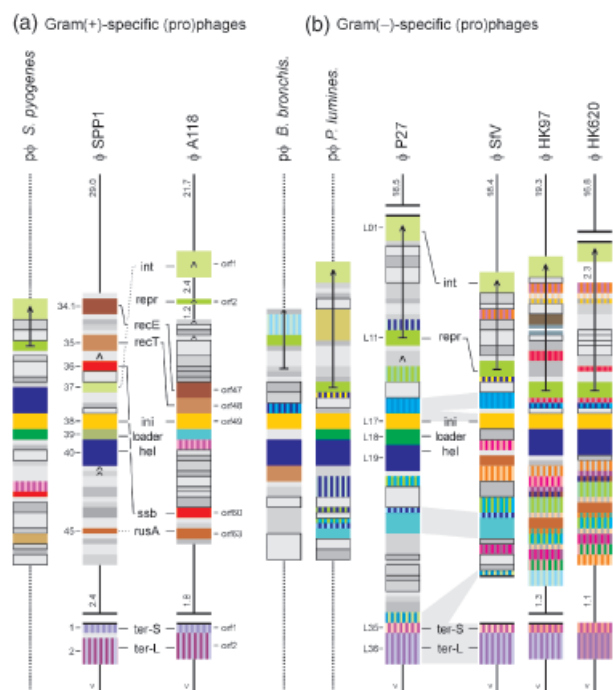


Fig. 16. 'Initiator-helicase loader-helicase' replication modules. Genes encoding replication and recombination functions are shown in their genomic context. ϕ A118 and ϕ SfV (I-solo type module), as well as ϕ HK97 and ϕ HK620 (IH-type module) were included for better comparison. The alignment is shown with homologues/functional analogues of ϕ P27 L17 (initiator) at fixed positions. Blocks with solid colours indicate (putative) gene functions detected by BLAST comparison: int, integrase; repr, λ cI-type phage repressor; recE, recT, RecE-, RecT-type recombinases; ini, initiator; loader, DnaC-type (dark green) or ϕ SPP1 G39P (pale olive) helicase loader; hel, F4-type helicase; rusA, RusA-type resolvase; ssb, single-strand DNA binding protein; terS, terL, terminase subunits. ORFs with significant similarity ($> 30\%$ identical residues) are indicated by striped colouring. Dark and light grey colouring indicates ORFs lacking homologues in any other phage genome. Dark and light grey colouring with black outline indicates ORFs with homologues in (completely sequenced) phages other than compared here. The ORF size is indicated by block height: ≤ 100 residues = 1 U, ≤ 200 residues = 2 U, ≤ 300 residues = 3 U, etc. The relative positions of the ORFs in the phage genomes are indicated by distances in kilobases (not to scale). Except where indicated by ' > ' or by an arrow spanning several ORFs, the direction of transcription is from up to down. The sequences were taken from the genome entries for ϕ HK97 (39.8 kb) [NC_002167], ϕ HK620 (38.3 kb) [NC_002730], ϕ SfV (37.1 kb) [NC_003444], ϕ A118 (40.8 kb) [NC_003216], ϕ SPP1 (44 kb) [NC_004166], *Streptococcus pyogenes* p ϕ (1.9 Mb) [NC_004070], ϕ P27 (42.6 kb) [NC_003356], *B. bronchiseptica* RB50 p ϕ (5.3 Mb) [NC_002927] and *Photobacterium luminescens* p ϕ (5.68 Mb) [NC_005126]. (a) Gram(+) specific (pro)phages; (b) Gram(-) specific (pro)phages.

rearrangements due to the prophage state seem unlikely (Fig. 16). The putative DnaC_{Eco}-type helicase loader (locus tag SpyM3_0692; 167 residues) of p ϕ 315.1 is considerably

shorter at the N-terminus than other helicase loaders of this type, and its functioning uncertain.

Escherichia coli phage P27 is the prototype Gram(-) specific phage with a ILH-type replication module, and encodes a helicase loader of the DnaC_{Eco}-type. The overall architectures of ϕ P27 and ϕ SfV (I-solo type) are very similar, and both phages share (limited) regions with almost perfect DNA sequence homology (Fig. 16; compare grey shaded regions). When the architecture of ϕ P27 is compared with that of ϕ HK97 and ϕ HK620, and with the architecture of ϕ SfV in addition, the emergent picture suggests that all three phages are closely related despite their different replication modules (Fig. 16). Because phages ϕ P27 and ϕ SfV lack detectable DNA sequence similarity directly downstream of the L19 helicase and the orf39 initiator, respectively, it is impossible to address the intriguing question of whether the ILH-type module of ϕ P27 came about by acquisition of a helicase-loader gene (moron; Hendrix *et al.*, 2000) by an IH-type phage, or whether the I-solo module of ϕ SfV came about by loss of the helicase loader/helicase gene pair (lesson) from an ILH-type phage.

Two other (putative) ILH-type modules of prophages could be identified in the genomes of *B. bronchiseptica* RB50 (*Betaproteobacteria*) and *P. luminescens* (*Gammaproteobacteria*), respectively (Fig. 16). Neither of these prophages is closely related to ϕ P27. The putative DnaC_{Eco}-type helicase loader (locus tag plu3472; 194 residues) of the *P. luminescens* prophage is considerably shorter at the N-terminus than other helicase loaders of this type, and its functioning uncertain. The putative helicase loader gene (locus tag bb2208; 123 residues) separating the initiator and helicase genes in the *B. bronchiseptica* prophage lacks homologues in the databases. It may represent another novel type of helicase loader, because it is similar in size to the ϕ SPP1 G39P helicase loader, and too large to be considered simply a recombination remnant.

Conclusions for 'Phages encoding initiator proteins'

We have described in the preceding paragraphs the replication modules of 40 bacteriophages with fully sequenced genomes and, in addition, 21 modules of several partially sequenced phage genomes and of various prophages. At the end of 2004, ~220 bacteriophage genome sequences were available in the databases, and the initiator-encoding phages of the 'lamdoid' type clearly made up a considerable percentage. With respect to statistical significance of the sample, however, their number is still too low to allow for a fair judgement as to whether one of the four module types has a particular selective advantage. In addition, there certainly exists a strong bias in phage sampling, e.g. the

fermentation industries have always had a strong interest in understanding dairy phages, while interest in pathogenicity determinants strongly favoured the study of phages of *Staphylococci* and *Streptococci*. It is perhaps only circumstantial that ILH-type replication modules are the least frequently found among Gram(+)– as well as Gram(–)–specific phages. However, the number of ~60 examples for I-solo, IL- IH- and ILH-type replication modules is sufficient to derive a unified model for a virtual ‘lambdoid’ replication module:

- (1) The (putative) initiator genes always contain the (putative) phage replication origin at a position that corresponds to the linker region separating the N-terminal DNA-binding domain from the C-terminal oligomerisation/interaction domain in the initiator protein.
- (2) In temperate phages, the integrase and phage repressor genes are invariably located upstream and transcribed divergently from the initiator gene. Integrase and/or phage repressor genes may not be present (e.g. ϕ SPP1) or inactive (e.g. ϕ Sf6) in the genomes of lytic phages, but the gene order is comparable with that of the temperate phages.
- (3) Helicase loader genes are (almost) invariably located directly downstream of the initiator genes within the same transcription unit. The genes may encode helicase loaders of the λ P-type, the DnaC_{Eco}-type, the ϕ SPP1 G39P-type or yet unknown types (e.g. ϕ Tuc2009).
- (4) Helicase genes are (almost) invariably located directly downstream of the initiator gene (IH-type module) or the helicase loader gene (ILH-type module) within the same transcription unit. Invariably, these genes encode F4 helicases of the DnaB_{Eco}-type.
- (5) Many phages encode exonuclease/SAP recombination genes, mostly in closely linked gene pairs. Although their absolute requirement has been shown in some cases (e.g. ϕ SPP1), these proteins are considered to have accessory functions because a number of phage genomes lack detectable exonuclease/SAP genes. Despite similarity on the protein level, one can distinguish between the Red α -type and RecE-type exonucleases on the one hand, and the Red β , RecT and Erf SAPs on the other hand. Gram(+)– and Gram(+)–specific phages can be distinguished by the preferential localisation of the exonuclease/SAP genes: they are (almost) invariably located downstream of the repressor gene and upstream of, and transcribed convergently to, the initiator gene in Gram(+)–specific phages, but are not necessarily part of the same transcription unit (e.g. ϕ SPP1). In Gram(–)–specific phages the exonuclease/SAP genes are located between the phage repressor and integrase genes and transcribed convergently with them. The positional conservation of the exonuclease/SAP genes in either phage group further supports the speculation that yet unknown recombination proteins may be identified in several phage genomes (e.g. ϕ HK620).

(6) Genes encoding Holliday junction resolvases are located (almost) invariably downstream of the initiator gene. The encoded proteins may be of the RusA-, RapA- or RuvC-type. These proteins provide accessory functions because several phage genomes lack a detectable resolvase gene. Given the already known diversity of phage-encoded resolvases, different novel types may be discovered in the future.

(7) *ssb* genes with similarity to *E. coli ssb* are found too frequently in ‘lambdoid’ phage genomes to be neglected, but not frequently enough to assume a more than accessory function for phage replication.

This definition of the virtual ‘lambdoid’ replication module would have been impossible to derive without a thorough analysis of the occurrence and the positioning of the high number of mostly small ORFs that separate the genes described above. Moreover, the positional conservation of several of these small ORFs in phage pairs with different replication modules (e.g. λ/ϕ VT2-Sa) was instrumental in developing the model that the four different module types are functionally equivalent although they mediate the link of phage replication to the host replication machinery at different stages of replisome formation.

Phages of the ϕ adh-type (see the subsequent section) encode entirely different replication proteins, including ϕ P4 α -type helicase-primases, but the gross architectures of these phages closely resemble that of λ , in particular with respect to the localisation and orientation of the replication genes to integrase and phage repressor genes. It seems possible that the definition of the virtual ‘lambdoid’ replication module requires an even wider approach than given here.

A unique organisation of replication genes is present in *E. coli* phage P1. This phage is a rare example for a naturally occurring ‘joint replicon’: plasmid replication of the P1 prophage is driven by the R replicon, while replication in the lytic cycle is driven by the L replicon (COM section C3.1.2.; Lobočka *et al.*, 2004). Both replicons are separated physically, use different initiators and require a different subset of host replication proteins. With the RepL initiator and the Ban helicase, the L replicon of ϕ P1 would formally fit to the ‘IH-type’ but we propose to exclude it from this definition because these two replication genes are not part of the same transcription unit. In addition, the gross architecture of ϕ P1 is hardly comparable with that of λ .

An important lesson of our comparison of the different types of ‘lambdoid’ replication modules may come from the observation that within a highly mosaic population an underlying pattern of similarity between individual members only becomes apparent when methods are applied that allow the (quasi-)simultaneous visualisation of the entire data set. Conventional one-by-one comparison does not suffice.

Replication module exchange among bacteriophages

The phenomenon of genetic mosaicism was first detected among close relatives of *E. coli* phage λ , and has been studied most extensively for members of this group (Highton *et al.*, 1990). Among many instructive studies, the genetic analyses of λ reverse and λ altSF show that infecting phages can pick up genes from cryptic prophages in the host genomes by recombination, thus resulting in a replacement of their recombination/replication modules by a different type (Kaiser & Murray, 1979; Friedman *et al.*, 1981). We were particularly interested to study to what extent mosaicism also affects replication functions in phages that are not closely related to λ (by DNA sequence).

Iandolo *et al.* (2002) have studied the genome organisation of the three *S. aureus* phages ϕ 11, ϕ 12 and ϕ 13 in detail. The three temperate phages have a comparable

genome size of ~ 40 kb, a highly similar overall genome architecture and share significant sequence similarity with other *S. aureus* phages, namely ϕ PVL, ϕ SLT, ϕ PV83 and ϕ ETA (Fig. 17). The 'replication module' of ϕ 11, ϕ 12, and ϕ 13 was localised between the 'lysogeny module' and the 'transcriptional regulation module' in one half of the phage genomes (Iandolo *et al.*, 2002). Within the 'replication modules' Iandolo and coworkers found helicase, polymerase and SSB encoding genes but the precise ends of the replication module could not be determined.

The replication modules of the three phages are clearly distinct, implying different replication mechanisms. Phages ϕ 11 and ϕ 13 contain (putative) replication origins for θ DR within their initiator genes (see 'Phages encoding initiator proteins' section). Both phages probably replicate via θ DR, therefore. Concerning the replication module and its flanking regions, ϕ 12 is apparently a chimera of phages most closely related to ϕ SLT on one side, and *Bordetella* sp. phage

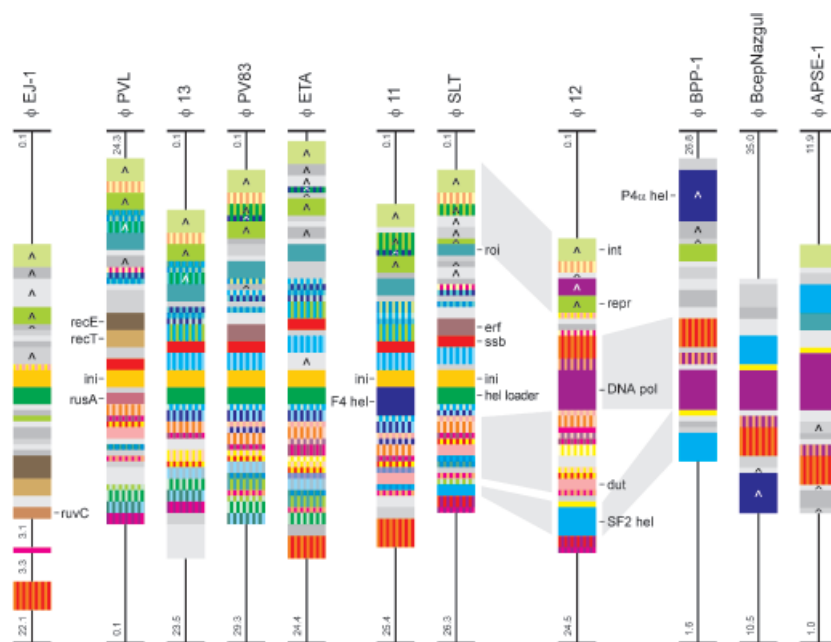


Fig. 17. Replication module exchange: part A. Genes encoding replication and recombination functions are shown in their genomic context. For easier comparison the alignment is shown with homologues of ϕ 12 p12 (DNA polymerase) and ϕ 11 gene 15 (initiator) at fixed positions. None of the proteins has been analysed biochemically and their putative function assigned by BLAST similarity to known proteins. Blocks with solid colours indicate gene functions: ini, initiator; hel loader, DnaC-type helicase loader; ssb, single-strand DNA binding protein; erf, ϕ P22 Erf-type protein; roi, Roi-type phage antirepressor; dut, dUTPase; int, phage integrase; repr, λ cI-type phage repressor; DNA pol, DNA polymerase (Pol I-type); SF2 hel, superfamily 2 helicase; F4 hel, F4 family helicase (DnaB_{Eco}-type); P4 α hel, ϕ P4 α -type primase-helicase; recE, recT, rusA, ruvC, putative recombinases. ORFs with significant similarity (> 30% identical residues) are indicated by striped colouring. Proteins with similarity to ORFs encoded by ϕ 12 are shown in red/yellow striped colours; proteins lacking homologues in the ϕ 12 genome are shown in blue/green striped colours (see text for details). Dark and light grey colouring indicates ORFs lacking homologues in any of the other phage genomes compared here. Light grey shaded blocks point to regions containing homologous ORFs in phages ϕ 12 and ϕ SLT, and in ϕ 12 and ϕ BPP-1. The ORF size is indicated by block height: ≤ 100 residues = 1 U, ≤ 200 residues = 2 U, ≤ 300 residues = 3 U, etc. The relative positions of the ORFs in the phage genomes are indicated by distances in kilobases (not to scale). Except where indicated by ' > ' the direction of transcription is from up to down. The sequences were taken from the genome entries for ϕ APSE-1 (36.5 kb) [NC_000935], ϕ BcepNazgul (58.1 kb) [NC_005091], ϕ BPP-1 (42.5 kb) [NC_005357], ϕ 12 (45 kb) [NC_004616], ϕ SLT (42.9 kb) [NC_002661], ϕ 11 (43.6 kb) [NC_004615], ϕ ETA (43.1 kb) [NC_003288], ϕ 13 (42.7 kb) [NC_004617], ϕ PVL (41.4 kb) [NC_002321], ϕ EJ-1 (42.9 kb) [NC_005294] and ϕ PV83 (45.6 kb) [NC_002486].

BPP-1 on the other (Fig. 17; compare shaded areas). The lack of a functional analogue of the (putative) primase-helicase protein encoded by ϕ BPP-1 leads us to speculate that ϕ 12 replicates by a mode similar to that known from ColE1-type plasmids: a primary ϕ 12 transcript synthesised by the host RNA polymerase is elongated by ϕ 12 DNA polymerase in a unidirectional strand-displacement reaction until the host replication machinery is recruited to replicate fully the phage replicon, probably at a PAS site (del Solar *et al.*, 1998). We discuss in COM section C3.3. the lack of experimental evidence for the participation of SF2-type helicases in replication. As with ϕ BPP-1, ϕ BcepNazgul and ϕ APSE-1, ϕ 12 also codes for a SF2-type helicase (p26). Intriguingly, a SF2-type helicase gene is also present in ϕ SLT, but in an apparently truncated form (116 residues; orf116a). Comparison of the ϕ BPP-1 gene arrangement with that of ϕ 12 suggests that independent recombination events led to the acquisition of the DNA polymerase and SF2 helicase genes by the parent(s) of ϕ 12. However, ϕ BPP-1 was certainly not the direct source of these genes, as judged from the moderate degree of similarity of the proteins (32% identical residues for DNA polymerase, 40% identical residues for SF2 helicase).

Yet another instructive example for replication module exchange is provided by a comparison of phages ϕ 11, ϕ 13 and ϕ PV83. We have shown in COM section C3.1.2. that the proteins encoded by gene 15 of ϕ 11 and ϕ 13, respectively, and the orf20 protein of ϕ PV83 are *bona fide* initiators for θ DR, and contain the (putative) replication origins of these phages. In all three phage genomes, a second (putative) replication gene is located directly downstream of a initiator gene: genes encoding a DnaC_{Eco}-type helicase loader in ϕ 13 and ϕ PV83 (IL-type replication module), and an F4-type helicase in ϕ 11 (IH-type module) (Fig. 17). In all three

phage genomes both (putative) replication genes are embedded in a highly similar gene context.

On the nucleotide level, the comparison of the region encompassing genes 13–18 of phages ϕ 11 and ϕ 13, respectively, suggests that genes 15 and 16 were replaced by a ‘cut+paste’-type recombination event in a common ancestor of both phages (Fig. 18a). However, when the region encompassing ORFs 18–22 of ϕ PV83 is also considered, the emerging picture is considerably more complex: although there is significant sequence homology between the 5′ halves of ϕ 11 gene 15 and ϕ PV83 orf20, the homology between ϕ 13 gene 15 and ϕ PV83 orf20 starts within the 3′-ends and remains uninterrupted until the end of ϕ 13 gene 16 (Fig. 18b, c). This sequence patchwork results in partial protein sequence homologies: whereas the N-terminal 104 residues of ϕ PVL orf20 protein are identical (one mismatch) with ϕ 11 gene 15 protein, the C-terminal 14 residues are identical with ϕ 13 gene 15 protein. The DNA and protein sequence similarity among the three proteins in the ‘middle’ part is low. Note that a DNA sequence patchwork can also be observed for the upstream and downstream neighbouring genes. Therefore, we have to assume several successive recombination events among related phages to explain this mosaicism, which makes it virtually impossible to trace the exact descent of the individual genes.

We derive confidence from the above observations that the concept of phage replication modules has a molecular basis and is not just a useful theoretical tool for the classification of the various replication module types of phages. However, recombination does not necessarily occur exactly between genes, but also at seemingly random points within genes. Subsequent selection of functional recombinants then leads to the impression that entire modules are replaced (Hendrix *et al.*, 2000). If this were true some useful

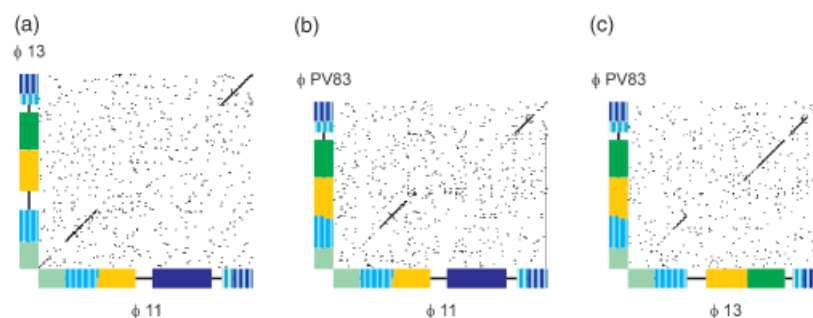


Fig. 18. Replication module exchange: part B. Dot-plot matrix analysis (nucleotide sequences) of the replication module plus adjacent regions of phages ϕ 11, ϕ 13 and ϕ PV83. ϕ 11, positions 8375–12 858, 4483 bp, genes 13–18 [NC_004615]; ϕ 13, positions 8133–12 214, 4081 bp, genes 13–18 [NC_004617]; ϕ PV83, positions 8367–11 867, 3500 bp, ORFs 18–22 [NC_002486]. The dot-plots were obtained using ‘method 2’ (K_tuple value = 8) of the dot matrix subprogram of the DNAMAN[®] software (version 4.0; Lynnon Inc.). (a), ϕ 11 \times ϕ 13; (b), ϕ 11 \times ϕ PV83; (c), ϕ 13 \times ϕ PV83. The coloured blocks indicate the individual proteins encoded by the regions of the phage genomes subjected to dot-plot analysis (exactly to scale). Colouring corresponds to that in Fig. 17; orange, initiator; green, DnaC-type helicase loader; dark blue, F4-type helicase.

information could be derived from a careful elucidation of such mosaics. We have discussed in COM section C3.1.2. that many initiators seem to be composed of a N-terminal DNA-binding domain and a C-terminal oligomerisation domain, which can be interchanged to a certain extent. With their homologous N-termini but only weakly homologous C-termini the $\phi 11$ gene 15 and $\phi PV83$ orf20 proteins belong to this category. More importantly, however, the conservation of the extreme C-termini of $\phi PV83$ orf20 and $\phi 13$ gene 15 proteins suggests that this part of the proteins is responsible for the interaction with the DnaC-type helicase loader encoded by the downstream gene. This is reminiscent of the situation observed for the extreme C-terminus of YdaU in the $p\phi Rac$ YdaUV IL-type replication module (see also COM Fig. C9).

In addition to the IL-type replication modules ($\phi PV83$, ϕETA , $\phi 13$, ϕSLT) and the IH-type ($\phi 11$), also the 'I-solo' type is found in this group of highly related phages: in ϕPVL . Recombination and SSB genes are found in some but not all of the I/IL/IH-phages (Fig. 17). Formally, this suggests that these are accessory rather than essential functions for phage replication (see previous section). To our

knowledge, however, it has not been rigorously examined whether all phages can propagate efficiently in their staphylococcal hosts in the absence of (helper) prophages providing such additional replication functions. Although expression of most prophage genes is repressed in lysogens by the cognate repressor, this situation may change upon infection by a hetero-immune phage (J. Iandolo, pers. commun.).

$\phi P4\alpha$ -type helicase-primase encoding replication modules

The ϕadh replication module was defined experimentally by Henrich and co-workers, who found autonomous replication for plasmids carrying a particular ~ 7 kb ϕadh fragment together with a selectable marker (Altermann *et al.*, 1999). This fragment contained in addition to the putative replication origin (downstream of orf771) the presumed replication genes orf223 (ntp), orf455 (SF2-helicase), orf175 (SSB), orf771 ($\phi P4\alpha$ -type helicase) and four small ORFs encoding polypeptides with unknown functions (indicated by an extended bracket in Fig. 19). Attempts to reduce the size of

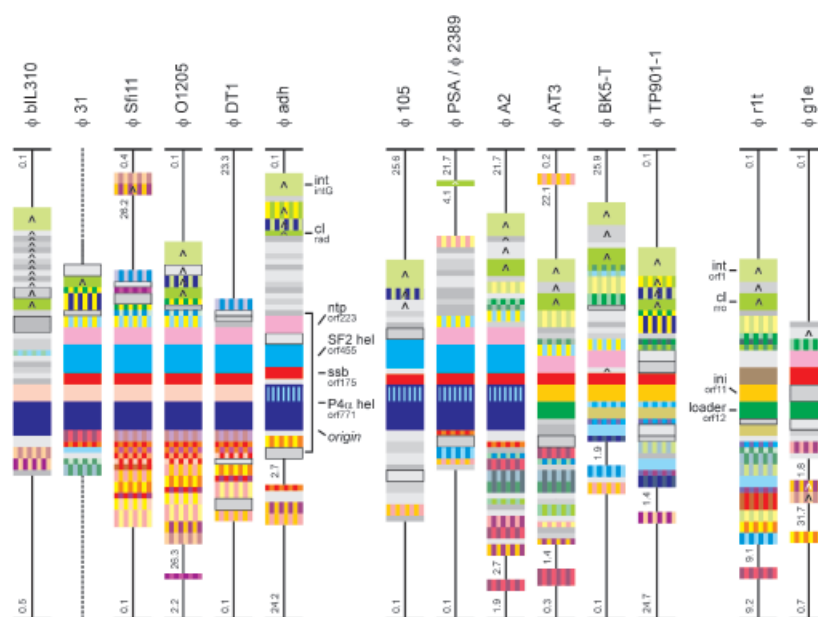


Fig. 19. $\phi P4\alpha$ -type helicase-primase encoding replication modules. Genes encoding replication functions are shown in their genomic context. For easier comparison the alignment is shown with homologues of ϕadh orf771 ($\phi P4\alpha$ -type helicase) at a fixed position. Blocks with solid colours indicate (assigned) gene functions. ORFs with significant similarity ($> 30\%$ identical residues) are indicated by identical striped colouring. Dark and light grey colouring indicates ORFs lacking homologues in any other phage genome. Dark and light grey colouring with black outline indicates ORFs with homologues in (completely sequenced) phages other than compared here. The ORF size is indicated by block height: ≤ 100 residues = 1 U, ≤ 200 residues = 2 U, ≤ 300 residues = 3 U, etc. The relative positions of the ORFs in the phage genomes are indicated by distances in kilobases (not to scale). Except where indicated by ' > ' the direction of transcription is from up to down. The sequences were taken from the genome entries for $\phi bIL310$ (14.9 kb) [NC_002669], $\phi Sfi11$ (39.8 kb) [NC_002214], $\phi O1205$ (43.1 kb) [NC_004303], $\phi DT1$ (34.8 kb) [NC_002072], ϕadh (43.8 kb) [NC_000896], $\phi -105$ (39.3 kb) [NC_004167], $\phi PSA/\phi 2389$ (37.7 kb) [NC_003291], $\phi A2$ (43.4 kb) [NC_004112], $\phi AT3$ (39.2 kb) [NC_005893], $\phi BK5-T$ (40 kb) [NC_002796], $\phi TP901-1$ (37.7 kb) [NC_002747], $\phi r1t$ (33.4 kb) [NC_004302] and $\phi g1e$ (42.3 kb) [NC_004305]. The partially known sequence of $\phi 31$ (10.8 kb) was taken from entry AJ292531.

this fragment were unsuccessful, probably because smaller or partly overlapping fragments lacked appropriate transcription signals, thus pointing to the limitations of the experimental approach (Altermann *et al.*, 1999). By a phage replication interference approach, Moscoso & Suarez (2000) located the (putative) ϕ A2 replication origin at a position corresponding to that of ϕ adh, van Sinderen and co-workers the replication origin of ϕ O1205 (Stanley *et al.*, 2000), Klaenhammer and co-workers the replication origin of ϕ 31 (Madsen *et al.*, 2001), and Brüssow and co-workers the replication origins of ϕ Sfi19 and ϕ Sfi21 (relatives of ϕ Sfi11 shown in Fig. 19) (Foley *et al.*, 1998).

A comparison of the ϕ adh replication module with the corresponding regions of phages ϕ 105, ϕ PSA and ϕ A2 suggests that the small ORFs between ϕ adh orf223 and orf771 are not essential and should not be considered as parts of the ϕ adh-type replication module (Fig. 19). The comparison of the flanking regions upstream of ϕ adh orf223 and downstream of orf771 with the corresponding regions in the other three phages corroborates the experimentally derived definition of the ϕ adh replication module, underlining the usefulness of combined experimental and genomic analyses.

Despite the lack of detectable DNA sequence similarity, the highly similar gene order and the homology of the proteins encoded by ϕ A2 and ϕ AT3 suggests that both phages have a common evolutionary origin (Fig. 19). It appears that the IL-type replication module of ϕ AT3 and the orf35 protein of ϕ A2 are functionally equivalent for replication of the two phages. Further examples of phages with partial similarities of their replication genes and flanking regions to ϕ A2 genes are ϕ BK5-T, ϕ TP901-1 and, to a lesser extent, ϕ r1t (Fig. 19). In all these phages, we could detect a (putative) replication origin in the intergenic region downstream of the gene encoding the ϕ P4 α -type helicase (COM section C2.2.).

Most probably, orf771 protein performs the same function for ϕ adh replication as the ϕ P4 α -protein for ϕ P4 replication, i.e. a combined initiator-helicase function, but experimental evidence for this hypothesis is not available. The C-termini of orf771 and ϕ P4 α -protein are homologous, but the function of the extended N-terminal domain of orf771 is not known (COM sections C3.3. + C3.4.). The set of (putative) replication proteins of ϕ DT1 corresponds to that of ϕ adh except that the ϕ DT1 orf36 protein (504 residues) lacks the extended N-terminus of its homologue, ϕ adh orf771 (771 residues). In addition, ϕ DT1 orf35 upstream of orf36 (pale pink label in Fig. 19) encodes a protein that lacks a homologue in ϕ adh. However, homologues of ϕ DT1 orf35 are invariably found in those phage genomes that encode the shorter variant of the ϕ P4 α -type helicase, i.e. ϕ bIL310, ϕ Sfi11, ϕ 31 and ϕ O1205. In COM Section C3.4., we speculate that the extended N-termini of ϕ adh orf771 and its homologues in ϕ A2, ϕ 105 and ϕ PSA may

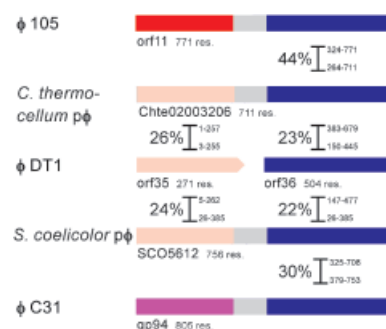


Fig. 20. Putative primase domains of ϕ P4 α -type helicases. The sources are indicated at the left side; gene names (locus tags) are shown underneath together with the sizes of the encoded proteins (not shown to size); the direction of transcription is indicated by arrowheads. The '%' values (BL2seq) shown give the percentage of identical residues for the compared regions of two proteins (Tatusova & Madden, 1999). Colour code: dark blue, ϕ P4 α -type helicase domain plus C-terminal DNA-binding domain; pink/red/violet, different types of (putative) primase domains.

represent yet uncharacterised primase domains. Here we extend this hypothesis, based on the results of BLAST searches, which identified two prophage-encoded N-terminal domains of ϕ P4 α -type helicases with similarity to orf35 (Fig. 20). This makes it likely that also ϕ DT1 orf35 encodes a yet uncharacterised primase. Phage-encoded DnaG-type primases occur either fused to a known helicase domain as in ϕ P4, ϕ C31 and ϕ T7, or as separate polypeptides as, for example, in ϕ KMV or ϕ VpV262 (see below: the ϕ T7-type replication module). It appears that the putative primases follow the same scheme: they occur fused to a ϕ P4 α -type helicase domain as in ϕ 105 or in the *S. coelicolor* prophage gene SCO5612, or as single polypeptides as in ϕ DT1 (Fig. 20). Fusions of a DnaG-type primase to a helicase are found for ϕ P4 α -type and F4-type helicases – but so far not for SF2-type helicases – and therefore seem to be a common theme among phage-encoded replicative helicases (see below). But despite these suggestive observations, the primase function of the ϕ adh orf39 N-terminus and the ϕ DT1 orf35 proteins have to be confirmed experimentally.

BLAST searches readily identify ϕ adh orf223 protein as a member of the AAA-family of NTPases but do not allow function to be predicted. Homologues of the ϕ adh orf223 protein are found in closely related replicon modules (e.g. ϕ A2), more distantly related replication modules (e.g. ϕ DT1), but also in unrelated modules (e.g. ϕ BK5-T). However, an orf223 homologue is not encoded by ϕ 105, which suggests an accessory rather than essential function of orf223 protein for the ϕ adh-type replication module.

ϕ adh orf455 encodes an SF2-type helicase with an as yet unknown function for phage replication. The SSB encoded by ϕ adh orf175 contains the characteristic acidic C-terminus of the ϕ A2 orf34-group of SSBs. These 'Group 3' SSBs

are related to the chromosomally encoded SSBs of bacilli, but probably not entirely interchangeable for functioning during phage replication (COM section C3.6.1.). Homologues of the ϕ adh orf455 SF2-type helicase and orf175 SSB are part of the ϕ adh-type replication module, but neither protein is encoded by ϕ bIL310 (Fig. 19). ϕ bIL310 (15 kb) is probably a satellite phage that depends on a helper phage for its propagation, as with *E. coli* phage P4 (Chopin *et al.*, 2001). It is likely, therefore, that ϕ bIL310 contains only the basic components of an ϕ adh-type replication module: replication origin+initiator function of the orf24 ϕ P4 α -type helicase, and the (putative) orf23 primase.

To conclude this subject, we wish to point out that in some cases the 'module concept' does not promote a better understanding of replication gene assortments than mere BLAST searches:

- (1) ϕ gle encodes a homologue of the ϕ adh orf223 NTP-binding protein (Hel), an unusually large (putative) SSB (224 residues; Rorf224) with similarity to ϕ 31 SSB, and a DnaC-type helicase loader (Ntp) similar to ϕ r1t orf12 protein (Fig. 19) (Kodaira *et al.*, 1997). No putative replication origin structure could be detected in this segment of the ϕ gle genome (not shown), and also a putative initiator gene could not be identified. The comparison of the regions upstream and downstream of these replication genes with the corresponding regions of other phages suggests that this is in fact the ϕ gle replication module, but a possible molecular mechanism for ϕ gle replication cannot be deduced from this highly atypical assortment of replication genes.
- (2) The *Streptomyces* sp. phages C31 and BT1 encode ϕ P4 α -type helicases, and the N-termini of these proteins are *bona fide* primase domains of the DnaG-type (COM section C3.4.), in contrast to ϕ adh orf771. No other genes are present in the ϕ C31 and ϕ BT1 genomes with similarity to the genes of the ϕ adh-type replication module. Both phages encode Pol I-type DNA polymerases (see Fig. 23). Even a distant similarity of the replication modules of these two phages with the ϕ adh-type replication module is hardly detectable, despite the ϕ P4 α -type helicase common to both.
- (3) The mycobacteriophage Barnyard encodes a ϕ P4 α -type helicase, and an SF2-type helicase. There is nevertheless no detectable similarity to the ϕ adh-type replication module: ϕ Barnyard encodes a Pol III-type DNA polymerase, and all replication genes are found at rather large distances from each other.

Phages-encoding DNA polymerases

The phage T4-type replication module

In order to identify a possible ϕ T4-type replication module, we aligned the genomes of phages ϕ T4, ϕ RB69, ϕ RB49, ϕ 44RR2.8t, ϕ Aeh1 and ϕ KVP40, also including the partially

known sequences of *Aeromonas* sp. phages ϕ 25 and ϕ 65 (Fig. 21). Because early and late replication of ϕ T4 involves several and different replication origins, it is not possible to include origins in the module concept for this phage group (Miller *et al.*, 2003). With one exception, all genes encoding the replication and recombination proteins discussed in the 'Recombination-dependent DNA replication' section could be localised in a comparable context: ϕ 44RR2.8t does not encode a homologue of the ϕ T4 UvsX SAP, and a functional analogue has yet to be identified.

There is a striking conservation of gene order, and direction of transcription, in all phage genomes despite highly varying numbers of ORFs separating the genes encoding essential replication/recombination functions. In some cases these intervening ORFs may be simply prediction artefacts and not actually expressed. For many of these ORFs, however, homologous sequences are found encoded by at least one of the other phages at a different genome position. This suggests that many of these ORFs are functional although certainly not essential for phage propagation. It seems possible that the recurring recombination events that are essential for replication of ϕ T4-type phages created these extensive mosaics, but without upsetting the overall gene order. An example of this recombinatorial chaos is the ϕ T4 *alt* gene, which is surrounded by one or more copies of complete and/or partial duplications in ϕ T4, ϕ RB69 and ϕ 44RR2.8t (see Fig. 21; orange label downstream of gp30). The observation that in the ϕ KVP40 genome entire gene blocks have been apparently transposed would be in line with this hypothesis.

We suggest that the ϕ T4-type replication module is composed of two gene clusters with additional 'orphan' genes (see Fig. 21): Cluster 1 contains the genes encoding (in order of transcription) the gp47 and gp46 recombination proteins, the gp45 clamp, the gp44 and gp62 clamp loaders, the gp43 DNA polymerase, the UvsX SAP, the gp41 helicase and the gp61 primase. The size of this cluster ranges from 15 kb (ϕ Aeh1) to ~18.5 kb (ϕ T4), depending on the number and sizes of intervening ORFs. Cluster 2 contains the genes encoding the 5' \rightarrow 3' exonuclease Rnh, the gp59 helicase loader and the gp32 SSB. The size of this cluster is about 3.8 kb. The genes encoding DNA ligase (gp30), UvsW (SF2 helicase) and endonuclease VII (gp49) are located at corresponding positions in all genomes, except for ϕ KVP40, but not in a larger context of other replication/recombination genes. This is also observed for Dda (SF1 helicase), which has a conserved position in the genomes of four of the six fully sequenced genomes. The conservation of the gene order in both gene clusters supports the identification of gene function by BLAST comparisons also for those proteins that have not been characterised biochemically.

Krisch and co-workers proposed a division of the ϕ T4-type phage group into three subgroups based on a sequence

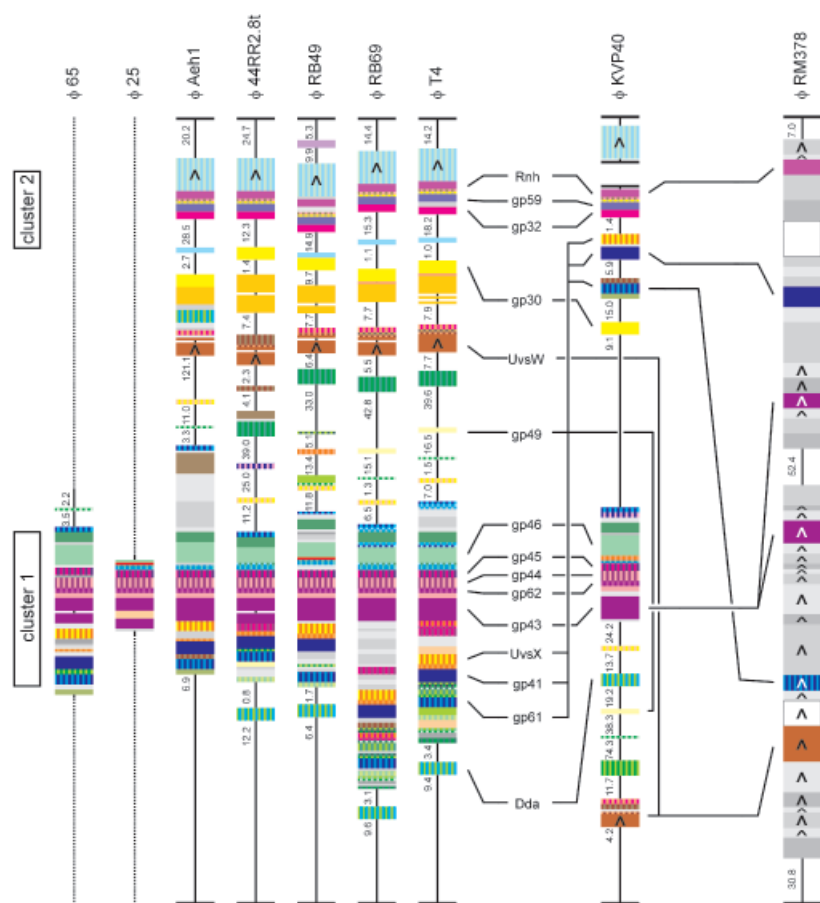


Fig. 21. The ϕ T4-type replication module. Genes encoding replication and (most) recombination functions are shown in their genomic context. For the definition of 'cluster 1' and 'cluster 2' see text. For easier comparison the alignment is shown with homologues of gp43 (DNA polymerase) at a fixed position. ORFs with significant similarity ($> 30\%$ identical residues) are indicated by identical colouring. Dark and light grey colouring indicates ORFs lacking homologues in any of the other phage genomes compared here. White blocks in the ϕ RM378 genome indicate ORFs with similarity to ORFs of (at least) one of the other phages, but not shown in this alignment. The ORF size is indicated by block height: ≤ 100 residues = 1 U, ≤ 200 residues = 2 U, ≤ 300 residues = 3 U, etc. The relative positions of the ORFs in the phage genomes are indicated by distances in kilobases (not to scale). Except where indicated by ' > ' the direction of transcription is from up to down. The sequences were taken from the genome entries for ϕ T4 (169 kb) [NC_000866], ϕ RB69 (167.5 kb) [NC_004928], ϕ RB49 (164 kb) [NC_005066], ϕ 44RR2.8t (173.6 kb) [NC_005135], ϕ Aeh1 (233.2 kb) [NC_005260] and ϕ KVP40 (244.8 kb) [NC_005083]. The partially known sequences for ϕ 25 (7.5 kb) and ϕ 65 (25.2 kb) were taken from entries AY497556 and AY303350, respectively.

comparison of the major head and tail proteins: T-even phages (ϕ T4, ϕ RB69), pseudo T-even phages (ϕ 44RR2.8t, ϕ RB49) and schizo T-even phages (ϕ 65, ϕ KVP40, ϕ Aeh1) (T  tart *et al.*, 2001). The BLAST similarities of the replication proteins discussed in COM section C3. support this subgrouping. However, the conserved direction of transcription and the conserved gene order in both clusters of replication/recombination genes of all eight phages, except for ϕ KVP40, justify the proposal of a 'common' replication module for the ϕ T4-group of phages.

The ϕ T4-type gene replication gene clusters could not be identified unambiguously in the ϕ RM378 genome (Fig. 21). In addition, this phage and its host *Rhodothermus marinus* (*Bacteroidetes/Chlorobium* group) are poorly characterised microbiologically. The p092 DNA polymerase of ϕ RM378 belongs to the Pol II-type DNA polymerases but lacks the canonical $3' \rightarrow 5'$ exonuclease domain, which is encoded by a separate gene (locus tag p024). Like the corresponding ϕ T4 proteins, the ϕ RM378 p019 helicase is a member of the F4-family, and the p101 primase a member of the DnaG-type family. In addition, a $5' \rightarrow 3'$ exonuclease gene (p012) and a SF2-type helicase gene (p104) are present in the ϕ RM378

genome. A homologue of the ϕ T4 UvsX gene was not detected, but p018 encodes a ϕ P22 Erf-like SAP (COM section C3.6.2). Genes encoding a helicase loader, a SSB and clamp+clamp loader proteins with similarity to the ϕ T4-type proteins were not detected. Despite the presence of several ORFs with similarity to predicted or known proteins of ϕ T4-type phages, neither the ORFs flanking the replication/recombination genes nor the gene order in ϕ RM378 suggest a relationship to T-even phages. Clearly, several more phages related to ϕ RM378 would have to be isolated and sequenced before conclusions concerning the relationship of this phage to the enterobacterial T-even phages can be made.

The phage T7-type replication module

T7 is the prototype of the 'T-odd' group of *E. coli* phages, following the traditional nomenclature. This group is 'odd' in several aspects: phages ϕ T1 and ϕ T5 have a genome architecture that deviates significantly from that of ϕ T7, and encode only very few ORFs with limited similarity to ϕ T7 proteins – they certainly do not belong to this group. In

addition, half of the ϕ Xp10 genome codes for structural and host lysis proteins that are related to λ in an arrangement typical for lactococcal phages (Yuzenkova *et al.*, 2003). However, the replication proteins of ϕ T7 are among the best understood examples of bacteriophage replication proteins, and we will refer to the ' ϕ T7-group' in the following for phages with a comparable set of replication genes, irrespective of their classification by systematic criteria.

Replication of ϕ T7 *in vitro* requires the assembly of the phage-encoded proteins SSB (gene 2.5), primase-helicase (gene 4A and 4B) and DNA polymerase (gene 5) at a preformed R- or D-loop, in addition to host thioredoxin as processivity factor. *In vivo*, host RNA polymerase transcribes the ϕ T7 RNA polymerase gene (gene 1), and ϕ T7 RNA polymerase subsequently transcribes the ϕ T7 genome from a (known) set of highly specific promoters. R-loops formed by these transcripts serve as assembly sites for the ϕ T7 replisome. In addition to the cognate RNA polymerase, ϕ T7 codes for a cognate DNA ligase (gene 1.3), a cognate 5' \rightarrow 3' exonuclease (gene 6) and a Holliday-junction resolvase (gene 3).

Figure 22 shows a set of 15 phage genomes, which resemble ϕ T7 with respect to types and arrangement of their replication genes. The alignment shows the considerable variation in number and size of small intervening ORFs with (mostly) unknown functions even for closely related phage pairs with identical gene order of their replication genes (e.g. ϕ T7/ ϕ A1122, ϕ T3/ ϕ YeO3-12, ϕ K1-5/ ϕ SP6). Also, there is no conservation of the order of the replication genes, although BLAST analysis suggested their (near) homology (COM section C3.). To 'distil' the ϕ T7 replication module from this complex picture we have to discuss the replication proteins individually.

The DNA polymerases encoded by all phages belong to the T7 gene 5 subfamily of Pol I-type DNA polymerases, except those of ϕ Xp10, ϕ KMV and ϕ VpV262. The latter are Pol I-type DNA polymerases but they lack the subfamily-specific residues within the signature motifs. The ϕ PaP3 DNA polymerase may be composed of two separate polypeptides: p39 contains the 3' \rightarrow 5' exonuclease and DNA polymerase signature motifs; the function of p32 protein, which is similar to the ϕ T7 gene 5 N-terminus, remains to be studied.

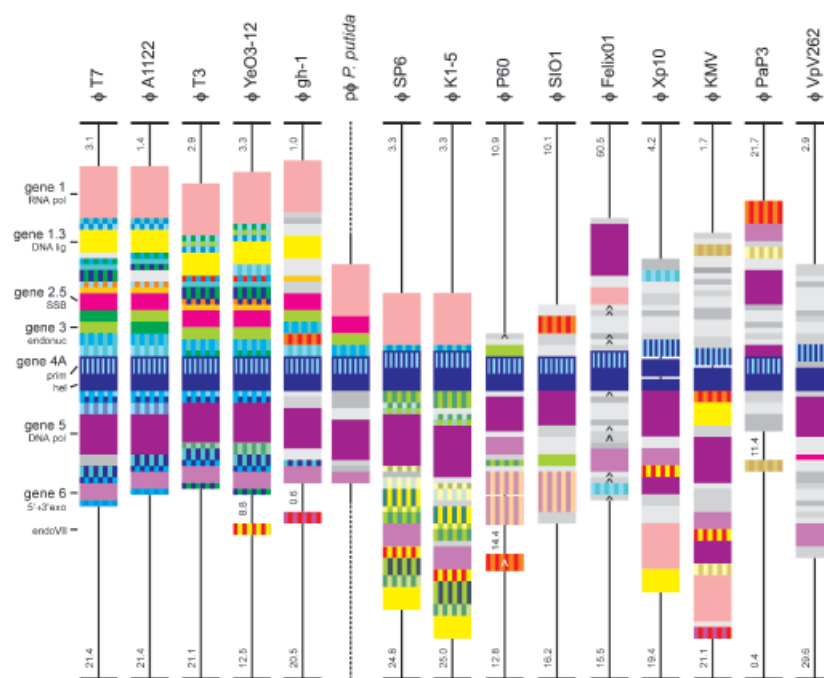


Fig. 22. The ϕ T7-type replication module. Genes encoding replication and recombination functions are shown in their genomic context. For easier comparison the alignment is shown with homologues of ϕ T7 gene 4A (DNA primase-helicase) at a fixed position. ORFs with significant similarity (> 30% identical residues) are indicated by identical colouring. Dark and light grey colouring indicates ORFs lacking homologues in any of the other phage genomes compared here. The ORF size is indicated by block height: ≤ 100 residues = 1 U, ≤ 200 residues = 2 U, ≤ 300 residues = 3 U, etc. The relative positions of the ORFs in the phage genomes are indicated by distances in kilobases (not to scale). Except where indicated by ' > ' the direction of transcription is from up to down. The sequences were taken from the genome entries for ϕ T7 (39.9 kb) [NC_001604], ϕ A1122 (37.5 kb) [NC_004777], ϕ T3 (38.2 kb) [NC_003298], ϕ YeO3-12 (39.6 kb) [NC_001271], ϕ gh-1 (37.4 kb) [NC_004665], ϕ SP6 (43.8 kb) [NC_004831], ϕ K1-5 (44.4 kb) [AY370674], ϕ P60 (47.9 kb) [NC_003390], ϕ SIO1 (39.9 kb) [NC_002519], ϕ Felix01 (86.2 kb) [NC_005282], ϕ Xp10 (44.4 kb) [NC_004902], ϕ PaP3 (45.5 kb) [NC_004466], ϕ KMV (42.5 kb) [NC_005045], ϕ VpV262 (46.0 kb) [NC_003907] and ϕ P. putida strain KT2440 (positions 2 586 633–2 597 674) [NC_002947].

Thirteen of the 15 phages encode a ϕ T7 gene 4A-type primase-helicase. In the ϕ XP10 and ϕ KMV genomes, the primase genes are located directly upstream of the helicase genes – a split helicase gene in the case of ϕ XP10 (see COM Table C16). The primase gene of ϕ VpV262 is located upstream of the helicase gene but separated by a small intervening ORF. The latter three examples demonstrate that although a fused primase-helicase is prevalent among the ϕ T7-type replication modules, different gene-arrangements are possible.

Proteins with similarity to the ϕ T7 gene 2.5 SSB are encoded by six of the 15 phages only. Given its essential role for ϕ T7 replication this is somewhat surprising. It is not known whether the phages that lack a cognate SSB gene recruit host SSB for their replication or whether they encode yet unknown SSB proteins. By simple eye-screening, however, we could not detect ORFs with the characteristic of SSBs, e.g. a highly acidic C-terminus. We propose that a cognate SSB is an accessory rather than essential component of the ϕ T7-type replication module.

Homologues of the $5' \rightarrow 3'$ exonuclease encoded by ϕ T7 gene 6 are found in all phage genomes, except ϕ SIO1. We therefore consider this gene as part of the ϕ T7-type replication module.

Ten of the 15 phages encode their own RNA polymerase (ϕ T7 gene 1). Because we were unable to identify in the remaining phages genes encoding replication proteins indicative of an alternative replication mode, we speculate that these phages use unknown mechanisms to redirect the host RNA polymerase to their own promoters, e.g. a specific σ factor. Until this question is answered experimentally we tend to consider the RNA polymerase gene as part of the ϕ T7-type replication module.

Cognate DNA ligases (ϕ T7 gene 1.3), endonucleases (ϕ T7 gene 3) or proteins similar to ϕ T4 endonuclease VII are encoded only by a subset of the phages compared here. These genes are therefore best classified as accessory functions for the basic ϕ T7-type replication module.

In two instances, we could identify small proteins with significant similarity to proteins encoded by other phages of the set compared here: (1) the ϕ Felix01 p181 protein (266 residues) is considerably shorter than ϕ T7 gene 1 RNA polymerase (883 residues) and therefore probably nonfunctional, and (2) the ϕ VpV262 p21 protein (85 residues) is probably an N-terminally truncated SSB and nonfunctional. From our type of analysis it is impossible to decide whether these apparently truncated genes represent unsuccessful attempts at gene acquisition ('moron') or remnants of a previously complete gene ('lesson') (Hendrix *et al.*, 2000). Cyanophage P60 encodes a small protein (87 residues; locus tag P60_19) with significant similarity to C3-type thioredoxins. As this gene is located between the genes encoding the primase-helicase and DNA polymerase, it is attractive to

speculate that this protein serves as processivity factor for the DNA polymerase, in analogy to *E. coli* thioredoxin for ϕ T7 gene 5 DNA polymerase. It is not known why ϕ KMV and ϕ Xp10 code for a second $3' \rightarrow 5'$ exonuclease, because their DNA polymerases (gp19 and p39, respectively) already contain the typical $3' \rightarrow 5'$ exonuclease domains within their Pol I-type DNA polymerases. The ϕ KMV orf24 and ϕ Xp10 p35 $3' \rightarrow 5'$ exonucleases are similar to each other (44% identical residues) but have no detectable similarity to the $3' \rightarrow 5'$ exonuclease domains of their Pol I-type DNA polymerases.

From the above, we conclude that the 'basic' ϕ T7-type replication module is composed of five genes: (1) a gene encoding a Pol I-type DNA polymerase lacking the $5' \rightarrow 3'$ exonuclease domain of *E. coli* PolA, (2) a gene encoding a $5' \rightarrow 3'$ exonuclease, (3) a gene encoding a DnaG_{Eco}-type primase, (4) a gene encoding a DnaB_{Eco}-type helicase (F4 family) and (5) a gene encoding an RNA polymerase. These five genes are arranged in a gene cluster ~15 kb in length in one half of the phage genome in most cases, arranged in the same direction of transcription, but with highly varying numbers of intervening ORFs. But, in contrast to the ϕ T4-type replication module, no prediction can be made about the gene order because all permutations are observed, the only conserved feature being the gene order: primase-helicase. In principle, this basic set of replication proteins would be able to initiate and drive replication of the phage genome by the mechanism known from ϕ T7. Accessory functions would, when present, increase replication specificity, i.e. render the phage replicon less dependent on the host transcription, recombination and replication machinery.

The phage D29-type replication module

The temperate mycobacteriophages ϕ D29, ϕ L5, ϕ Bxz2, ϕ Bxb1, ϕ Rosebush and ϕ PG1, as well as the *Streptomyces* sp. phages ϕ BT1 and ϕ C31 encode a DNA polymerase with significant similarity to ϕ D29 gp44 (COM section C3.5.). For none of the phages has the replication mechanism been studied in detail. David *et al.* (1992) reported that cloning of a 2.6-kb *Pst*I fragment from ϕ D29 into a selectable plasmid resulted in efficient and stable transformation of *Mycobacterium smegmatis*. They concluded that this fragment carries the ϕ D29 replication origin. Because this fragment carries the intact ϕ D29 gp33 (putative) integrase gene together with the *att* (attachment) site, the observed plasmid stability could have been also due to efficient integration into the *M. smegmatis* chromosome. In order to identify a possible ϕ D29-type replication module, and replication mechanism, we aligned the genomes of the phages for the region flanking the DNA polymerase genes (Fig. 23).

We exclude phages ϕ BT1 and ϕ C31 from the further discussion of a ' ϕ D29-type replication module' for two reasons. The functional equivalent of the primase and

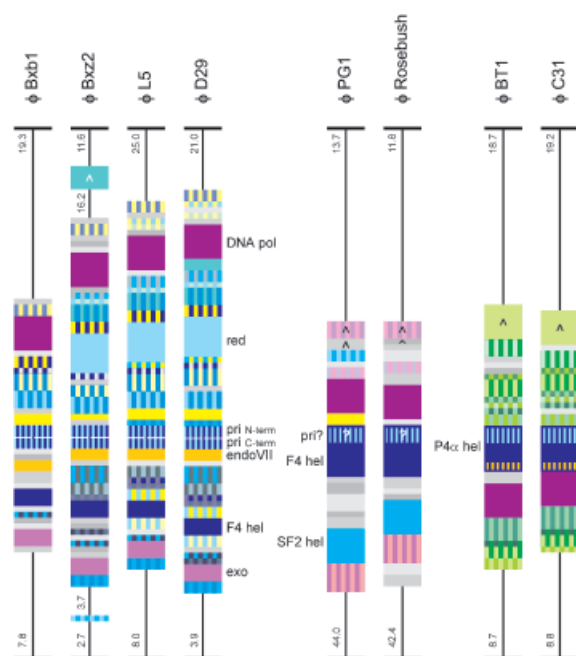


Fig. 23. The ϕ D29-type replication module. Genes encoding replication and recombination functions are shown in their genomic context. For easier comparison the alignment is shown with homologues of ϕ D29 gp57 (DNA primase) at a fixed position. None of the proteins has been analysed biochemically and their putative function assigned by BLAST similarity to known proteins. Blocks with solid colours indicate gene functions: DNA pol, DNA polymerase (Pol I-type); pri N-term, primase, N-terminus; pri C-term, primase, C-terminus; endo VII, endonuclease VII; exo, RecB-type exonuclease; F4 hel, F4 family helicase (DnaB_{Eco}-type); P4 α hel, ϕ P4 α -type primase-helicase; SF2 hel, superfamily 2 helicase; red, ribonucleotide reductase. ORFs with significant similarity ($> 30\%$ identical residues) are indicated by striped colouring. Proteins with similarity to ORFs encoded by ϕ D29 are shown in blue/yellow striped colours; proteins with similarity to ORFs encoded by ϕ PG1 are shown in red/violet striped colours; proteins with similarity to ORFs encoded by ϕ C31 are shown in green striped colours. Dark and light grey colouring indicates ORFs lacking homologues in any of the other phage genomes compared here. The ORF size is indicated by block height: ≤ 100 residues = 1 U, ≤ 200 residues = 2 U, ≤ 300 residues = 3 U, etc. The relative positions of the ORFs in the phage genomes are indicated by distances in kilobases (not to scale). Except where indicated by ' > ' the direction of transcription is from up to down. The sequences were taken from the genome entries for ϕ D29 (49.1 kb) [NC_001900], ϕ L5 (52.3 kb) [NC_001335], ϕ Bxz2 (50.9 kb) [NC_004682], ϕ Bxb1 (50.5 kb) [NC_002656], ϕ BT1 (41.8 kb) [NC_004664], ϕ C31 (41.2 kb) [NC_001978] and ϕ Rosebush (67.5 kb) [NC_004684].

helicase of ϕ D29 are the ϕ P4 α -type primase-helicase proteins of these two phages. This suggests that ϕ BT1 and ϕ C31 replicate by initiator-dependent θ DR, but this remains to be confirmed. In addition, ϕ BT1 and ϕ C31 do not encode a (detectable) endonuclease VII and exonuclease, and also no ORFs with similarity to any of the smaller ORFs of ϕ D29 were found. Therefore, both phages are discussed together

with other phages encoding ϕ P4 α -type primase-helicase (see above).

For ϕ PG1, but not for ϕ Rosebush, we could detect two smaller ORFs with homologues in ϕ D29. In addition, these two phages lack detectable endonuclease VII and exonuclease genes, and they encode a SF2-type helicase, in contrast to ϕ D29. It is not known whether the very large F4-type helicases of these phages contain an N-terminal primase domain (COM section C3.4.). Although it may seem somewhat arbitrary, we exclude these two phages from the discussion of the ' ϕ D29-type replication module'.

The remaining four phages are closely related, even on the DNA sequence level (Bruessow & Desiere, 2001) (see Fig. 23). Other phage replicons with a split primase gene are not known, and all four encode a particularly small version of an F4-type helicase (COM sections C3.3. + C3.4.). When the ' ϕ D29-type replication module' is examined for the genes flanking the replication/recombination genes, it appears that the ϕ Bxb1 genome contains the 'minimal version': the region from the gp41 DNA polymerase gene to the gp62 exonuclease gene spans 10.2 kb, in contrast to 15.5 kb in the ϕ D29 genome. In all four phage genomes, the gene order of the replication/recombination genes and the direction of their transcription are conserved.

When the set of replication/recombination genes of ϕ D29 is compared with the ϕ T7-type replication module, a striking similarity becomes apparent. Both contain a Pol I-type DNA polymerase, a DnaG_{Eco}-type primase and an F4-family helicase. In addition, the ' ϕ D29-type replication module' contains an endonuclease VII with similarity to the ϕ Xp10 p36 protein, and a (putative) ribonucleotide reductase with similarity to the ϕ SIO1 p21 protein – the latter not being a replication protein in the strict sense. However, ϕ D29 lacks a cognate RNA polymerase and SSB, and the (putative) 5' \rightarrow 3' exonuclease is clearly nonhomologous, being more related to RecB-type exonucleases [COG 2887]. There is thus no reason to differentiate between a ' ϕ D29-type replication module' and a ' ϕ T7-type replication module', and we propose to include the former in a more relaxed definition of the latter. Clearly, the replication mechanism of ϕ D29 has to be analysed experimentally to justify this classification.

The replication modules of the phages K, Bxz1 and T5

Escherichia coli phage T5, mycobacteriophage Bxz1 and *Staphylococcus* sp. phage K are among the relatively few (known) phages with large genomes (> 100 kb) that are, with respect to the types of replication proteins they encode, not closely related to the T-even phages. Like the ϕ T4-type phages, however, all three encode DNA polymerases, primases, and F4- as well as SF2-helicases. The observation by

electron microscopy of multiple origins and branched concatemeric structures of ϕ T5 replication intermediates suggested a replication mechanism similar to ϕ T4 (Bourguignon *et al.*, 1976). The replication mechanisms of ϕ K and ϕ Bxz1 are not known.

In the ϕ T5, ϕ K and ϕ Bxz1 genomes the replication genes are arranged in a ~ 15 kb (ϕ T5, ϕ Bxz1) to ~ 25 kb (ϕ K) segment together with (putative) recombination genes and with ORFs of unknown function (Fig. 24). ϕ T5 and ϕ K encode Pol I-type DNA polymerases (the ϕ K sequence became available after completion of COM section C3.5.). Unlike most other Pol I-type polymerases the ϕ T5 DNA polymerase does not require additional factors for processivity. The processivity factor requirements of the ϕ K orf86/88/90 DNA polymerase are unknown. ϕ Bxz1 encodes a Pol III-type DNA polymerase. Genes encoding the proofreading activity ($3' \rightarrow 5'$ exonuclease) and the $5' \rightarrow 3'$ exonuclease have not yet been identified in the ϕ Bxz1 genome. Given the high degree of conservation of these proteins it seems likely that ϕ Bxz1 recruits the respective host proteins for replication. In addition, genes encoding DNA polymerase accessory proteins could not be identified in the ϕ Bxz1 genome (Pedulla *et al.*, 2003). It is presently not known whether the three polymerases perform strand-displacement synthesis or assemble into dimeric replisomes for coupled leading- and lagging-strand synthesis. All three phages encode F4-type (replicative) helicases and DnaG-type primases. ϕ Bxz1 gp192 located upstream of the gp193 helicase gene encodes a DnaC-type helicase loader, an arrangement also found in ϕ P27. In addition to the F4-type helicases, all three phages encode SF2-type helicases. The second SF2-type helicase of ϕ T5 (locus tag T5.108) shows significant similarity to the UL9 helicase involved in replication initiation of Herpes simplex virus 1 (HSV1) but its role for ϕ T5 replication is unknown. As noted in COM section C3.3., the presence of SF2-type helicase genes in so many phage replicons – in many cases located within the replication modules – makes it necessary to understand the role of this helicase for the replication process better. ϕ K orf24 and ϕ T5 *rmh* encode the cognate $5' \rightarrow 3'$ exonucleases for primer removal. Both phages encode DNA ligases, although the assignment of ϕ K orf21 as DNA ligase is questioned by the observation that this protein is similar to the ϕ RB69 RnlB RNA ligase. A putative ligase gene was not detected in the ϕ Bxz1 genome.

ϕ K orf70 encodes a protein for which the assignment 'putative Rep protein' was chosen, but experimental results confirming this role are not available (O'Flaherty *et al.*, 2004). Rep has a (predicted) helix–turn–helix (HTH) motif in its N-terminus but shows no similarity to known phage or plasmid initiators. We could detect a pronounced AT-peak approximately in the middle of the orf70 gene but iterons could not be identified. ϕ T5 encodes the DNA-binding protein D5 that has attracted some attention in the

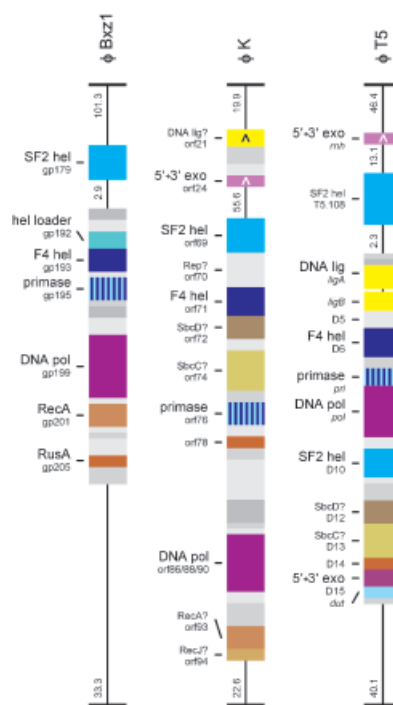


Fig. 24. The ϕ K, ϕ T5 and ϕ Bxz1 replication modules. Genes encoding replication and recombination functions are shown in their genomic context. Few of the proteins have been analysed biochemically; in most cases, their putative function could be assigned by BLAST similarity to known proteins. Gene names/locus tags are shown below the (assigned) functions. Dark and light grey colouring indicates ORFs lacking homologues in any of the other phage genomes compared here. The ORF size is indicated by block height: ≤ 100 residues = 1 U, ≤ 200 residues = 2 U, ≤ 300 residues = 3 U, etc. The relative positions of the ORFs in the phage genomes are indicated by distances in kilobases (not to scale). Except where indicated by ' > ' the direction of transcription is from up to down. The sequences were taken from the genome entries for ϕ K (127.4 kb) [NC_005880], ϕ T5 (121.8 kb) [NC_005859] and ϕ Bxz1 (156.1 kb) [NC_004687].

past. Although D5 binds to ssDNA, it binds with higher affinity to dsDNA, which it covers stoichiometrically. D5 is required for the regulation of late transcription in ϕ T5, but a role for replication has also been proposed. The DNA binding properties of D5 make a role for this protein as replication initiator unlikely. D5 actually inhibits ϕ T5 DNA polymerase *in vitro* possibly through direct interaction (Fujimura & Roop, 1983). Interestingly, none of the three phages encodes a (detectable) SSB. However, there may exist more than the known SSB types (COM section C3.6.1.).

The replication modules of ϕ Bxz1, ϕ K and ϕ T5 provide further examples for the co-localisation of (putative) recombination genes and replication genes in phage replicons. In most cases a direct participation of cognate recombination proteins in phage replication has not been demonstrated. Their exact enzymatic function remains uncertain and their assignment is based mostly on BLAST similarities.

However, the clustering of recombination and replication genes suggests that mechanisms related to RDR are operating during replication of most phages. Although the similarity of ϕ K orf93 protein to RecA proteins is weak, ϕ Bxz1 gp201 is one of the rare examples of a phage-encoded RecA homologue (COM section C3.6.2.). Whether ϕ K orf94 protein is a (putative) σ -factor (O'Flaherty *et al.*, 2004) or a member of the RecJ-family of recombination proteins as also suggested by BLAST searches remains to be seen. ϕ Bxz1 gp205 shows all signature motifs of a RuvA-type Holliday junction resolvase, and also the ϕ K orf78 and ϕ T5 D14 proteins give BLAST hits with Holliday junction resolvases. ϕ K and ϕ T5 encode proteins with significant similarity to the SbcDC-type recombinases. ϕ T5 encodes, in addition, the D15 exonuclease with high affinity for fork structures.

From the phylogenetic viewpoint ϕ K, ϕ T5 and ϕ Bxz1 are, at best, only distantly related to each other, and also only distantly related to the T-even phages despite the comparable genome sizes. In addition, their hosts belong to different branches of the bacterial kingdom. However, all three phages possess highly similar replication modules with respect to the set of replication/recombination genes they encode. If only the replication module is considered, a hypothetical common ancestor of all DNA polymerase-encoding phage replicons could be envisaged that carried genes for a primase, a replicative helicase and a DNA polymerase. The highly differentiated types of extant replicons would then reflect gene replacement, e.g. in the case of the different DNA polymerase types, and acquisition of additional genes during evolution, e.g. genes encoding processivity factors, RNA polymerase genes and *ssb* genes. Alternatively, we could assume that the fortuitous colocalisation of a DNA polymerase gene of any of the three known polymerase types and an F4- or ϕ P4 α -type primase-helicase on a DNA string creates the potential for its autonomous replication. According to the latter model, DNA polymerase-encoding phage replicons could have evolved independently several times during evolution. Also in this model, the different types of extant replicons would reflect the acquisition of additional genes and their eventual exchange by recombination. Whether the over-simplified version of a tree-like phylogeny of phage replicons is more inspiring for future research than the concept of a web-like phylogeny remains to be seen.

The phage ϕ 29-type replication module

As discussed in detail in the 'Initiation at the ends of linear DNA: protein-primed DNA replication', ppDR of the *B. subtilis* phage ϕ 29 requires the cognate DNA polymerase (gene 2 protein), the TP (gene 3 protein), the cognate SSB (gene 5A) and the DNA-binding protein (DBP; gene 6 protein) (reviewed in Meijer *et al.*, 2001). The genes encod-

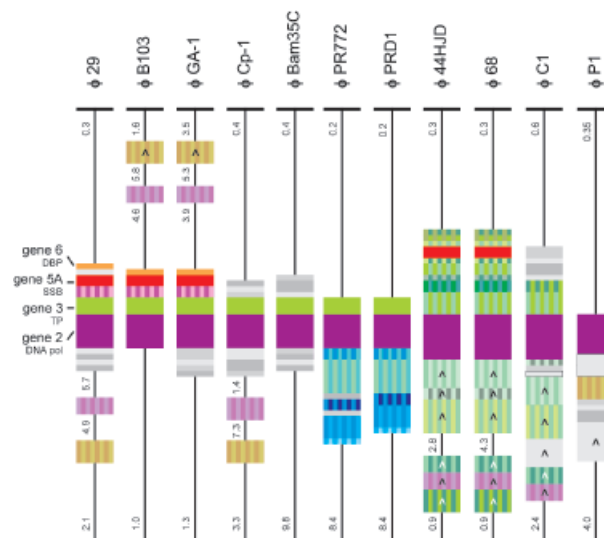


Fig. 25. The replication module of the ϕ 29-type phages. Genes encoding replication functions are shown in their genomic context. ORFs encoding proteins with known replication functions are shown in solid colour. Striped colours indicate similar proteins in the phages. Dark and light grey colouring indicates ORFs lacking homologues in any other phage genome. Dark and light grey colouring with black outline indicates ORFs with homologues in (completely sequenced) phages other than compared here. The ORF size is indicated by block height: ≤ 100 residues = 1 U, ≤ 200 residues = 2 U, ≤ 300 residues = 3 U, etc. The relative positions of the ORFs in the phage genomes are indicated by distances in kilobases (not to scale). Except where indicated by ' > ' the direction of transcription is from up to down. The sequences were taken from: ϕ 29 (19.4 kb) [NC_001423], ϕ B103 (18.6 kb) [NC_004165], ϕ GA-1 (21.1 kb) [NC_002649], ϕ Cp-1 (19.3 kb) [NC_001825], ϕ Bam35C (14.9 kb) [NC_005258], ϕ PR772 (14.9 kb) [AY441783], ϕ PRD1 (14.9 kb) [NC_001421], ϕ 44HJD (16.8 kb) [NC_004678], ϕ 68 (18.2 kb) [NC_004679], ϕ C1 (16.7 kb) [NC_004814] and ϕ P1 (11.7 kb) [NC_002515] (not to be confused with *Escherichia coli* phage P1).

ing these replication factors constitute the replication module of ϕ 29 together with the ends of its linear genome. We could detect 10 proteins with similarity to the ϕ 29 DNA polymerase that are encoded by phages with linear dsDNA genomes, ranging in size from 12 to 21 kb (COM section C3.5.1.). It is, however, difficult to trace the other replication proteins of ϕ 29 in the entire set of 11 phages (Fig. 25). The TP gene can be detected in the majority of the sequences, but homologues of the ϕ 29 DBP gene could only be detected in ϕ B103 and ϕ GA-1. Therefore, we cannot discuss the possible variations of the ppDR mechanism driving ϕ 29 replication in detail.

Replication modules of phages replicating by RCR

The replication modules of the phages that replicate by RCR are composed of: (1) an initiator gene, (2) a double-strand

origin (*dso*) and (3) a single-strand origin (*sso*) (see 'Initiation by nicking: 'rolling circle'-type DNA replication' section).

We can distinguish four groups of initiator proteins, which, despite poor overall similarity, have the conserved 'active tyrosine' motif 3 in common and perform identical functions for the replication of their cognate replicons. The initiators of ϕ X174, ϕ fd, ϕ P2 and CTX ϕ have been studied in detail (COM section C3.1.1.).

The known *ssos* contain a nick-site for the initiation of replication, and a region to which the initiator binds (COM section C2.1.). We can distinguish three different types of initiator-binding sites: (1) an array of repeats (ϕ fd), (2) a region of ~ 25 bp with pronounced dyad symmetry and the potential to form a stemloop (CTX ϕ) and (3) a stretch of ~ 30 bp lacking any detectable sequence or structural motifs (ϕ P2, ϕ X174). Although a particular type of *sso* is recognised by a particular type of initiator, the structural basis for this interaction is presently not well enough understood to derive rules from it for reliable predictions. There are four known localisations for the *sso* with respect to the initiator gene: (1) in the 5'-part of the initiator gene (ϕ X174), (2) in the 3'-part of the initiator gene (ϕ P2), (3) in the intergenic region 5'-upstream of the initiator gene (ϕ fd) and (4) in an intergenic region elsewhere in the phage genome (CTX ϕ).

There are three known structures in the single-stranded form of these phages that can serve as *sso*: (1) a secondary structure 'mimicking' a promoter (ϕ fd), (2) a secondary structure with the quality of a primosome-assembly site (ϕ X174) or (3) a primase binding-site (ϕ G4). The co-localisation of *sso* and *dso* in the filamentous *E. coli* phages fd, f1 and M13 is exceptional: in all other known systems, *sso* and *dso* are not linked. For the phages with mid-sized genomes that replicate by RCR, the ϕ P2 B-type helicase loader should be considered as part of the replication module. However, the exact mechanism of complementary-strand synthesis of ϕ P2 is presently not known, and other cognate proteins might be involved in addition to B (Liu *et al.*, 1993). This topic requires further research.

Phage replicons lacking replication protein genes

Comparing the number of ~ 220 completely sequenced phage genomes with the number of phages discussed in this review, we realise that for approximately 30 – excluding the phages with RNA genomes – no putative replication protein genes could be identified by comparison with known examples. Among the phages lacking known replication protein genes are the 'relatives' of *E. coli* phage Mu and a number of small phages with ssDNA genomes that may also replicate via transposition, e.g. *Spiroplasma* ϕ 1-R8A2B. The replication of the mycobacteriophages Che9c, Corndog and

Omega remains enigmatic, and the 'incomplete' sets of replication genes in the phages Barnyard and Rosebush do not provide a clue at present (Pedulla *et al.*, 2003). Probably the most intriguing 'white spot' is the *Pseudomonas aeruginosa* phage KZ with the largest phage genome known so far (280 kb) (Mesyanzhinov *et al.*, 2002).

The identification of the replication origins of lactococcal phages belonging to the groups of ϕ bIL67-, ϕ 923- and ϕ c2-type phages underlines the importance of established experimental strategies for gaining insight into the replication mechanism of any phage under study (COM section C2.3) (Rakonjac *et al.*, 2003). These phages do not encode cognate replication proteins. Their replication depends instead on the synthesis of an untranslated transcript that 'initiates' replication via tDR by host factors. Thus, also the replication of phages can occur by a mechanism that is known for a long time from the ColE1-type plasmids. Further search for phage replication genes and mechanisms will have to take into account that the failure to identify replication genes by protein homology/similarity to known examples calls for experiments to determine the host factors required for replication of a phage under study.

Evolutionary considerations

Bacteriophages present a wider spectrum of replication mechanisms than bacterial plasmids or chromosomes and an impressive variety of different types of enzymes that perform particular steps during replication. When, in addition, the variability of *related* phage replication modules is considered, it becomes immediately clear that the evolution of phage replicons cannot be discussed in depth in the context of this review. Therefore, we confine ourselves to a discussion of two particularly interesting types of replication proteins: helicases and helicase loaders.

We first discuss possibilities to identify the replicative helicase in phage genomes that encode more than one helicase (see 'The different types of phage-encoded helicases' section). We then discuss the evolutionary origin of phage-encoded homologues of the *E. coli* DnaB helicase proposed by Moreira (2000) (see 'Phage-encoded homologues of the *E. coli* DnaB helicase' section). Lastly, we present a hypothesis on the evolutionary origin of bacterial helicase loaders: the DnaC-type helicase loader of several *Gammaproteobacteria*, and the DnaB1 and DnaD helicase loaders of the bacillales (see 'Chromosomally encoded homologues of phage helicase loaders' section).

The different types of phage-encoded helicases

Our survey of 87 completely sequenced (pro)phage genomes (dsDNA) that encode one or more helicases revealed the presence F4-type helicases in 50 of them, either as the only

helicase (39) or together with a SF2-type helicase (11). ϕ P4 α -type helicases were found less frequently (20), and in the majority of the cases in combination with a SF2-type helicase (13). We could not find an example for a phage genome that encodes a F4-type helicase together with a ϕ P4 α -type (primase-)helicase. Altogether, 40 phage genomes were found to encode SF2-type helicases. SF1-type helicases were only found in the group of ϕ T4-like phages (6). This confusing scenario of different types of helicases encoded by bacteriophages provokes the questions: (1) whether an underlying pattern exist for the occurrence of particular types of helicase gene(s) within the specific set of replication genes in a given phage replicon, and (2) whether such a pattern allows us to pinpoint the replicative helicase, i.e. the helicase associated with the replication fork during the elongation step of replication.

As the basis for an answer to these two questions, we collected from the fully sequenced (dsDNA-)phage genomes that encode helicases the data for initiators (COM section C3.1.), helicase loaders (COM section C3.2.), primases and DNA polymerases (COM section C3.5.). We present in Fig. 26 a scheme that includes examples for all detected variants of replication gene assortments, omitting only the SF1-type Dda helicases of phages from the ϕ T4 group. For completeness, we added the data for λ and ϕ A118, neither of which encodes helicases.

There are two possible approaches to answer our initial questions: (1) a strict approach demanding that every (putative) replicative helicase is experimentally analysed for this property prior to a decisive statement, and (2) a more relaxed approach that allows us to classify and hypothesise on the basis of reliable experimental results obtained only for a subset of the systems to be compared. The first approach is inherently less error-prone and therefore more attractive. We cannot neglect the second approach, however, because it is better suited to promote a deeper understanding of the fundamental biological process of replication by allowing the prediction of interesting model systems for experimental studies. This becomes particularly important when one considers that only a small percentage of the many phage replicons known to date will ever be analysed by genetic or biochemical experiments.

The 'strict approach' gains support when the helicases of eukaryotic viruses are also taken into consideration – in a brief survey. The helicase subunit (UL5) of the trimeric UL5-UL8-UL52 primase-helicase complex of HSV1 is a member of the helicase superfamily 1 (SF1). The UL9 protein of HSV1, which is responsible for origin recognition and unwinding together with the UL29 (ICP8) SSB, belongs to the helicase superfamily 2 (SF2) (Marintcheva & Weller, 2001a,b). However, the SF1- and SF2-type helicases have probably a higher degree of similarity in their three-dimensional structure than the comparison of their primary



Fig. 26. Phage-encoded helicases in the context of other replication proteins. Helicase genes detected in the (completely sequenced) genomes of the listed phages are shown by their gene/gene product/locus tag names. The three helicase/helicase domain types are shown by different colouring (see colour code in Fig.). Initiators, helicase loaders, primases and DNA polymerases detected in the genomes of the listed phages are included to allow for a comparison of the phage-encoded replication functions. Gene products are not shown to size, and the gene order is not reflected in the figure. Bars indicate the failure to detect the respective genes by signature motif and BLAST searches of known helicases and polymerases, respectively. For initiators, helicase loaders and primases, a blank indicates the failure to detect one of the respective genes by comparison with one of the known types. The corresponding *Escherichia coli* proteins are shown on top to allow for an easy comparison. The role of the *E. coli* YehH protein is not known; it has been included here solely for completeness. Phage genomes NCBI accession numbers are: ϕ A118 [NC_003216], λ [NC_001416], ϕ P22 [NC_002371], ϕ SPP1 [NC_004166], ϕ MAV1 [NC_001942], ϕ Bxz1 [NC_004687], ϕ KMV [NC_005045], ϕ T7 [NC_001604], ϕ D29 [NC_001900], ϕ PG1 [NC_005259], ϕ K [AY176327], ϕ T4 [NC_000866], ϕ RM378 [NC_004735], ϕ P4 [NC_001609], ϕ C31 [NC_001978], ϕ BPP-1 [NC_005357], ϕ Barnyard [NC_004689], ϕ Sfi21 [NC_000872], ϕ Bcep1 [NC_005263], ϕ T1 [NC_005833] and ϕ N15 [AF064539]. The genome sizes of the phages are shown in the rightmost column.

sequence would suggest (Bird *et al.*, 1998; Korolev *et al.*, 1998). Several viruses employ for their replication superfamily 3 helicases (SF3), for which no members were detected in phage genomes. Examples include the Rep40 helicase of adeno-associated virus 2 (AAV2), the helicase domain of simian virus 40 (SV40) T antigen (James *et al.*, 2003) and the E1 origin-binding protein of human papillomavirus (HPV) (Masterson *et al.*, 1998). Several virus-encoded SF2-type helicases are involved in the replication of viral RNA genomes, e.g. hepatitis C virus (HCV) NS3

helicase (Lam *et al.*, 2003). We could not detect (by BLAST searches) proteins with significant similarity to the F4-type helicases of prokaryotes and bacteriophages in the genomes of eukaryotes and their viruses, except for some candidate proteins in genomes of chloroplasts and mitochondria. Also for the ϕ P4 α -type (primase-) helicase type, reasonably similar proteins were not detected in virus genomes, except for a protein of unknown function of the *Ectocarpus siliculosus* (marine brown alga) EsV-1 virus. The application of more refined bioinformatic methods than the crude BLAST approach would be necessary to ascertain a relationship of several (putative) pox virus helicases with the ϕ P4 α helicase domain. Apparently, the already complex pattern of different helicases encoded by phage replicons is only a part of the puzzle, and one can hardly repress the notion that virtually all types of helicases can be adapted to the specific requirements of a particular step in a nucleic acid metabolic pathway, e.g. to the role as replicative helicase.

The relaxed approach would start from the background of the classical 'staged initiation' model for chromosome replication in *E. coli* that positions the loading step of the replicative helicase DnaB at the interface of open complex formation and replisome formation (Kaguni & Kornberg, 1984; Kornberg & Baker, 1992). *Escherichia coli* DnaB (F4-type) is the prototype replicative helicase of prokaryotes with orthologues in all sequenced bacterial genomes; its participation in recombination processes (e.g. branch migration) has been suggested only recently (Kaplan & O'Donnell, 2002). It is therefore reasonable to assume that all phage-encoded DnaB homologues function as replicative helicases for their cognate replicons; this has been shown experimentally for ϕ T4 gp61, ϕ T7 gene4A and ϕ SPP1 G40P. The α protein is the replicative helicase for ϕ P4 replication (Briani *et al.*, 2001). The orf382 and orf504 proteins of ϕ Sfi21 and ϕ Sfi11, respectively, lack the N-terminal primase domain of their homologue, ϕ P4 α . Presumably, the ϕ Sfi21 orf382-type proteins function as combined 'initiator-helicase' (COM section C3.1.2.). There is presently no phage genome known that encodes an F4-type helicase together with a ϕ P4 α -type (primase-) helicase, and we therefore hypothesise that phages encode for one particular type of replicative helicase only. The replication protein sets encoded by the phages ϕ KMV and ϕ K differ only with respect to the presence/absence of a SF2-type helicase (see Fig. 26). Together, these observations suggest that the phage-encoded SF2-type helicases are unlikely to function as replicative helicases, despite their importance for phage replication, e.g. for a switch to RDR. Although the sequence similarity between the phage-encoded UvsW-type helicases and the bacterial PriA homologues is very low, it may be informative that PriA, the 'initiator' of the *E. coli* restart primosome, is also a SF2-type helicase (Marians, 1996) (see 'Replication restart' section).

The 'staged initiation' model postulates that the origin-bound replicative helicase DnaB (preprimosome) recruits the primase DnaG for the synthesis of the leading-strand primer (primosome). A physical interaction of DnaB and DnaG could be shown *in vitro* for the *E. coli* system (Tougu & Marians, 1996), and also that the binding of ϕ SPP1 G40P helicase to ssDNA is stabilised by the addition of host DnaG (Ayora *et al.*, 1998). Replisome assembly then occurs at the DNA · DnaB · DnaG complex, involving yet another set of multiple protein interactions with Pol III holoenzyme subunits. The tight interaction of replicative helicase and primase is particularly important for the repeated priming of Okazaki fragments during co-ordinated leading- and lagging-strand synthesis by the replisome. The finding of primase domains fused to helicase domains, e.g. ϕ T7 gene 4 protein and ϕ P4 α , emphasises the importance of the interaction of the replicative helicase with the primase. There is presently only one example of a SF2-type helicase fused to a primase domain: the ϕ N15 RepA protein and its homologues in ϕ KO2, ϕ PY54 and ϕ VP882. The modular architecture of the ϕ N15 RepA protein is strikingly similar to that of ϕ P4 α , as is its role for ϕ N15 replication, but the primase and helicase functions have yet to be established (Ravin *et al.*, 2003). Also, it is not known precisely to what extent ϕ N15 replication is independent of host functions.

A tendency becomes apparent if the working hypothesis is accepted that the monofunctional SF2-type helicases do not function as the replicative helicases of bacteriophages: replicons that code for a helicase and, in addition, a DNA polymerase also encode a primase/primase domain, but phage replicons devoid of a DNA polymerase gene are also lacking a primase gene. This observation holds for phage replicons encoding F4-type helicases and ϕ P4 α -type helicases, likewise. Probably the assortment of replication genes simply reflects the degree of dependence of a given phage on host replication proteins:

- (1) Phage replicons lacking a cognate helicase gain access to the host replication machinery by attracting the host replicative helicase through interaction with their initiator (ϕ A118) or helicase loader (λ). The host replicative helicase then attracts the host primase and DNA polymerase.
- (2) Phage replicons encoding a helicase gain access to the host replication machinery by attracting the host primase through interaction with their helicase, after loading of the latter to the phage replication origin by interaction with the initiator (ϕ P22) or the helicase loader (ϕ SPP1). The phage helicase · host primase complex then recruits the host DNA polymerase.
- (3) Phage replicons that encode a replicative helicase, primase and DNA polymerase are independent from the host replication machinery for elongation (ϕ T4, ϕ T7, ϕ C31).

With bacteriophages, there are always exceptions from rules or patterns (quoting the complaint of Casjens 2003): ϕ P4 and ϕ N15 encode a primase as domain of the α or RepA proteins, respectively, but no DNA polymerase. Therefore, we cannot claim that a 'helicase pattern' exists, and prefer to talk of a tendency, rather. This tendency is weakened, however, by the fact that for three phages shown in Fig. 26 the primase domains can only be called 'putative', at best. These phages encode F4-type (ϕ PG1) or ϕ P4 α -type (ϕ BPP-1, ϕ Barnyard) helicases with extended N-termini that lack similarity to known protein sequences, except to other closely related phage proteins. In analogy to the ϕ T7 gene 4 and ϕ P4 α primase-helicase fusion proteins, we speculate that these extended N-termini represent yet unknown primase domains. Experimental evidence in support of this speculation is not available. It should be noted that, for example, the RepB' primase of plasmid RSF1010 also lacks any sequence similarity with DnaG_{Eco}-type primases, showing that much has still to be learned about the existing primase families.

The 'relaxed approach' cannot provide a conclusive model at present, but it allows us to suggest experiments: promising candidates for a test of whether SF2-type helicases can act as replicative helicases for their cognate replicons are ϕ N15 RepA, ϕ Bcep1 gp66 and ϕ T1 gene 22 (see Fig. 26). Recent results indicate that the SF2-type RepA helicases are indeed the replicative helicases of ϕ N15 and the related phage ϕ PY54 (Mardanov *et al.*, 2004; Ziegelin *et al.*, 2005). ϕ Bcep1 has not yet been studied genetically, and the available experimental data for ϕ T1 replication are somewhat contradictory: the phage can be propagated in *E. coli* *dnaB*(ts) and *dnaC*(ts) hosts (Bourque & Christensen, 1980) but ϕ T1 mutants with an inactivated gene 22 could be isolated (cited in Roberts *et al.*, 2004). In addition, ϕ T1 replication requires a functional DnaG primase of its host, although gene 24 encodes a DnaG homologue (Bourque & Christensen, 1980).

Finally, there is no apparent correlation between the type of helicase(s) and the type of DNA polymerase encoded by phage replicons. However, for a thorough discussion we would have to consider also the processivity factors. For example, it has been shown for *E. coli* that all three DNA polymerases can interact with the DNA polymerase III β -sliding clamp, and form a replication fork together with the DnaB helicase through multiple molecular interactions (Lopez de Saro & O'Donnell, 2001). At present, the processivity factors have only been characterised for the ϕ T4 and ϕ T7 replicons, while the ϕ T5 DNA polymerase possesses high intrinsic processivity. Virtually nothing is known about the possible interaction of ϕ P4 α -type helicases with DNA polymerases. Therefore, we have to leave this point open.

Phage-encoded homologues of the *E. coli* DnaB helicase

Although a review does not usually allow the authors to present their own data as part of the discussion, we would like to point out that the phylogenetic tree for DnaB homologues presented by Moreira based on a maximum-likelihood analysis seems highly questionable: a ϕ T7/ ϕ P22/ ϕ HK022 group splits from the *E. coli* branch very early, and then differentiates into the ϕ T7/ ϕ HK022 and ϕ P22 sub-branches; the branching points of *E. coli*/ ϕ P1 and ϕ T7/ ϕ HK022 are at roughly equal distance from the common origin (see Fig. 1 in Moreira, 2000). Our analysis of signature motif conservation among F4-type helicases suggests, in contrast, that the P and gene 12 proteins of phages ϕ HK022 and ϕ P22, respectively, belong to the DnaB subfamily whereas the gene 4A helicase of ϕ T7 is the prototype of a distinct subfamily (compare COM Tables C11 and C14). We performed a BLAST sequence comparison of individual domains of DnaB homologues analysed by Moreira based on the known domain structure of *E. coli* DnaB (Biswas & Biswas, 1999). In this comparison, we included the site within the N-terminus of the β domain, which is responsible for the primary interaction with *E. coli* DnaA and several plasmid initiator proteins (Datta *et al.*, 1999; Seitz *et al.*, 2000). We found the highest degree of conservation for the C-terminal DNA-binding γ domain (signature motifs 2–4), followed by slightly lower conservation of the nucleotide-binding β domain (signature motifs 1 and 1a) (Table 3). Similarity in the N-termini was in general \sim 20% lower than in the two C-terminal domains. We were not surprised to find the lowest degree of conservation for the primary site of DnaA–DnaB interaction (residues 154–210). The still significant degree of similarity between ban and DnaB in this region certainly contributes to the successful substitution of DnaB by ban described by Lemonnier *et al.* (2003). But all other phage-encoded helicases (compared in Table 3) encode, at best, distantly related initiator proteins, which in turn would require a specific adaptation of the helicase, assuming that the same site is responsible for interaction.

All compared helicases showed the lowest degree of similarity to ϕ T7 gene 4A helicase, and similarity was only detectable in the C-terminal DNA-binding domain. Although it is difficult to determine the evolutionary rate of phage-encoded genes exactly, this suggests that the divergence of ϕ T7 gene 4A from the other phage-encoded helicases occurred before the divergence of the latter from *E. coli* DnaB and *B. subtilis* DnaC. Clearly, the P and gene 12 proteins of ϕ HK022 and ϕ P22, respectively, are more closely related to DnaB_{Eco} than to ϕ T7 gene 4A protein, in support of the subfamily grouping by signature motifs (see also Ilyina *et al.*, 1992). We can confirm, however, the result of Moreira that Ban is evolutionary closer to DnaB of its host

Table 3. Cross-comparison (BLAST) of family 4 helicases by their domains

Replicon	Gene	<i>Escherichia coli</i> <i>dnaB</i>				ϕ P1 <i>ban</i>				<i>Bacillus subtilis</i> <i>dnaC</i>				ϕ SPP1 gene 40				ϕ HK022 P				ϕ P22 gene 12				ϕ P27 L19			
		1–153	154–210	211–302	303–471	1–135	136–191	192–284	285–454	1–136	137–189	190–281	282–454	1–132	133–181	182–247	248–442	1–137	138–193	194–305	306–478	1–124	125–173	174–273	274–458	1–126	127–179	180–276	277–463
<i>Escherichia coli</i>	<i>dnaB</i>	x	x	x	x																								
ϕ P1	<i>ban</i>	73	66	85	82	x	x	x	x																				
<i>Bacillus subtilis</i>	<i>dnaC</i>	39	38	48	55	36	35	51	55	x	x	x	x																
ϕ SPP1	gene 40	28	–	31	49	29	–	34	52	30	–	48	63	x	x	x	x												
ϕ HK022	P	50	–	37	45	44	–	37	42	37	–	31	39	25	–	–	38	x	x	x	x								
ϕ P22	gene 12	23	–	36	49	–	–	36	47	–	–	38	41	–	–	38	40	30	–	29	38	x	x	x	x				
ϕ P27	L19	–	–	38	44	–	–	36	43	–	–	37	40	–	–	39	37	25	–	36	32	28	–	51	49	x	x	x	x
ϕ T7	gene 4A	–	–	–	23	–	–	–	28	–	–	–	22	–	–	–	26	–	–	–	–	–	–	–	–	–	–	–	25

Domains of the helicases were compared by the bl2seq program using the blastportion with default settings [gap open: 11; gap extension: 1; gap \times dropoff: 50; expect: 10.0; word size: 3; no (low complexity) filter] Tatusova & Madden (1999). The compared domains were adjusted to match the known domain structure of *Escherichia coli* DnaB: residues 1–153 = N-terminal α domain; residues 154–210 = interaction site with DnaA Seitz *et al.* (2000) and with several plasmid initiators Datta *et al.* (1999), residues 211–471 = β domain containing the signature motifs 1 plus 1a of the F4-type helicases; C-terminal γ domain containing the signature motifs 2, 3 and 4 of the F4-type helicases (Hall & Matson, 1999; Biswas & Biswas, 1999). Values shown are percentage identical residues; a dash indicates no detectable similarity (< 20% identical residues).

E. coli than ϕ SPP1 G40P is to the DnaC replicative helicase of its host *B. subtilis*.

Despite disputable branching points, the (theoretical) evidence that DnaB homologues were acquired from their enterobacterial hosts independently by phage replicons related to the extant phages ϕ P1 on one side, and ϕ Sf6, ϕ P22, and ϕ P27 on the other side, is compelling. It also seems reasonable to conclude that the ancestor of ϕ SPP1 acquired gene 40 from a Gram-positive host. The root of the 'DnaB tree', however, requires revision, including not only the ϕ T7 gene 4A-type but also the ϕ D29 gp65-type and ϕ T4 gp41-type helicase subfamilies. To improve the value of such phylogenetic studies, we suggest including available data on protein domain architecture, conserved structural and functional motifs, and experimentally defined interaction sites, rather than relying exclusively on the raw protein sequences.

Interestingly, ϕ HK022 P protein, ϕ P22 gene 12 protein and ϕ P27 L19 protein seem to be at roughly equal evolutionary distance from each other (Table 3). BLAST analysis is therefore not sensitive enough to detect subtle differences in the primary sequence of the helicases that would allow us to predict the protein-specific mechanism for helicase loading: ϕ P27 encodes a DnaC_{Eco}-type helicase loader (COM section C3.2.), but not ϕ HK620. In addition, ϕ P22 does not encode a cognate helicase loader, and Wickner (1984a, b) was able to show that loading of gene 12 helicase to ssDNA *in vitro* does not require DnaC_{Eco}.

Chromosomally encoded homologues of phage helicase loaders

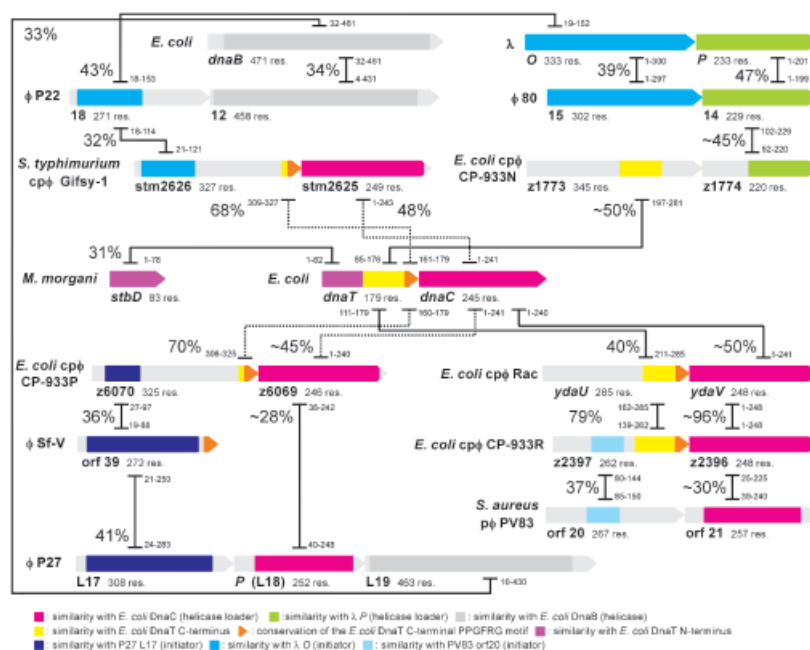
We have discussed in COM section C3.2. the significant similarities between the two primosomal proteins encoded

by the *E. coli* *dnaTC* gene pair and the initiator plus helicase loader proteins encoded by the *ydaUV* replication module of the *E. coli* K12 Rac prophage, respectively (Wrobel & Wegrzyn, 2002). Because of the elaborate functional interplay of DnaT and DnaC with other replication proteins the *dnaT* and *dnaC* genes are considered housekeeping genes of *E. coli* (see 'Replication restart' section). Several lines of evidence lead us to speculate, however, that a progenitor of *E. coli* acquired the *dnaTC* gene pair approximately 10⁸ years ago from a bacteriophage replicon, i.e. by horizontal gene transfer (HGT):

(1) Initiator plus helicase loader gene pairs are common in replication modules of Gram(–)-specific phages replicating by θ DR (see 'Initiator-helicase loader' replication modules' section). The propagation of these phage replicons depends strictly on the recruitment of the replicative helicase of their hosts. Conversely, the chromosomal replicons of Gram-negative bacteria have to compete with invading phage replicons for the replicative helicase. The acquisition of a phage-encoded helicase loader by the host chromosomal replicon could potentially improve its fitness for competition with an invading phage replicon. Although a newly acquired helicase loader would be already shaped for optimal functioning with the host helicase through a long-lasting coevolution of phage and host, it would have to replace the established molecular mechanisms for helicase loading in order to confer any selective advantage to the host replicon. Therefore, the acquisition of the helicase loader gene, *dnaC*, can only be one aspect of why the progenitor of *E. coli* kept the *dnaTC* gene pair.

(2) DnaT shows significant similarity with the C-terminus of YdaU (see 'Bacteriophage replication modules' section) but no characteristics of a phage initiator protein in its

Fig. 27. Similarity network for the *Escherichia coli* *dnaTC* genes. The bacterial and (pro)phage genes are shown roughly to size with the source indicated on the left side; the direction of transcription of the genes is indicated by arrowheads; gene names (locus tags) are shown underneath together with the sizes of the encoded proteins. The '%' values shown indicate the percentage of identical residues for the compared regions of two proteins.



N-terminus (Fig. 27; see also COM sections C2.2. + C3.1.2.). To account for this observation, we speculate that a fortuitous recombination event in the *E. coli* progenitor created a translational fusion between a chromosomal gene and the 3'-half of a prophage-encoded initiator gene, *dnaT**. The chromosomal recipient gene had the potential to interact with one of the PriABC primosomal proteins, possibly PriA (Sandler, 2000). Interestingly, a gene was found in the genomes of two enteric bacteria (not in *E. coli*) encoding a small protein that shares in its N-terminus ~30% identical residues with the DnaT N-terminal half (Hayes, 1998). The presumed recombination event resulted in *dnaT* in its present form, relieved the transcription of the newly created gene fusion from the control of the prophage repressor, and eliminated – by deletion of the *dnaT** 5'-half – the replication origin of the prophage, activity of which would otherwise be detrimental to the host. The novel fusion protein DnaT could recruit the replicative helicase DnaB to PriA-primosomes at stalled replication forks by interaction with its cognate helicase loader, DnaC. We assume that this statistically highly improbable recombination event became genetically fixed because it resulted in a selective advantage for the chromosomal replicon, adding considerably to the improvement of competition fitness by the acquisition of a helicase loader alone. Note that these extensive speculations are in principle amenable to experimental analysis.

(3) Orthologues of the *E. coli* *dnaTC* gene pair are found exclusively in the genomes of enterobacteriales, and within this subfamily of the gammaproteobacteria so far only in the genomes of all *E. coli*, *S. flexneri*, *S. enterica*, *K. pneumoniae* and *B. aphidicola* strains, but not in others, e.g. *Yersinia* sp.

or *P. luminescens* (Thomson *et al.*, 2002; Duchaud *et al.*, 2003). Interestingly, the former four species, but not *Yersinia* sp., belong to one of three clusters of enteric bacteria that could be grouped by their specific combinations of aromatic amino acid synthesis enzymes, and are therefore believed to represent very closely related species (Ahmad *et al.*, 1990). The presence of a *dnaTC* gene pair in the genomes of very few closely related species makes it likely that their common ancestor had acquired it. It is less likely, by contrast, that an ancestral *dnaTC* gene pair was lost from the chromosomal replicons of the other gammaproteobacteria during their evolution.

(4) According to phylogenetic studies, *S. enterica* and *E. coli* diverged as distinct species ~1.2–1.6 × 10⁸ years ago (Ochman & Wilson, 1987). The *dnaTC* gene pair is embedded in a larger well-conserved context in the chromosomes of both species. This makes it likely that the genome of their common ancestor already contained the *dnaTC* gene pair. The alternative, that one species has inherited the *dnaTC* gene pair from the other by HGT at some later time, seems less likely.

(5) Given the great timely distance, it is not surprising that, except for the *dnaTC* gene pair, all other parts of the presumed ancient prophage have been eliminated from the *E. coli* genome in its present version. In addition, differences in the G+C content or the codon usage between the presumed prophage-encoded *dnaTC* genes and the chromosome have been reduced. Accordingly, the *dnaTC* gene pair escaped detection in analyses aimed to identify recent acquisitions of the *E. coli* genome, and which were successful in the cases of Rac, other prophages or the *lac* operon

(Lawrence & Ochman, 1998; Canchaya *et al.*, 2003; Casjens, 2003). Several lysogenic phages of *E. coli* integrate into its chromosome preferentially at or close to tRNA genes, and many phage-derived genes are found at such sites (Lawrence & Ochman, 1998). Intriguingly, the *dnaTC* gene pair is located at a distance of ~5 kb from the *leuVPQ* genes encoding tRNAs.

We were encouraged to speculate about the evolutionary origin of the *E. coli dnaTC* gene pair by the observation that also the *B. subtilis dnaBI* gene pair has probably a similar origin, together with the *dnaD* gene. The details of the molecular mechanisms for helicase loading are strikingly different in both species, but the DnaT and DnaC proteins of *E. coli* and the DnaB, DnaD and DnaI proteins of *B. subtilis* perform analogous functions as primosomal proteins (COM section C3.2.). As discussed in 'Phages encoding initiator proteins', the initiator genes of several θ -replicating (pro)phages of bacilli show significant sequence similarities in their C-terminal half with the C-termini of the DnaB protein (e.g. ϕ BK5-T), the DnaD protein (e.g. ϕ SM1) or both proteins (e.g. ϕ A118) of their specific hosts. Also, (putative) helicase loaders, which show significant similarity with *B. subtilis* DnaI, are frequently found in initiator plus helicase loader replication modules of *Bacilli* phages. We speculate that the *dnaBI* gene pair of *B. subtilis* was acquired from a prophage because it resembles the initiator plus helicase loader gene pairs found in phage replicons. We cautiously extend this hypothesis to the *dnaD* gene, which resembles a phage initiator gene. As we found for *dnaT* in the *dnaTC* gene pair of enteric bacteria, the *dnaB* and *dnaD* genes of *B. subtilis* lack the main characteristics of an origin-

containing initiator gene (e.g. 'AT-peak', iterons) although both proteins bind to DNA (Marsin *et al.*, 2001).

Only the C-termini of the DnaB and DnaD proteins of *S. aureus* contain a stretch of reasonable similarity, but not those of *Listeria* or *Bacillus* species. In addition, the conservation of the *dnaBI* and *dnaD* genes in the bacilli genomes is considerably lower than that of several other replication proteins, e.g. DnaA, DnaC_{Bsu}. It is therefore difficult to decide whether the *dnaD* gene was acquired independently from the *dnaBI* gene pair or whether it is a duplicated *dnaB* gene. Although the replication modules of bacilli phages and prophages frequently encode helicase loaders with similarity to DnaI, in all these cases the similarity of the respective phage initiator with either DnaB or DnaD was low or not detectable. Owing to this complex pattern of mutual relatedness – reminiscent of mosaicism – no simple line of descent can be drawn (Fig. 28).

All sequenced genomes of the bacillales subfamily of the bacilli (i.e. *Bacillus* sp., *Listeria* sp., *Staphylococcus* sp., *Oceanobacillus iheyensis*) contain in a fairly well-conserved genomic context a *dnaBI* gene pair, and – unlinked to it – a *dnaD* gene directly upstream of the *nth* gene encoding endonuclease III. In addition, all sequenced genomes of the second bacilli subfamily, the lactobacillales, contain a *dnaBI* gene pair in a genomic context, which is reasonably similar to that in the bacilli genomes. The situation is different for the *dnaD* genes of lactobacillales: (1) in the first group (*Enterococcus faecium*, *E. faecalis* V583, *Streptococcus epidermidis*, *S. pyogenes* MGAS8232, *L. lactis*) the *dnaD* genes are located next to the *nth* genes, as in the genomes of bacillales; (2) in the second group (*S. pneumoniae* TIGR4,

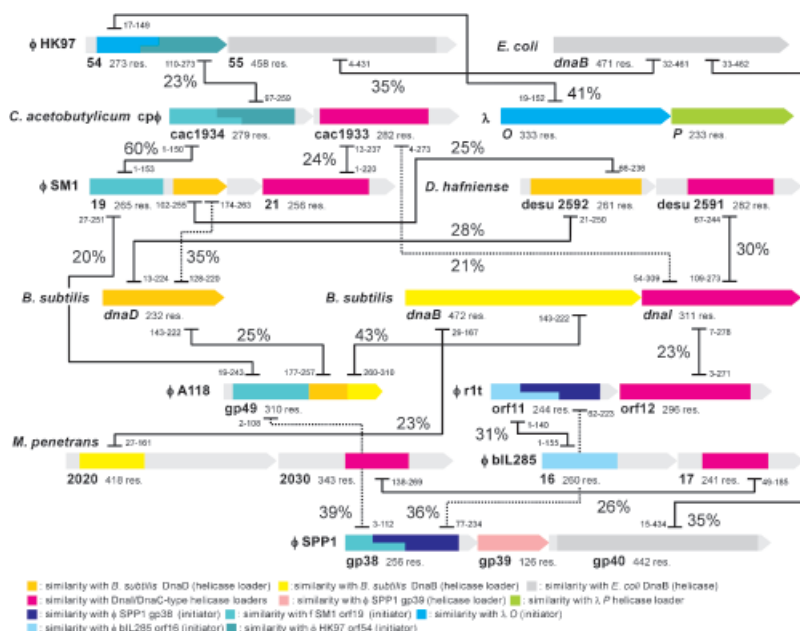


Fig. 28. Similarity network for the *Bacillus subtilis* *dnaBI* and *dnaD* genes. The bacterial and (pro)phage genes are shown roughly to size with the source indicated at the left side; the direction of transcription of the genes is indicated by arrowheads; gene names (locus tags) are shown underneath together with the sizes of the encoded proteins. The '%' values shown indicate the percentage of identical residues for the compared regions of two proteins.

S. pneumoniae R6, *Streptococcus mutans*, *S. agalactiae*) the *dnaD* genes are not linked to the *nth* genes; (3) in the third group (*L. plantarum*, *L. gasseri*, *Leuconostoc mesenteroides*, *Oenococcus oeni*) the *dnaD* genes are present but *nth* orthologues missing. However, these three groups do not correspond to the intrabranched clustering of the lactobacillales species based on the conservation of their 16S rRNA (Ludwig *et al.*, 1985).

We could not detect homologues of the *dnaBI* gene pair in the sequenced genomes of species of the clostridia branch of the firmicutes. Genes with significant similarity to *dnaD* were rare, and in all cases coupled to a gene related to phage helicase loaders (DnaI_{Bsu}-type). These gene pairs may rather be prophage replication modules (COM section C3.1.2.). In the sequenced genomes of species of the mollicutes branch of the firmicutes we could not detect *dnaD* homologues. All sequenced genomes contain a homologue of the *B. subtilis* *dnaB* gene, but only in *Mycoplasma penetrans* and *Ureaplasma urealyticum* linked to a *dnaI* homologue. Interestingly, the similarity to DnaB_{Bsu} is confined to the N-termini of these proteins, in contrast to the C-terminal similarity of DnaB_{Bsu} to phage initiators (see Fig. 28). We were unable to locate a *dnaI* homologue in the highly rearranged and size-reduced genomes of *Mycoplasma pneumoniae*, *M. gallisepticum* and *M. genitalium*. The *M. penetrans* gene pair MYPE2020 and MYPE2030, encoding proteins similar to the *B. subtilis* DnaB and DnaI proteins, respectively, is located closely upstream of the MYPE2050 gene, encoding the replicative helicase. This arrangement is reminiscent of the ILH-type replication module found in ϕ SPP1 and few other phages. However, this arrangement is not found in *U. urealyticum*, and may therefore represent a fortuitous gene-coupling event.

We infer from these observations that the common ancestor of the bacilli branch of the firmicutes acquired the *dnaBI* gene pair and the *dnaD* gene from different prophages. How the three phage-derived genes were subsequently reshaped to create the intricate three-protein helicase loading mechanism of *B. subtilis* remains 'in the cloud of unknowing' at present (Donachie, 2001). Our hypothesis that chromosomal replicons of bacteria acquired genes encoding initiator and helicase loader proteins from bacteriophages twice during the evolution of this domain of life lacks an experimental basis, and the evidence obtained by mere comparisons of protein sequences and gene contexts is circumstantial, although consistent. Admittedly, the most speculative part of our hypothesis is the presumed conversion of phage initiator proteins into host primosomal proteins.

It has frequently been claimed that there is an apparent lack of orthologous replication proteins in the prokaryota domain when compared with archaea and eukarya, in marked contrast to the considerable orthology found for

proteins of the other two major information processing systems – translation and transcription – in all three domains (Edgell & Doolittle, 1997; Leipe *et al.*, 1999; Tye, 2000; Woese, 2000, 2002). Forterre's (1999) suggested solution to this puzzle was that in the prokaryota many replication genes were replaced by nonorthologous plasmid- or virus (phage)-derived genes shortly after the divergence of the three domains of life from LUCA, the last universal common ancestor (Penny & Poole, 1999). This hypothesis was questioned by Moreira (2000), who found that several plasmid- or phage-encoded replication proteins resemble the chromosomal counterparts of their specific hosts more than other plasmid- or phage-encoded orthologues, even when taking higher mutational rates of virus genes into account (Drake *et al.*, 1998). *Prima vista*, our hypothesis seems to be in line with the proposal of Forterre, who explicitly mentions DnaC_{Eco} and DnaI_{Bsu}. There are three points of reserve, however: (1) the helicase loader genes were introduced into the chromosomal replicons of enterobacteriales and firmicutes independently, and in both cases apparently at a later stage of branch differentiation; (2) the helicase loader genes were introduced together with initiator genes into the chromosomal replicons – the latter, however, did not replace the cognate initiator of the recipient replicons, DnaA, but were converted into primosomal proteins instead; (3) it is presently unclear whether the newly introduced helicase loaders replaced the genes driving the primordial mechanism for helicase loading to ssDNA, or simply added more specificity. All three types of prokaryotic replicons provide us with examples suggesting that it is likely that helicase loaders added specificity to the existing molecular mechanisms rather than replacing the gene(s) responsible for primordial pathways for helicase loading.

Among the various types of phage replication modules discussed in the previous section, the initiator plus helicase modules are particularly informative. They are found in Gram(+) specific phages (ϕ 3626, ϕ 11) and Gram(–) specific phages (ϕ D3, ϕ ST64T, ϕ P22). It has not been shown experimentally, but it is reasonable to assume that the replication of these phage replicons depends on the specific interaction of their initiators with their cognate helicases.

Helinski, Konieczny and coworkers elucidated the intricate mechanism evolved by the broad host range plasmid RK2 (IncP-group) to ensure its propagation in different hosts (reviewed in Konieczny, 2003). The RK2 initiator gene *trfA* allows for the synthesis of a longer TrfA-44 protein, and, using an internal secondary start site, a shorter TrfA-33 protein. For the *in vitro* formation of a prepriming complex with *E. coli* replication proteins, both TrfA proteins are fully active. DnaA, DnaB and DnaC are strictly required, and DnaA was shown to recruit the helicase from the DnaB₆C₆ double-hexamer to the replication origin *oriV* of RK2 (Jiang

et al., 2003). Using *P. putida* DnaB (60% identical residues with DnaB_{Eco}) in the assay, TrfA-44 is more active than TrfA-33 in prepriming complex formation. The cognate DnaA is required when using TrfA-33; DnaA_{Ppu} and DnaB_{Ppu} were shown to interact (Caspi *et al.*, 2001). With *P. aeruginosa* DnaB (85% identical residues with DnaB_{Ppu}; 61% identical residues with DnaB_{Eco}), however, only the longer TrfA-44 is active and recruits the DnaB_{Pae} hexamer. Notably, an accessory helicase loader protein is not required for the recruitment of both *Pseudomonas* helicases to RK2 *oriV*. In addition, helicase loading during *in vitro* initiation of chromosome replication in *Pseudomonas* sp. does not require an accessory protein(s) (Y. Jiang, D. Helinski and A. Toukdarian, pers. commun.). Scherzinger *et al.* (1991, 1997) showed that the *in vitro* replication initiation of another broad host range plasmid, RSF1010 (IncQ-group), depends on three plasmid-encoded genes: *repC* (initiator), *repA* (helicase) and *repB'* (primase). Strand opening at *oriV* of RSF1010 requires the RepC protein, and the hexameric RepA helicase is recruited to *oriV* without an accessory helicase loader. The initiation of RSF1010 replication is independent from the host replicative helicase and the replication linked to the host machinery at a later stage of replisome formation.

Both sequenced *Yersinia pestis* genomes lack a *dnaTC* gene pair (see above), and BLAST searches failed to produce matches with other known helicase loaders (COM section C3.2.). Except for DnaT and DnaC, the entire set of replication and recombination proteins of *Y. pestis* is highly similar to that in *E. coli* ($\geq 70\%$ identical residues) with the exception of the PriC primosomal protein (36% identical residues). Because also the replication origins of the *E. coli* and *Y. pestis* chromosomes are virtually identical, it is likely that replication initiation and restart follow the same routes. Whether the missing DnaC – and also the missing DnaT, we would add – is responsible for the long generation times of *Y. pestis* as suggested by Thomson *et al.* (2002) has not been thoroughly examined. *Y. pestis* is an evolutionary young species ($< 2 \times 10^4$ years), and conclusions drawn by comparing sequence data should therefore be regarded with caution (Achtman *et al.*, 1999). However, the genomes of *Y. pseudotuberculosis* IP 32953 (ancestor of *Y. pestis*), and of *Yersinia enterocolitica* (almost completely sequenced) lack a *dnaTC* gene pair. At present, we may assume that helicase loading during replication initiation and restart does not require accessory protein(s) in *Yersinia*.

In conclusion, we propose to exclude proteins other than those known to be directly involved in the essential steps of replication from interdomain comparisons aimed to elucidate the evolutionary origin of this basic information processing system of all extant cells. Helicase loaders do not belong to this group of essential proteins of prokaryotic replicons because initiation reactions have been described (others are plausible) which dispense with them. An exam-

ple of an approach concentrating on one essential replication protein instead of using poorly defined protein families was recently presented by Giraldo, who was able to trace the relationship of prokaryotic, archaeal and eukaryal replication initiator proteins very close to LUCA (Giraldo, 2003).

Perspectives

The ongoing race/rage of genomic sequencing will provide bacteriophage research with many promising novel objects, but only their exploration by genetic, biochemical and structural studies can transform the present data overflow into something coming closer to knowledge (Brenner, 2000). In particular, knowing more about phage and plasmid replication will contribute to a better understanding of the spread of virulence genes among human pathogens, a task of obvious importance (for recent publications see, e.g., Davis & Waldor, 2003; Ferretti *et al.*, 2004; Munson *et al.*, 2004; Nair *et al.*, 2004; Nelson *et al.*, 2004; Summer *et al.*, 2004).

The long-standing observation of mosaicism among phage genomes also extends to their replication/recombination genes. Our study of bacteriophage replication modules suggests that probably every theoretically possible combination exists with respect to replication protein assortment, and also with respect to the various types of proteins that carry out a particular enzymatic function. This is to some degree in conflict with one of the fundamental dogmas of evolution theory: that selective pressure drives evolution in favour of the best-adapted genome (replicon) and results in the 'survival of the fittest', as Darwin's contemporary Spencer phrased it. However, if we tentatively understand 'the fittest' as plural instead of singular we could start to define a 'fitness threshold', and study its variation over time depending on the ever-changing environmental conditions. A thorough discussion of this topic is far beyond the scope of this review. But we speculate that the recombination functions encoded by many phages are, in addition to their role for phage replication, responsible for the creation and the maintenance of mosaicism: they could promote the repeated re-creation of different replication modules with a comparable selective fitness. Temperate phage replicons could be of particular importance for any experimental approach to define a 'fitness threshold' because they represent a unique class of replicons: a decrease in their fitness is not necessarily accompanied by their extinction as long as 'backup' prophage copies reside in host genomes. In general, evolutionary considerations by microbiologists are not readily accepted by evolutionary biologists, including the authoritative Mayr (1998). However, we wish to remind the reader that the 'phage fluctuation test' presented by Luria & Delbrück (1943) is still the most convincing experimental proof for yet another fundamental dogma of evolution theory: the

random occurrence of mutations. After all, phage biology may still be a promising route to study *experimentally* the principles of the evolution of life.

The enzymology of chromosome replication in *E. coli* was established in the 1980s by the Kornberg lab among others, and readily accepted as the comprehensive and valid grammar, syntax and vocabulary – literally spoken – of one of the central cellular pathways of nucleic acid transactions. Genetic and biochemical studies of the propagation of the two other types of prokaryotic replicons, plasmids and phages, provided evidence for a common grammar: initiation, priming, elongation and termination of DNA synthesis. However, various syntactic variants, i.e. molecular mechanisms, were revealed by studying initiation, priming, elongation and termination of DNA replication in individual plasmids and phages. Research on plasmid replication contributed broad knowledge to the different mechanisms for initiation and, in particular, copy number control and segregation (del Solar *et al.*, 1998; Giraldo, 2003). Studies of phage replication revealed further mechanisms for initiation, priming and replication fork restart involving recombination processes. The analysis of phage-encoded replication proteins teaches us that it is the vocabulary that varies most: a plethora of different types of initiators, helicase loaders, helicases, primases, DNA polymerases and polymerase accessory proteins fulfil their enzymatic roles within a strikingly similar grammatical and fairly similar syntactic framework. Therefore, the *E. coli* way of chromosome replication can be considered a local dialect rather than the *lingua franca*.

If we abandon thinking about *E. coli* as ‘the model prokaryote’ we can approach a more abstract model for replication by defining enzymatic steps, rather than demanding that particular types of enzymes carry out these steps (Benkovic *et al.*, 2001). This ‘universal textbook of replication’ would have to include the still missing descriptions of: (1) the DNA-unwinding by initiators in thermodynamically solid terms, and (2) the involvement of cell membrane (components) in replication. Therefore, we are not yet in a position to close this chapter of molecular biology.

Acknowledgements

We thank Walter Messer for generous support over all these years – and for sharing our curiosity. In addition, we thank Ramón Díaz Orejas for continuous encouragement and inspiring discussions. We gratefully acknowledge helpful comments and communication of unpublished work by Erich Lanka, Juan C. Alonso, Sandra Weller, Don Helinski, John Iandolo, Grzes Wegrzyn, Kenneth Mariani, Elisabeth Haggård and Jasna Rakonjac. Special thanks are due to J. C. Alonso, P. Tavares, R. Lurz and T. Trautner for making their

review on phage SPP1 available to us prior to publication. This work was supported by the European Union (MolTool, grant LSHG-CT-2004-503155) and the Max-Planck-Society.

Supplementary material

For all figures, high resolution versions are available from the authors on request.

The following supplementary material is available for this article online.

Appendix S1. A compendium of phage replication origins and phage replication/recombination proteins (including 37 Tables and 22 Figures).

The material is available as part of the online article from <http://www.blackwell-synergy.com>.

References

- Abril AM, Salas M, Andreu JM, Hermoso JM & Rivas G (1997) Phage phi29 protein p6 is in a monomer-dimer equilibrium that shifts to higher association states at the millimolar concentrations found in vivo. *Biochemistry* **36**: 11901–11908.
- Abril AM, Marco S, Carrascosa JL, Salas M & Hermoso JM (1999) Oligomeric structures of the phage phi29 histone-like protein p6. *J Mol Biol* **292**: 581–588.
- Achtman M, Zurth K, Morelli G, Torrea G, Guiyoule A & Carniel E (1999) *Yersinia pestis*, the cause of plague, is a recently emerged clone of *Yersinia pseudotuberculosis*. *Proc Natl Acad Sci USA* **96**: 14043–14048.
- Ahmad S, Weisburg WG & Jensen RA (1990) Evolution of aromatic amino acid biosynthesis and application to the fine-tuned phylogenetic positioning of enteric bacteria. *J Bacteriol* **172**: 1051–1061.
- Allison DP, Ganesan AT, Olson AC, Snyder CM & Mitra S (1977) Electron microscopic studies of bacteriophage M13 DNA replication. *J Virol* **24**: 673–684.
- Alonso JC, Tavares P, Lurz R & Trautner TA (2005) Bacteriophage SPP1. *The Bacteriophages*. 2nd edn (Calendar R, ed), pp. 331–349. Oxford University Press, Oxford, UK.
- Altmann E, Klein JR & Henrich B (1999) Primary structure and features of the genome of the *Lactobacillus gasseri* temperate bacteriophage (phi)adh. *Gene* **236**: 333–346.
- Aoyama A & Hayashi M (1986) Synthesis of bacteriophage phi X174 in vitro: mechanism of switch from DNA replication to DNA packaging. *Cell* **47**: 99–106.
- Ayora S, Langer U & Alonso JC (1998) *Bacillus subtilis* DnaG primase stabilises the bacteriophage SPP1 G40P helicase-ssDNA complex. *FEBS Lett* **439**: 59–62.
- Ayora S, Missich R, Mesa P, Lurz R, Yang S, Egelman EH & Alonso JC (2002) Homologous-pairing activity of the *Bacillus subtilis* bacteriophage SPP1 replication protein G35P. *J Biol Chem* **277**: 35969–35979.
- Banks DJ, Beres SB & Musser JM (2002) The fundamental contribution of phages to GAS evolution, genome

- diversification and strain emergence. *Trends Microbiol* **10**: 515–521.
- Bao K & Cohen SN (2003) Recruitment of terminal protein to the ends of *Streptomyces* linear plasmids and chromosomes by a novel telomere-binding protein essential for linear DNA replication. *Genes Dev* **17**: 774–785.
- Baranska S, Gabig M, Wegrzyn A, *et al.* (2001) Regulation of the switch from early to late bacteriophage lambda DNA replication. *Microbiology* **147**: 535–547.
- Baranska S, Konopa G & Wegrzyn G (2002) Directionality of lambda plasmid DNA replication carried out by the heritable replication complex. *Nucleic Acids Res* **30**: 1176–1181.
- Barry J & Alberts B (1994a) A role for two DNA helicases in the replication of T4 bacteriophage DNA. *J Biol Chem* **269**: 33063–33068.
- Barry J & Alberts B (1994b) Purification and characterization of bacteriophage T4 gene 59 protein. A DNA helicase assembly protein involved in DNA replication. *J Biol Chem* **269**: 33049–33062.
- Benkovic SJ, Valentine AM & Salinas F (2001) Replisome-mediated DNA replication. *Annu Rev Biochem* **70**: 181–208.
- Bird LE, Subramanya HS & Wigley DB (1998) Helicases: a unifying structural theme? *Curr Opin Struct Biol* **8**: 14–18.
- Biswas EE & Biswas SB (1999) Mechanism of DnaB helicase of *Escherichia coli*: structural domains involved in ATP hydrolysis, DNA binding, and oligomerization. *Biochemistry* **38**: 10919–10928.
- Blanco L, Bernad A, Esteban JA & Salas M (1992) DNA-independent deoxynucleotidylatation of the phi 29 terminal protein by the phi 29 DNA polymerase. *J Biol Chem* **267**: 1225–1230.
- Bleuit JS, Ma Y, Munro J & Morrical SW (2004) Mutations in a conserved motif inhibit single-stranded DNA binding and recombination mediator activities of bacteriophage T4 UvsY protein. *J Biol Chem* **279**: 6077–6086.
- Bollback JP & Huelsenbeck JP (2001) Phylogeny, genome evolution, and host specificity of single-stranded RNA bacteriophage (family Leviviridae). *J Mol Evol* **52**: 117–128.
- Botstein D & Matz MJ (1970) A recombination function essential to the growth of bacteriophage P22. *J Mol Biol* **54**: 417–440.
- Bouché JP, Rowen L & Kornberg A (1978) The RNA primer synthesized by primase to initiate phage G4 DNA replication. *J Biol Chem* **253**: 765–769.
- Bourguignon GJ, Sweeney TK & Delius H (1976) Multiple origins and circular structures in replicating T5 bacteriophage DNA. *J Virol* **18**: 245–259.
- Bourque LW & Christensen JR (1980) The synthesis of coliphage T1 DNA: requirement for host dna genes involved in elongation. *Virology* **102**: 310–316.
- Bravo A, Serrano-Heras G & Salas M (2005) Compartmentalization of prokaryotic DNA replication. *FEMS Microbiol Rev* **29**: 25–47.
- Brenner S (2000) Biochemistry strikes back. *Trends Biochem Sci* **25**: 584.
- Briani F, Dehò G, Forti F & Ghisotti D (2001) The plasmid status of satellite bacteriophage P4. *Plasmid* **45**: 1–17.
- Broker TR (1973) An electron microscopic analysis of pathways for bacteriophage T4 DNA recombination. *J Mol Biol* **81**: 1–16.
- Broker TR & Doermann AH (1975) Molecular and genetic recombination of bacteriophage T4. *Annu Rev Genet* **9**: 213–244.
- Bruessow H & Desiere F (2001) Comparative phage genomics and the evolution of Siphoviridae: insights from dairy phages. *Mol Microbiol* **39**: 213–222.
- Bruessow H, Canchaya C & Hardt WD (2004) Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. *Microbiol Mol Biol Rev* **68**: 560–602.
- Burger KJ & Trautner TA (1978) Specific labelling of replicating SPP1 DNA: analysis of viral DNA synthesis and identification of phage DNA-genes. *Mol Gen Genet* **166**: 277–285.
- Burke RL, Munn M, Barry J & Alberts BM (1985) Purification and properties of the bacteriophage T4 gene 61 RNA priming protein. *J Biol Chem* **260**: 1711–1722.
- Cadman CJ & McGlynn P (2004) PriA helicase and SSB interact physically and functionally. *Nucleic Acids Res* **32**: 6378–6387.
- Cairns J (1963) The bacterial chromosome and its manner of replication as seen by autoradiography. *J Mol Biol* **6**: 208–213.
- Campbell A (1994) Comparative molecular biology of lambdoid phages. *Annu Rev Microbiol* **48**: 193–222.
- Campbell A & Botstein D (1983) Evolution of the lambdoid phages. *Lambda II* (Hendrix RW, Roberts JW, Stahl FW & Weisberg RA, eds), pp. 365–380. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Canchaya C, Proux C, Fournous G, Bruttin A & Bruessow H (2003) Prophage genomics. *Microbiol Mol Biol Rev* **67**: 238–276.
- Carles-Kinch K & Kreuzer KN (1997) RNA–DNA hybrid formation at a bacteriophage T4 replication origin. *J Mol Biol* **266**: 915–926.
- Casjens S (2003) Prophages and bacterial genomics: what have we learned so far? *Mol Microbiol* **49**: 277–300.
- Caspi R, Pacek M, Consiglieri G, Helinski DR, Toukdarian A & Konieczny I (2001) A broad host range replicon with different requirements for replication initiation in three bacterial species. *EMBO J* **20**: 3262–3271.
- Chai S, Bravo A, Lüder G, Nedlin A, Trautner TA & Alonso JC (1992) Molecular analysis of the *Bacillus subtilis* bacteriophage SPP1 region encompassing genes 1 to 6. The products of gene 1 and gene 2 are required for pac cleavage. *J Mol Biol* **224**: 87–102.
- Chakraborty T, Yoshinaga K, Lothar H & Messer W (1982) Purification of the *E. coli* dnaA gene product. *EMBO J* **1**: 1545–1549.
- Chang P & Marians KJ (2000) Identification of a region of *Escherichia coli* DnaB required for functional interaction with DnaG at the replication fork. *J Biol Chem* **275**: 26187–26195.

- Chang PC, Kim ES & Cohen SN (1996) *Streptomyces* linear plasmids that contain a phage-like, centrally located, replication origin. *Mol Microbiol* **22**: 789–800.
- Chao KL & Lohman TM (1991) DNA-induced dimerization of the *Escherichia coli* Rep helicase. *J Mol Biol* **221**: 1165–1181.
- Chetverin AB (2004) Replicable and recombinogenic RNAs. *FEBS Lett* **567**: 35–41.
- Chopin A, Bolotin A, Sorokin A, Ehrlich SD & 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.
- Chung YB & Hinkle DC (1990) Bacteriophage T7 DNA packaging. I. Plasmids containing a T7 replication origin and the T7 concatemer junction are packaged into transducing particles during phage infection. *J Mol Biol* **216**: 911–926.
- Collins J & Hohn B (1978) Cosmids: a type of plasmid gene-cloning vector that is packageable in vitro in bacteriophage lambda heads. *Proc Natl Acad Sci USA* **75**: 4242–4246.
- Corpet F (1988) Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res* **16**: 10881–10890.
- Cox MM (2001) Historical overview: searching for replication help in all of the rec places. *Proc Natl Acad Sci USA* **98**: 8173–8180.
- Cox MM, Goodman MF, Kreuzer KN, Sherratt DJ, Sandler SJ & Marians KJ (2000) The importance of repairing stalled replication forks. *Nature* **404**: 37–41.
- Cuff JA, Clamp ME, Siddiqui AS, Finlay M & Barton GJ (1998) JPred: a consensus secondary structure prediction server. *Bioinformatics* **14**: 892–893.
- Dannenbergh R & Mosig G (1983) Early intermediates in bacteriophage T4 DNA replication and recombination. *J Virol* **45**: 813–831.
- Datta HJ, Khatri GS & Bastia D (1999) Mechanism of recruitment of DnaB helicase to the replication origin of the plasmid pSC101. *Proc Natl Acad Sci USA* **96**: 73–78.
- Davis BM & Waldor MK (2003) Filamentous phages linked to virulence of *Vibrio cholerae*. *Curr Opin Microbiol* **6**: 35–42.
- David M, Lubinsky-Mink S, Ben Zvi A, Ulitzur S, Kuhn J & Suissa M (1992) A stable *Escherichia coli*–*Mycobacterium smegmatis* plasmid shuttle vector containing the mycobacteriophage D29 origin. *Plasmid* **28**: 267–271.
- Deneke J, Ziegelin G, Lurz R & Lanka E (2000) The protelomerase of temperate *Escherichia coli* phage N15 has cleaving-joining activity. *Proc Natl Acad Sci USA* **97**: 7721–7726.
- Deneke J, Ziegelin G, Lurz R & Lanka E (2002) Phage N15 telomere resolution. Target requirements for recognition and processing by the protelomerase. *J Biol Chem* **277**: 10410–10419.
- Deneke J, Burgin AB, Wilson SL & Chaconas G (2004) Catalytic residues of the telomere resolvase ResT: a pattern similar to, but distinct from, tyrosine recombinases and type IB topoisomerases. *J Biol Chem* **279**: 53699–53706.
- Díaz R & Pritchard RH (1978) Cloning of replication origins from the *E. coli* K12 chromosome. *Nature* **275**: 561–564.
- Dodson M, Echols H, Wickner S, *et al.* (1986) Specialized nucleoprotein structures at the origin of replication of bacteriophage lambda: localized unwinding of duplex DNA by a six-protein reaction. *Proc Natl Acad Sci USA* **83**: 7638–7642.
- Donachie WD (2001) Co-ordinate regulation of the *Escherichia coli* cell cycle or The cloud of unknowing. *Mol Microbiol* **40**: 779–785.
- Dotto GP, Horiuchi K & Zinder ND (1984) The functional origin of bacteriophage f1 DNA replication. Its signals and domains. *J Mol Biol* **172**: 507–521.
- Drake JW, Charlesworth B, Charlesworth D & Crow JF (1998) Rates of spontaneous mutation. *Genetics* **148**: 1667–1686.
- Dressler D, Wolfson J & Magazin M (1972) Initiation and reinitiation of DNA synthesis during replication of bacteriophage T7. *Proc Natl Acad Sci USA* **69**: 998–1002.
- Drolet M, Bi X & Liu LF (1994) Hypernegative supercoiling of the DNA template during transcription elongation in vitro. *J Biol Chem* **269**: 2068–2074.
- Duchaud E, Rusniok C, Frangeul L, *et al.* (2003) The genome sequence of the entomopathogenic bacterium *Photobacterium luminescens*. *Nat Biotechnol* **21**: 1307–1313.
- Dudas KC & Kreuzer KN (2001) UvsW protein regulates bacteriophage T4 origin-dependent replication by unwinding R-loops. *Mol Cell Biol* **21**: 2706–2715.
- Edgell DR & Doolittle WF (1997) Archaea and the origin(s) of DNA replication proteins. *Cell* **89**: 995–998.
- Espeli O & Marians KJ (2004) Untangling intracellular DNA topology. *Mol Microbiol* **52**: 925–931.
- Esteban JA, Blanco L, Villar L & Salas M (1997) In vitro evolution of terminal protein-containing genomes. *Proc Natl Acad Sci USA* **94**: 2921–2926.
- Feiss M, Sippy J & Miller G (1985) Processive action of terminase during sequential packaging of bacteriophage lambda chromosomes. *J Mol Biol* **186**: 759–771.
- Feng JN, Russel M & Model P (1997) A permeabilized cell system that assembles filamentous bacteriophage. *Proc Natl Acad Sci USA* **94**: 4068–4073.
- Ferretti JJ, Ajdic D & McShan WM (2004) Comparative genomics of streptococcal species. *Indian J Med Res* **119**(Suppl): 1–6.
- Foley S, Lucchini S, Zwahlen MC & Bruessow H (1998) A short noncoding viral DNA element showing characteristics of a replication origin confers bacteriophage resistance to *Streptococcus thermophilus*. *Virology* **250**: 377–387.
- Formosa T & Alberts BM (1986) DNA synthesis dependent on genetic recombination: characterization of a reaction catalyzed by purified bacteriophage T4 proteins. *Cell* **47**: 793–806.
- Forterre P (1999) Displacement of cellular proteins by functional analogues from plasmids or viruses could explain puzzling phylogenies of many DNA informational proteins. *Mol Microbiol* **33**: 457–465.
- Freire R, Serrano M, Salas M & Hermoso JM (1996) Activation of replication origins in phi29-related phages requires the recognition of initiation proteins to specific nucleoprotein complexes. *J Biol Chem* **271**: 31000–31007.

- Friedman D, Tomich P, Parsons C, Olson E, Deans R & Flamm E (1981) lambda altSF: a phage variant that acquired the ability to substitute specific sets of genes at high frequency. *Proc Natl Acad Sci USA* **78**: 410–414.
- Fujikawa N, Kurumizaka H, Nureki O, Terada T, Shirouzu M, Katayama T & Yokoyama S (2003) Structural basis of replication origin recognition by the DnaA protein. *Nucleic Acids Res* **31**: 2077–2086.
- Fujimura RK & Roop BC (1983) Interaction of a DNA-binding protein, the product of gene D5 of bacteriophage T5, with double-stranded DNA: effects on T5 DNA polymerase functions in vitro. *J Virol* **46**: 778–787.
- Fuller CW & Richardson CC (1985a) Initiation of DNA replication at the primary origin of bacteriophage T7 by purified proteins. Initiation of bidirectional synthesis. *J Biol Chem* **260**: 3197–3206.
- Fuller CW & Richardson CC (1985b) Initiation of DNA replication at the primary origin of bacteriophage T7 by purified proteins. Site and direction of initial DNA synthesis. *J Biol Chem* **260**: 3185–3196.
- Fuller RS, Kaguni JM & Kornberg A (1981) Enzymatic replication of the origin of the *E. coli* chromosome. *Proc Natl Acad Sci USA* **78**: 7370–7374.
- Fuller RS, Funnell BE & Kornberg A (1984) The *dnaA* protein complex with the *E. coli* chromosomal origin (*oriC*) and other sites. *Cell* **38**: 889–900.
- George JW, Stohr BA, Tomso DJ & Kreuzer KN (2001) The tight linkage between DNA replication and double-strand break repair in bacteriophage T4. *Proc Natl Acad Sci USA* **98**: 8290–8297.
- Gilbert W & Dressler D (1968) DNA replication: the rolling circle model. *Cold Spring Harbor Symp Quant Biol* **33**: 473–484.
- Giraldo R (2003) Common domains in the initiators of DNA replication in Bacteria, Archaea and Eukarya: combined structural, functional and phylogenetic perspectives. *FEMS Microbiol Rev* **26**: 533–554.
- Giraldo R & Fernandez-Tresguerres ME (2004) Twenty years of the pPS10 replicon: insights on the molecular mechanism for the activation of DNA replication in iteron-containing bacterial plasmids. *Plasmid* **52**: 69–83.
- Glinkowska M, Majka J, Messer W & Wegrzyn G (2003) The mechanism of regulation of bacteriophage lambda *p_R* promoter activity by *Escherichia coli* DnaA protein. *J Biol Chem* **278**: 22250–22256.
- Grompone G, Ehrlich SD & Michel B (2003) Replication restart in *gyrB* *Escherichia coli* mutants. *Mol Microbiol* **48**: 845–854.
- Hall MC & Matson SW (1999) Helicase motifs: the engine that powers DNA unwinding. *Mol Microbiol* **34**: 867–877.
- Hase T, Nakai M & Masamune Y (1989) Transcription of a region downstream from lambda *ori* is required for replication of plasmids derived from coliphage lambda. *Mol Gen Genet* **216**: 120–125.
- Hayes F (1998) A family of stability determinants in pathogenic bacteria. *J Bacteriol* **180**: 6415–6418.
- Heller RC & Mariani KJ (2005) The disposition of nascent strands at stalled replication forks dictates the pathway of replisome loading during restart. *Mol Cell* **17**: 733–743.
- Hendrix RW, Lawrence JG, Hatfull GF & Casjens S (2000) The origins and ongoing evolution of viruses. *Trends Microbiol* **8**: 504–508.
- Herman-Antosiewicz A, Obuchowski M & Wegrzyn G (2001) A plasmid cloning vector with precisely regulatable copy number in *Escherichia coli*. *Mol Biotechnol* **17**: 193–199.
- Hermoso JM, Mendez E, Soriano F & Salas M (1985) Location of the serine residue involved in the linkage between the terminal protein and the DNA of phage phi 29. *Nucleic Acids Res* **13**: 7715–7728.
- Hiasa H & Mariani KJ (1999) Initiation of bidirectional replication at the chromosomal origin is directed by the interaction between helicase and primase. *J Biol Chem* **274**: 27244–27248.
- Hiasa H, Sakai H, Tanaka K, Honda Y, Komano T & Godson GN (1989) Mutational analysis of the primer RNA template region in the replication origin (*oric*) of bacteriophage G4: priming signal recognition by *Escherichia coli* primase. *Gene* **84**: 9–16.
- Higashitani N, Higashitani A & Horiuchi K (1993) Nucleotide sequence of the primer RNA for DNA replication of filamentous bacteriophages. *J Virol* **67**: 2175–2181.
- Higashitani A, Greenstein D, Hirokawa H, Asano S & Horiuchi K (1994) Multiple DNA conformational changes induced by an initiator protein precede the nicking reaction in a rolling circle replication origin. *J Mol Biol* **237**: 388–400.
- Highton PJ, Chang Y & Myers RJ (1990) Evidence for the exchange of segments between genomes during the evolution of lambdoid bacteriophages. *Mol Microbiol* **4**: 1329–1340.
- Horiuchi K (1997) Initiation mechanisms in replication of filamentous phage DNA. *Genes Cells* **2**: 425–432.
- Hours C & Denhardt DT (1979) 'Nick translation' in *Escherichia coli* rep strains deficient in DNA polymerase I activities. *Mol Gen Genet* **172**: 73–80.
- Huang WM, Robertson M, Aron J & Casjens S (2004) Telomere exchange between linear replicons of *Borrelia burgdorferi*. *J Bacteriol* **186**: 4134–4141.
- Iandolo JJ, Worrell V, Groicher KH, *et al.* (2002) Comparative analysis of the genomes of the temperate bacteriophages phi 11, phi 12 and phi 13 of *Staphylococcus aureus* 8325. *Gene* **289**: 109–118.
- Ilyina TV, Gorbalenya AE & Koonin EV (1992) Organization and evolution of bacterial and bacteriophage primase–helicase systems. *J Mol Evol* **34**: 351–357.
- Ishmael FT, Alley SC & Benkovic SJ (2001) Identification and mapping of protein–protein interactions between gp32 and gp59 by cross-linking. *J Biol Chem* **276**: 25236–25242.
- Jacob F, Brenner S & Cuzin F (1963) On the regulation of DNA replication in bacteria. *Cold Spring Harbor Symp Quant Biol* **28**: 329–348.
- James JA, Escalante CR, Yoon-Roberts M, Edwards TA, Linden RM & Aggarwal AK (2003) Crystal structure of the SF3

- helicase from adeno-associated virus type 2. *Structure* **11**: 1025–1035.
- Jiang Y, Pacek M, Helinski DR, Konieczny I & Toukdarian A (2003) A multifunctional plasmid-encoded replication initiation protein both recruits and positions an active helicase at the replication origin. *Proc Natl Acad Sci USA* **100**: 8692–8697.
- Jones JM & Nakai H (1999) Duplex opening by primosome protein PriA for replisome assembly on a recombination intermediate. *J Mol Biol* **289**: 503–516.
- Jones CE, Mueser TC & Nossal NG (2000) Interaction of the bacteriophage T4 gene 59 helicase loading protein and gene 41 helicase with each other and with fork, flap, and cruciform DNA. *J Biol Chem* **275**: 27145–27154.
- Kadyrov FA & Drake JW (2001) Conditional coupling of leading-strand and lagging-strand DNA synthesis at bacteriophage T4 replication forks. *J Biol Chem* **276**: 29559–29566.
- Kaguni JM & Kornberg A (1982) The rho subunit of RNA polymerase holoenzyme confers specificity in priming M13 viral DNA replication. *J Biol Chem* **257**: 5437–5443.
- Kaguni JM & Kornberg A (1984) Replication initiated at the origin (*oriC*) of the *E. coli* chromosome reconstituted with purified enzymes. *Cell* **38**: 183–190.
- Kaguni JM, Fuller RS & Kornberg A (1982) Enzymatic replication of *E. coli* chromosomal origin is bidirectional. *Nature* **296**: 623–627.
- Kaiser K & Murray NE (1979) Physical characterisation of the 'Rac prophage' in *E. coli* K12. *Mol Gen Genet* **175**: 159–174.
- Kaplan DL & O'Donnell M (2002) DnaB drives DNA branch migration and dislodges proteins while encircling two DNA strands. *Mol Cell* **10**: 647–657.
- Katayama T (2001) Feedback controls restrain the initiation of *Escherichia coli* chromosomal replication. *Mol Microbiol* **41**: 9–17.
- Kawasaki Y, Matsunaga F, Kano Y, Yura T & Wada C (1996) The localized melting of mini-F by the combined action of the mini-F initiator protein (RepE) and HU and DnaA of *Escherichia coli*. *Mol Gen Genet* **253**: 42–49.
- Khan SA (1997) Rolling-circle replication of bacterial plasmids. *Microbiol Mol Biol Rev* **61**: 442–455.
- Khan SA (2000) Plasmid rolling-circle replication: recent developments. *Mol Microbiol* **37**: 477–484.
- Kodaira KI, Oki M, Kakikawa M, Watanabe N, Hirakawa M, Yamada K & Taketo A (1997) Genome structure of the *Lactobacillus* temperate phage phi g1e: the whole genome sequence and the putative promoter/repressor system. *Gene* **187**: 45–53.
- Kogoma T (1997) Stable DNA replication: interplay between DNA replication, homologous recombination, and transcription. *Microbiol Mol Biol Rev* **61**: 212–238.
- Kogoma T & Lark KG (1975) Characterization of the replication of *Escherichia coli* DNA in the absence of protein synthesis: stable DNA replication. *J Mol Biol* **94**: 243–256.
- Kogoma T, Hong X, Cadwell GW, Barbard KG & Asai T (1993) Requirement of homologous recombination functions for viability of the *Escherichia coli* cell that lacks RNase H and exonuclease V activities. *Biochimie* **75**: 89–99.
- Kohiyama M, Cousin D, Ryter A & Jacob F (1966) Mutants thermosensibles d'*E. coli* K-12. I. Isolement et caractérisation rapide. *Ann Inst Pasteur* **110**: 465–486.
- Konieczny I (2003) Strategies for helicase recruitment and loading in bacteria. *EMBO Rep* **4**: 37–41.
- Konieczny I, Doran KS, Helinski DR & Blasina A (1997) Role of TrfA and DnaA proteins in origin opening during initiation of DNA replication of the broad host range plasmid RK2. *J Biol Chem* **272**: 20173–20178.
- Kornberg A & Baker TA (1992) *DNA Replication*. W.H. Freeman and Company, New York, NY.
- Korolev S, Yao N, Lohman TM, Weber PC & Waksman G (1998) Comparisons between the structures of HCV and Rep helicases reveal structural similarities between SF1 and SF2 super-families of helicases. *Protein Sci* **7**: 605–610.
- Krause M, Lurz R, Rückert B & Messer W (1997) Complexes at the replication origin of *Bacillus subtilis* with homologous and heterologous DnaA protein. *J Mol Biol* **274**: 365–380.
- Kuzminov A (1999) Recombinational repair of DNA damage in *Escherichia coli* and bacteriophage lambda. *Microbiol Mol Biol Rev* **63**: 751–813.
- Lam AMI, Keeney D & Frick DN (2003) Two novel conserved motifs in the hepatitis C virus NS3 protein critical for helicase action. *J Biol Chem* **278**: 44514–44524.
- Lawrence JG & Ochman H (1998) Molecular archaeology of the *Escherichia coli* genome. *Proc Natl Acad Sci USA* **95**: 9413–9417.
- LeClerc JE & Richardson CC (1979) Gene 2 protein of bacteriophage T7: purification and requirement for packaging of T7 DNA in vitro. *Proc Natl Acad Sci USA* **76**: 4852–4856.
- Leipe DD, Aravind L & Koonin EV (1999) Did DNA replication evolve twice independently? *Nucleic Acids Res* **27**: 3389–3401.
- Lemonnier M, Ziegelin G, Reick T, Gomez AM, Diaz Orejas R & Lanka E (2003) Bacteriophage P1 Ban protein is a hexameric DNA helicase that interacts with and substitutes for *Escherichia coli* DnaB. *Nucleic Acids Res* **31**: 3918–3928.
- Linderth NA, Ziermann R, Haggård-Liungquist E, Christie GE & Calendar R (1991) Nucleotide sequence of the DNA packaging and capsid synthesis genes of bacteriophage P2. *Nucleic Acids Res* **19**: 7207–7214.
- Liu Y & Haggård-Liungquist E (1994) Studies of bacteriophage P2 DNA replication: localization of the cleavage site of the A protein. *Nucleic Acids Res* **22**: 5204–5210.
- Liu J & Mariani KJ (1999) PriA-directed assembly of a primosome on D loop DNA. *J Biol Chem* **274**: 25033–25041.
- Liu LF & Wang JC (1987) Supercoiling of the DNA template during transcription. *Proc Natl Acad Sci USA* **84**: 7024–7027.
- Liu Y, Saha S & Haggård-Liungquist E (1993) Studies of bacteriophage P2 DNA replication. The DNA sequence of the cis-acting gene A and ori region and construction of a P2 mini-chromosome. *J Mol Biol* **231**: 361–374.

- Liu J, Xu L, Sandler SJ & Marians KJ (1999) Replication fork assembly at recombination intermediates is required for bacterial growth. *Proc Natl Acad Sci USA* **96**: 3552–3555.
- Llsoa M, Gomis-Ruth FX, Coll M & de la Cruz Fd F (2002) Bacterial conjugation: a two-step mechanism for DNA transport. *Mol Microbiol* **45**: 1–8.
- Lobocka MB, Rose DJ, Plunkett G III, Rusin M, Samojedny A, Lehnher H, Yarmolinsky MB & Blattner FR (2004) Genome of bacteriophage P1. *J Bacteriol* **186**: 7032–7068.
- Lopez de Saro FJ & O'Donnell M (2001) Interaction of the beta sliding clamp with MutS, ligase, and DNA polymerase I. *Proc Natl Acad Sci USA* **98**: 8376–8380.
- Lu YB, Datta HJ & Bastia D (1998) Mechanistic studies of initiator–initiator interaction and replication initiation. *EMBO J* **17**: 5192–5200.
- Luder A & Mosig G (1982) Two alternative mechanisms for initiation of DNA replication forks in bacteriophage T4: priming by RNA polymerase and by recombination. *Proc Natl Acad Sci USA* **79**: 1101–1105.
- Ludwig W, Seewaldt E, Kilpper-Balz R, Schleifer KH, Magrum L, Woese CR, Fox GE & Stackebrandt E (1985) The phylogenetic position of *Streptococcus* and *Enterococcus*. *J Gen Microbiol* **131**(Part 3): 543–551.
- Luria SE & Delbrück M (1943) Mutations of bacteria from virus sensitivity to virus resistance. *Genetics* **28**: 491–511.
- Lusetti SL & Cox MM (2002) The bacterial RecA protein and the recombinational DNA repair of stalled replication forks. *Annu Rev Biochem* **71**: 71–100.
- Maaloe O & Kjeldgaard NO (1966) *Control of Macromolecular Synthesis*. Benjamin, New York.
- Mackiewicz P, Zakrzewska-Czerwinska J, Zawilak A, Dudek MR & Cebat S (2004) Where does bacterial replication start? Rules for predicting the oriC region. *Nucleic Acids Res* **32**: 3781–3791.
- Madsen SM, Mills D, Djordjevic G, Israelsen H & Klaenhammer TR (2001) Analysis of the genetic switch and replication region of a P335-type bacteriophage with an obligate lytic lifestyle on *Lactococcus lactis*. *Appl Environ Microbiol* **67**: 1128–1139.
- Maisnier-Patin S, Nordström K & Dasgupta S (2001) Replication arrests during a single round of replication of the *Escherichia coli* chromosome in the absence of DnaC activity. *Mol Microbiol* **42**: 1371–1382.
- Makeyev EV & Grimes JM (2004) RNA-dependent RNA polymerases of dsRNA bacteriophages. *Virus Res* **101**: 45–55.
- Mardanov AV, Strakhova TS & Ravin NV (2004) RepA protein of the bacteriophage N15 exhibits activity of DNA helicase. *Dokl Biochem Biophys* **397**: 217–219.
- Marians KJ (1996) Replication fork propagation. *Escherichia coli and Salmonella, Cellular and Molecular Biology* (Neidhardt FC, Curtiss III R, Ingraham J, Lin ECC, Low KB, Magasanik B, Reznikoff WS, Riley M, Schaechter M & Umberger HE, eds), pp. 749–763. ASM, Washington, DC.
- Marintcheva B & Weller SK (2001a) A tale of two HSV1 helicases: roles of phage and animal virus helicases in DNA replication and recombination. *Prog Nucleic Acid Res Mol Biol* **70**: 77–118.
- Marintcheva B & Weller SK (2001b) Residues within the conserved helicase motifs of UL9, the origin-binding protein of herpes simplex virus-1, are essential for helicase activity but not for dimerization or origin binding activity. *J Biol Chem* **276**: 6605–6615.
- Marsin S, McGovern S, Ehrlich SD, Bruand C & Polard P (2001) Early steps of *Bacillus subtilis* primosome assembly. *J Biol Chem* **276**: 45818–45825.
- Marszalek J & Kaguni JM (1994) DnaA protein directs the binding of DnaB protein in initiation of DNA replication in *Escherichia coli*. *J Biol Chem* **269**: 4883–4890.
- Martin G, Lázaro JM, Méndez E & Salas M (1989) Characterization of the phage phi 29 protein p5 as a single-stranded DNA binding protein. Function in phi 29 DNA-protein p3 replication. *Nucleic Acids Res* **17**: 3663–3672.
- Masai H & Arai K (1988) Operon structure of dnaT and dnaC genes essential for normal and stable DNA replication of *Escherichia coli* chromosome. *J Biol Chem* **263**: 15083–15093.
- Masai H, Nomura N & Arai K (1990) The ABC-primosome. A novel priming system employing dnaA, dnaB, dnaC, and primase on a hairpin containing a dnaA box sequence. *J Biol Chem* **265**: 15134–15144.
- Masai H, Asai T, Kubota Y, Arai K & Kogoma T (1994) *Escherichia coli* PriA protein is essential for inducible and constitutive stable DNA replication. *EMBO J* **13**: 5338–5345.
- Masterson PJ, Stanley MA, Lewis AP & Romanos MA (1998) A C-terminal helicase domain of the human papillomavirus E1 protein binds E2 and the DNA polymerase alpha-primase p68 subunit. *J Virol* **72**: 7407–7419.
- Matsubara K & Kaiser AD (1968) Lambda dv: an autonomously replicating DNA fragment. *Cold Spring Harbor Symp Quant Biol* **33**: 769–775.
- Mayr E (1998) Two empires or three? *Proc Natl Acad Sci USA* **95**: 9720–9723.
- McGarry KC, Ryan VT, Grimwade JE & Leonard AC (2004) Two discriminatory binding sites in the *Escherichia coli* replication origin are required for DNA strand opening by initiator DnaA-ATP. *Proc Natl Acad Sci USA* **101**: 2811–2816.
- McGlynn P & Lloyd RG (2000) Modulation of RNA polymerase by (p)ppGpp reveals a RecG-dependent mechanism for replication fork progression. *Cell* **101**: 35–45.
- Meijer M, Beck E, Hansen FG, Bergmans HE, Messer W, von Meyenburg K & Schaller H (1979) Nucleotide sequence of the origin of replication of the *E. coli* K-12 chromosome. *Proc Natl Acad Sci USA* **76**: 580–584.
- Meijer WJ, Horcajadas JA & Salas M (2001) Phi29 family of phages. *Microbiol Mol Biol Rev* **65**: 261–287.
- Méndez J, Blanco L, Esteban JA, Bernad A & Salas M (1992) Initiation of phi 29 DNA replication occurs at the second 3' nucleotide of the linear template: a sliding-back mechanism for protein-primed DNA replication. *Proc Natl Acad Sci USA* **89**: 9579–9583.
- Méndez J, Blanco L & Salas M (1997) Protein-primed DNA replication: a transition between two modes of priming by a unique DNA polymerase. *EMBO J* **16**: 2519–2527.

- Meselson M & Stahl F (1958) The replication of DNA in *E. coli*. *Proc Natl Acad Sci USA* **44**: 671–682.
- Messer W (2002) The bacterial replication initiator DnaA. DnaA and oriC, the bacterial mode to initiate DNA replication. *FEMS Microbiol Rev* **26**: 355–374.
- Messer W & Weigel C (1996) Initiation of chromosome replication. *Escherichia coli and Salmonella, Cellular and Molecular Biology* (Neidhardt FC, Curtiss III R, Ingraham J, Lin ECC, Low KB, Magasanik B, Reznikoff WS, Riley M, Schaechter M & Umberger HE, eds), pp. 1579–1601. ASM Press, Washington, DC.
- Messer W & Zakrewska-Czerwinska J (2002) *Streptomyces* and *Escherichia coli*, model organisms for the analysis of the initiation of bacterial chromosome replication. *Arch Immunol Ther Exp (Warsz)* **50**: 393–398.
- Mesyanzhinov VV, Robben J, Grymonprez B, Kostyuchenko VA, Bourkaltseva MV, Sykilinda NN, Krylov VN & Volckaert G (2002) The genome of bacteriophage phiKZ of *Pseudomonas aeruginosa*. *J Mol Biol* **317**: 1–19.
- Meyer TF & Geider K (1982) Enzymatic synthesis of bacteriophage fd viral DNA. *Nature* **296**: 828–832.
- Michel B, Ehrlich SD & Uzzell M (1997) DNA double-strand breaks caused by replication arrest. *EMBO J* **16**: 430–438.
- Miller ES, Kutter E, Mosig G, Arisaka F, Kunisawa T & Ruger W (2003) Bacteriophage T4 genome. *Microbiol Mol Biol Rev* **67**: 86–156.
- Mindich L (2004) Packaging, replication and recombination of the segmented genome of bacteriophage Phi6 and its relatives. *Virus Res* **101**: 83–92.
- Missich R, Weise F, Chai S, Lurz R, Pedré X & Alonso JC (1997) The replisome organizer (G38P) of *Bacillus subtilis* bacteriophage SPP1 forms specialized nucleoprotein complexes with two discrete distant regions of the SPP1 genome. *J Mol Biol* **270**: 50–64.
- Moreira D (2000) Multiple independent horizontal transfers of informational genes from bacteria to plasmids and phages: implications for the origin of bacterial replication machinery. *Mol Microbiol* **35**: 1–5.
- Morgan GJ, Hatfull GF, Casjens S & Hendrix RW (2002) Bacteriophage Mu genome sequence: analysis and comparison with Mu-like prophages in *Haemophilus*, *Neisseria* and *Deinococcus*. *J Mol Biol* **317**: 337–359.
- Moriya S, Imai Y, Hassan AK & Ogasawara N (1999) Regulation of initiation of *Bacillus subtilis* chromosome replication. *Plasmid* **41**: 17–29.
- Moscato M & Suarez JE (2000) Characterization of the DNA replication module of bacteriophage A2 and use of its origin of replication as a defense against infection during milk fermentation by *Lactobacillus casei*. *Virology* **273**: 101–111.
- Mosig G (1994) Homologous recombination. *Molecular Biology of Bacteriophage T4* (Karam J, Drake JW, Kreuzer KN, Mosig G & Hall DH, eds), pp. 54–82. ASM, Washington, DC.
- Mosig G (1998) Recombination and recombination-dependent DNA replication in bacteriophage T4. *Annu Rev Genet* **32**: 379–413.
- Mosig G, Colowick N, Gruidl ME, Chang A & Harvey AJ (1995) Multiple initiation mechanisms adapt phage T4 DNA replication to physiological changes during T4's development. *FEMS Microbiol Rev* **17**: 83–98.
- Munson RS Jr, Harrison A, Gillaspay A, et al. (2004) Partial analysis of the genomes of two nontypeable *Haemophilus influenzae* otitis media isolates. *Infect Immun* **72**: 3002–3010.
- Nair S, Alokam S, Kothapalli S, et al. (2004) *Salmonella enterica* serovar typhi strains from which SPI7, a 134-kilobase island with genes for Vi exopolysaccharide and other functions, has been deleted. *J Bacteriol* **186**: 3214–3223.
- Nakai H, Doseeva V & Jones JM (2001) Handoff from recombinase to replisome: insights from transposition. *Proc Natl Acad Sci USA* **98**: 8247–8254.
- Nelson KE, Fouts DE, Mongodin EF, et al. (2004) Whole genome comparisons of serotype 4b and 1/2a strains of the food-borne pathogen *Listeria monocytogenes* reveal new insights into the core genome components of this species. *Nucleic Acids Res* **32**: 2386–2395.
- Ng JY & Mariani KJ (1996) The ordered assembly of the phiX174-type primosome .2. Preservation of primosome composition from assembly through replication. *J Biol Chem* **271**: 15649–15655.
- Nordström K (2003) The replicon theory 40 years. An EMBO workshop held in Villefranche sur Mer, France, January 18–23, 2003. *Plasmid* **49**: 269–280.
- Nossal NG, Dudas KC & Kreuzer KN (2001) Bacteriophage T4 proteins replicate plasmids with a preformed R loop at the T4 ori (uvsY) replication origin in vitro. *Mol Cell* **7**: 31–41.
- O'Flaherty S, Coffey A, Edwards R, Meaney W, Fitzgerald GF & Ross RP (2004) Genome of staphylococcal phage K: a new lineage of myoviridae infecting Gram-positive bacteria with a low G+C content. *J Bacteriol* **186**: 2862–2871.
- Obregon V, Garcia JL, Garcia E, Lopez R & Garcia P (2003) Genome organization and molecular analysis of the temperate bacteriophage MM1 of *Streptococcus pneumoniae*. *J Bacteriol* **185**: 2362–2368.
- Ochman H & Wilson AC (1987) Evolution in bacteria: evidence for a universal substitution rate in cellular genomes. *J Mol Evol* **26**: 74–86.
- Odegrip R & Haggård-Liungquist E (2001) The two active-site tyrosine residues of the a protein play non-equivalent roles during initiation of rolling circle replication of bacteriophage p2. *J Mol Biol* **308**: 147–163.
- Park K & Chatteraj DK (2001) DnaA boxes in the P1 plasmid origin: the effect of their position on the directionality of replication and plasmid copy number. *J Mol Biol* **310**: 69–81.
- Pedulla ML, Ford ME, Houtz JM, et al. (2003) Origins of highly mosaic mycobacteriophage genomes. *Cell* **113**: 171–182.
- Penny D & Poole A (1999) The nature of the last universal common ancestor. *Curr Opin Genet Dev* **9**: 672–677.
- Petit MA & Ehrlich SD (2002) Essential bacterial helicases that counteract the toxicity of recombination proteins. *EMBO J* **21**: 3137–3147.

- Petit MA, Dervyn E, Rose M, Entian KD, McGovern S, Ehrlich SD & Bruand C (1998) PcrA is an essential DNA helicase of *Bacillus subtilis* fulfilling functions both in repair and rolling-circle replication. *Mol Microbiol* **29**: 261–273.
- Pfeiffer P & Hohn T (1983) Involvement of reverse transcription in the replication of cauliflower mosaic virus: a detailed model and test of some aspects. *Cell* **33**: 781–789.
- Poranen MM & Tuma R (2004) Self-assembly of double-stranded RNA bacteriophages. *Virus Res* **101**: 93–100.
- Postow L, Hardy CD, Arsuaaga J & Cozzarelli NR (2004) Topological domain structure of the *Escherichia coli* chromosome. *Genes Dev* **18**: 1766–1779.
- Rahmouni AR & Wells RD (1992) Direct evidence for the effect of transcription on local DNA supercoiling in vivo. *J Mol Biol* **223**: 131–144.
- Rakonjac J, Ward LJH, Schiemann AH, Gardner PP, Lubbers MW & O'Toole PW (2003) Sequence diversity and functional conservation of the origin of replication in lactococcal prolate phages. *Appl Environ Microbiol* **69**: 5104–5114.
- Ravin NV (2003) Mechanisms of replication and telomere resolution of the linear plasmid prophage N15. *FEMS Microbiol Lett* **221**: 1–6.
- Ravin NV, Strakhova TS & Kuprianov VV (2001) The protelomerase of the phage-plasmid N15 is responsible for its maintenance in linear form. *J Mol Biol* **312**: 899–906.
- Ravin NV, Kuprianov VV, Gilcrease EB & Casjens SR (2003) Bidirectional replication from an internal ori site of the linear N15 plasmid prophage. *Nucleic Acids Res* **31**: 6552–6560.
- Rawlings DE & Tietze E (2001) Comparative biology of IncQ and IncQ-like plasmids. *Microbiol Mol Biol Rev* **65**: 481–496.
- Richardson CC (1983) Bacteriophage T7: minimal requirements for the replication of a duplex DNA molecule. *Cell* **33**: 315–317.
- Richter S, Hagemann M & Messer W (1998) Transcriptional analysis and mutation of a *dnaA*-like gene in *Synechocystis* sp. PCC6803. *J Bacteriol* **180**: 4946–4949.
- Roberts MD, Martin NL & Kropinski AM (2004) The genome and proteome of coliphage T1. *Virology* **318**: 245–266.
- Roth A & Messer W (1995) The DNA binding domain of the initiator protein DnaA. *EMBO J* **14**: 2106–2111.
- Saito H, Tabor S, Tamanoi F & Richardson CC (1980) Nucleotide sequence of the primary origin of bacteriophage T7 DNA replication: relationship to adjacent genes and regulatory elements. *Proc Natl Acad Sci USA* **77**: 3917–3921.
- Salinas F & Benkovic SJ (2000) Characterization of bacteriophage T4-coordinated leading- and lagging-strand synthesis on a minicircle substrate. *Proc Natl Acad Sci USA* **97**: 7196–7201.
- Sandler SJ (2000) Multiple genetic pathways for restarting DNA replication forks in *Escherichia coli* K-12. *Genetics* **155**: 487–497.
- Sandler SJ & Mariani KJ (2000) Role of PriA in replication fork reactivation in *Escherichia coli*. *J Bacteriol* **182**: 9–13.
- Sandler SJ, Mariani KJ, Zavitz KH, Coutu J, Parent MA & Clark AJ (1999) *dnaC* mutations suppress defects in DNA replication- and recombination-associated functions in *priB* and *priC* double mutants in *Escherichia coli* K-12. *Mol Microbiol* **34**: 91–101.
- Schekman R, Weiner JH, Weiner A & Kornberg A (1975) Ten proteins required for conversion of phiX174 single-stranded DNA to duplex form in vitro. Resolution and reconstitution. *J Biol Chem* **250**: 5859–5865.
- Scherzinger E, Haring V, Lurz R & Otto S (1991) Plasmid RSF1010 DNA replication in vitro promoted by purified RSF1010 RepA, RepB and RepC proteins. *Nucleic Acids Res* **19**: 1203–1211.
- Scherzinger E, Ziegelin G, Bárcena M, Carazo JM, Lurz R & Lanka E (1997) The RepA protein of plasmid RSF1010 is a replicative DNA helicase. *J Biol Chem* **272**: 30228–30236.
- Schnos M, Zahn K, Inman RB & Blattner FR (1988) Initiation protein induced helix destabilization at the lambda origin: a prepriming step in DNA replication. *Cell* **52**: 385–395.
- Seigneur M, Bidnenko V, Ehrlich SD & Michel B (1998) RuvAB acts at arrested replication forks. *Cell* **95**: 419–430.
- Seitz H, Weigel C & Messer W (2000) The interaction domains of the DnaA and DnaB replication proteins of *Escherichia coli*. *Mol Microbiol* **37**: 1270–1279.
- Serrano M, Gutiérrez J, Prieto I, Hermoso JM & Salas M (1989) Signals at the bacteriophage phi 29 DNA replication origins required for protein p6 binding and activity. *EMBO J* **8**: 1879–1885.
- Serrano M, Gutiérrez C, Salas M & Hermoso JM (1993) Superhelical path of the DNA in the nucleoprotein complex that activates the initiation of phage phi 29 DNA replication. *J Mol Biol* **230**: 248–259.
- Serwer P, Watson RH & Son M (1990) Role of gene 6 exonuclease in the replication and packaging of bacteriophage T7 DNA. *J Mol Biol* **215**: 287–299.
- Seufert W & Messer W (1986) Initiation of *Escherichia coli* minichromosome replication at *oriC* and at protein n' recognition sites. Two modes for initiating DNA synthesis in vitro. *EMBO J* **5**: 3401–3406.
- Sippy J & Feiss M (2004) Initial cos cleavage of bacteriophage lambda concatemers requires proheads and gpFI in vivo. *Mol Microbiol* **52**: 501–513.
- Smith MP & Feiss M (1993) Sites and gene products involved in lambdoid phage DNA packaging. *J Bacteriol* **175**: 2393–2399.
- Soengas MS, Gutiérrez C & Salas M (1995) Helix-destabilizing activity of phi 29 single-stranded DNA binding protein: effect on the elongation rate during strand displacement DNA replication. *J Mol Biol* **253**: 517–529.
- del Solar G, Giraldo R, Ruiz-Echevarria MJ, Espinosa M & Díaz Orejas R (1998) Replication and control of circular bacterial plasmids. *Microbiol Mol Biol Rev* **62**: 434–464.
- Son M, Watson RH & Serwer P (1993) The direction and rate of bacteriophage T7 DNA packaging in vitro. *Virology* **196**: 282–289.
- Speck C & Messer W (2001) Mechanism of origin unwinding: sequential binding of DnaA initiator protein to double-stranded and single-stranded DNA in the AT-rich region of the replication origin. *EMBO J* **20**: 1469–1476.

- Stahl FW, Kobayashi I & Stahl MM (1985) In phage lambda, *cos* is a recombinator in the red pathway. *J Mol Biol* **181**: 199–209.
- Stahl MM, Thomason L, Potete AR, Tarkowski T, Kuzminov A & Stahl FW (1997) Annealing vs. invasion in phage lambda recombination. *Genetics* **147**: 961–977.
- Stanley E, Walsh L, van der Zwet A, Fitzgerald GF & van Sinderen D (2000) Identification of four loci isolated from two *Streptococcus thermophilus* phage genomes responsible for mediating bacteriophage resistance. *FEMS Microbiol Lett* **182**: 271–277.
- Stayton MM & Kornberg A (1983) Complexes of *Escherichia coli* primase with the replication origin of G4 phage DNA. *J Biol Chem* **258**: 13205–13212.
- Stayton MM, Bertsch L, Biswas S, *et al.* (1983) Enzymatic recognition of DNA replication origins. *Cold Spring Harb Symp Quant Biol* **47**(II): 693–700.
- Su'tsugu M, Shimuta TR, Ishida T, Kawakami H & Katayama T (2004) Protein associations in DnaA-ATP hydrolysis mediated by the replicase clamp-Hda complex. *J Biol Chem* **280**: 6528–6.
- Sugimoto K, Oka A, Sugisaki H, Takanami M, Nishimura A, Yasuda Y & Hirota Y (1979) Nucleotide sequence of *Escherichia coli* replication origin. *Proc Natl Acad Sci USA* **76**: 575–579.
- Summer EJ, Gonzalez CF, Carlisle T, Mebane LM, Cass AM, Savva CG, LiPuma JJ & Young R (2004) *Burkholderia cenocepacia* phage BcepMu and a family of Mu-like phages encoding potential pathogenesis factors. *J Mol Biol* **340**: 49–65.
- Takahashi S (1975) The starting point and direction of rolling-circle replicative intermediates of coliphage lambda DNA. *Mol Gen Genet* **142**: 137–153.
- Takahashi S, Hours C, Chu A & Denhardt DT (1979) The *rep* mutation. VI. Purification and properties of the *Escherichia coli rep* protein, DNA helicase III. *Can J Biochem* **57**: 855–866.
- Tamanoi F, Saito H & Richardson CC (1980) Physical mapping of primary and secondary origins of bacteriophage T7 DNA replication. *Proc Natl Acad Sci USA* **77**: 2656–2660.
- Tatusova TA & Madden TL (1999) BLAST 2 sequences, a new tool for comparing protein and nucleotide sequences. *FEMS Microbiol Lett* **174**: 247–250.
- Taylor K & Wegrzyn G (1995) Replication of coliphage lambda DNA. *FEMS Microbiol Rev* **17**: 109–119.
- Tétart F, Desplats C, Kutateladze M, Monod C, Ackermann HW & Krisch HM (2001) Phylogeny of the major head and tail genes of the wide-ranging T4-type bacteriophages. *J Bacteriol* **183**: 358–366.
- Thomson N, Sebahia M, Cerdano-Tarraga A & Parkhill J (2002) Sibling rivalry. *Trends Microbiol* **10**: 396–397.
- Tougu K & Mariani KJ (1996) The extreme C terminus of primase is required for interaction with DnaB at the replication fork. *J Biol Chem* **271**: 21391–21397.
- Toussaint A & Merlin C (2002) Mobile elements as a combination of functional modules. *Plasmid* **47**: 26–35.
- Tye BK (2000) Insights into DNA replication from the third domain of life. *Proc Natl Acad Sci USA* **97**: 2399–2401.
- Venkatesan M, Silver LL & Nossal NG (1982) Bacteriophage T4 gene 41 protein, required for the synthesis of RNA primers, is also a DNA helicase. *J Biol Chem* **257**: 12426–12434.
- Watson JD & Crick F (1953) Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature* **171**: 737.
- Weigel C & Seitz H (2002) Strand-specific loading of DnaB helicase by DnaA to a substrate mimicking unwound oriC. *Mol Microbiol* **46**: 1149–1156.
- Weigel C, Schmidt A, Seitz H, Tuengler D, Welz M & Messer W (1999) The N-terminus promotes oligomerisation of the *Escherichia coli* initiator protein DnaA. *Mol Microbiol* **34**: 53–66.
- Weise F, Chai S, Lüder G & Alonso JC (1994) Nucleotide sequence and complementation studies of the gene 35 region of the *Bacillus subtilis* bacteriophage SPP1. *Virology* **202**: 1046–1049.
- Wever GH, Fischer H & Hinkle DC (1980) Bacteriophage T7 DNA replication in vitro. Electron micrographic analysis of T7 DNA synthesized with purified proteins. *J Biol Chem* **255**: 7965–7972.
- White JH & Richardson CC (1987) Processing of concatemers of bacteriophage T7 DNA in vitro. *J Biol Chem* **262**: 8851–8860.
- Wickner S (1984a) DNA-dependent ATPase activity associated with phage P22 gene 12 protein. *J Biol Chem* **259**: 14038–14043.
- Wickner S (1984b) Oligonucleotide synthesis by *Escherichia coli* dnaG primase in conjunction with phage P22 gene 12 protein. *J Biol Chem* **259**: 14044–14047.
- Woese CR (2000) Interpreting the universal phylogenetic tree. *Proc Natl Acad Sci USA* **97**: 8392–8396.
- Woese CR (2002) On the evolution of cells. *Proc Natl Acad Sci USA* **99**: 8742–8747.
- Worcel A & Burgi E (1972) On the structure of the folded chromosome of *Escherichia coli*. *J Mol Biol* **71**: 127–147.
- Wrobel B & Wegrzyn G (2002) Evolution of lambdoid replication modules. *Virus Genes* **24**: 163–171.
- Yuzenkova J, Nechaev S, Berlin J, Rogulja D, Kuznedelov K, Inman R, Mushegian A & Severinov K (2003) Genome of *Xanthomonas oryzae* bacteriophage Xp10: an odd T-odd phage. *J Mol Biol* **330**: 735–748.
- Zavitz KH & Mariani KJ (1992) ATPase-deficient mutants of the *Escherichia coli* DNA replication protein PriA are capable of catalyzing the assembly of active primosomes. *J Biol Chem* **267**: 6933–6940.
- Zechner EL, Wu CA & Mariani KJ (1992) Coordinated leading- and lagging-strand synthesis at the *Escherichia coli* DNA replication fork. III. A polymerase-primase interaction governs primer size. *J Biol Chem* **267**: 4054–4063.
- Ziegelin G, Tegtmeyer N, Lurz R, Hertwig S, Hammerl J, Appel B & Lanka E (2005) The *repA* gene of the linear *Yersinia enterocolitica* prophage PY54 functions as a circular minimal replicon in *Escherichia coli*. *J Bacteriol* **187**: 3445–3454.
- Zylicz M, Liberek K, Wawrzynow A & Georgopoulos C (1998) Formation of the preprimosome protects lambda O from RNA transcription-dependent proteolysis by ClpP/ClpX. *Proc Natl Acad Sci USA* **95**: 15259–15263.