Mu *in vitro* Transposition Technology in Functional Genetics and Genomics: Applications on Mouse and Bacteriophages

HEIKKI VILEN

Institute of Biotechnology
and
Department of Biological and Environmental Sciences
Division of Genetics
Faculty of Biosciences
and
Helsinki Graduate School in Biotechnology and
Molecular Biology
University of Helsinki

Dissertationes bioscientiarum molecularium Universitatis Helsingiensis in Viikki

7/2006

Mu *in vitro* Transposition Technology in Functional Genetics and Genomics: Applications on Mouse and Bacteriophages

Heikki Vilen

Institute of Biotechnology
And
Department of Biological and Environmental Sciences
Division of Genetics
Faculty of Biosciences
And
Helsinki Graduate School in Biotechnology
and Molecular Biology

University of Helsinki

ACADEMIC DISSERTATION

To be presented, with the permission of the Faculty of Biosciences of the University of Helsinki, for public criticism in auditorium B3 at Forestry Sciences Building (Entrance from Viikki Campus A-building, Latokartanonkaari 9) on 5 May 2006, at 12 noon.

Supervisor

Docent Harri Savilahti,
Institute of Biotechnology,
University of Helsinki

Reviewers

Professor Jari Valkonen
Department of Applied Biology,
University of Helsinki

And

Docent Päivi Ojala
Institute of Biomedicine/Biochemistry,
Biomedicum Helsinki,
University of Helsinki

Opponent

Docent Sirkka Keränen
University of Helsinki

7/2006

ISBN 952-10-3099-2 (paperback)
ISBN 952-10-3100-X (PDF, online)
ISSN 1795-7079 (paperback)
ISSN 1795-8229 (PDF, online)

Gummerus Kirjapaino Oy
Vaajakoski 2006

ORIGINAL PUBLICATIONS

This thesis is based on the following articles, which are referred to in the text by their Roman numerals.

I Vilen H., Eerikäinen S., Tornberg J., Airaksinen M.S., and Savilahti H. 2001: Construction of gene-targeting vectors: a rapid Mu *in vitro* DNA transposition - based strategy generating null, potentially hypomorphic, and conditional alleles. *Transgenic Res.* **10**: 69-80.

II Vilen H., Aalto J-M., Kassinen A., Paulin L., and Savilahti H. 2003: A direct transposon insertion tool for modification and functional analysis of viral genomes. *J. Virol.* **77**: 123-34.

III Kiljunen S., Vilen H., Pajunen M., Savilahti H., and Skurnik M. 2005: Nonessential genes of phage ΦYeO3-12 include genes involved in adaptation to growth on *Yersinia enterocolitica* serotype O:3. *J. Bacteriol.* **187**: 1405-14.

IV Krupovič M., Vilen H., Bamford J.K.H., Kivelä H.M., Aalto J-M., Savilahti H., and Bamford D.H.: Genome characterisation of lipid-containing marine bacteriophage PM2 by transposon insertional mutagenesis. Submitted.

TABLE OF CONTENTS

ABBREVIATIONS

AGE	agarose gel electrophoresis
ATP	adenosine triphosphate
BAC	bacterial artificial chromosome
bp	base pair(s)
CDC	cleaved donor complex
DNA	deoxyribonucleic acid
DMSO	dimethyl sulfoxide
ds-	double-stranded
ER	enhancer - right end complex
ES cells	embryonic stem cells
GAMBIT	genome analysis and mapping by *in vitro* transposition
Gb	Gigabase(pairs), 10^9 base pairs
HPLC	high pressure liquid chromatography
IAS	internal activation sequence
IR	inverted repeat(s)
IS	insertion sequence
kb	kilobase(pairs), 10^3 base pairs
LER	left end - enhancer - right end complex
LTR	long terminal repeat(s)
Mb	Megabase(pairs), 10^6 base pairs
MITE	miniature inverted-repeat transposable element
nt	nucleotide
ORF	open reading frame
PCR	polymerase chain reaction
RNA	ribonucleic acid
RT	reverse transcriptase
ss-	single-stranded
SSC	stable synaptic complex
STC	strand transfer complex
STM	signature-tagged mutagenesis
TE	transposable element
Tn	transposon
wt	wild type

A. ABSTRACT

Transposons are mobile elements of genetic material that are able to move in the genomes of their host organisms using a special form of recombination called transposition that does not require extensive target sequence homology. Bacteriophage Mu was the first transposon for which a cell-free *in vitro* transposition reaction was developed. Subsequently, the reaction has been refined and the minimal Mu *in vitro* reaction is useful in the generation of comprehensive libraries of mutant DNA molecules that can be used in a variety of applications. However, to date, the functional genetics applications of Mu *in vitro* technology have been subjected to either plasmids or genomic regions and entire genomes of viruses cloned on specific vectors.

The present study expands the use of Mu *in vitro* transposition in functional genetics and genomics by describing novel methods applicable to the targeted transgenesis of mouse (*Mus musculus*) and the whole-genome analysis of bacteriophages. The methods described here are rapid, efficient, and easily applicable to a wide variety of organisms, demonstrating the potential of the Mu *in vitro* transposition technology in the functional analysis of genes and genomes.

First, an easy-to-use, rapid strategy to generate construct for the targeted mutagenesis of mouse genes was developed. To test the strategy, a fragment of *KCC2*, a gene encoding a neuronal K^+/Cl^- cotransporter was mutagenised. After a highly efficient *in vitro* transpositional mutagenesis with a transposon containing a neomycin resistance marker gene flanked by *loxP* sites, the gene fragments mutagenised were separated, cloned into a vector backbone, and subsequently transferred into bacterial cells. The desired constructs were screened with polymerase chain reaction using an effective three-dimensional matrix system. In addition to traditional knock-out constructs, the method developed yields hypomorphic alleles that lead into reduced expression of the target gene in transgenic mice. These alleles have since been used in a follow-up study. Moreover, a scheme is devised to rapidly produce conditional alleles from the constructs produced.

Next, an efficient strategy for the whole-genome analysis of bacteriophages was developed based on the transpositional mutagenesis of uncloned, infective virus genomes and their subsequent transfer into susceptible host cells. Mutant viruses able to produce viable progeny were collected using the blue/white selection, autoradiography, or based on mobility in agarose gel electroforesis and their transposon integration sites were determined to map genomic regions nonessential to the viral life cycle. This method, applied here to three very different bacteriophages, PRD1, ΦYeO3-12, and PM2 infecting different host bacteria, does not require the target genome to be cloned and is directly applicable to all DNA and RNA viruses that have infective genomes. The method developed yielded valuable novel information on the three bacteriophages studied and whole-genome data can be complemented with concomitant studies on individual genes. Moreover, end-modified transposons constructed for this study can be used to manipulate genomes devoid of suitable restriction sites.

B. INTRODUCTION

1. TRANSPOSONS

Transposons (or transposable elements, TEs) are mobile elements of genetic material that can move from one site in a genome into another site or even into another genome by transposition, a special form of recombination that generally does not require extensive target sequence homology.

Transposable elements are almost universal: they have been found in the genomes of virtually all the organisms that have been studied in any detail including Bacteria (reviewed in Campbell, 2002), Archaea (reviewed in Brugger *et al.*, 2002), and eukaryotic kingdoms of Fungi (reviewed in Kempken and Kuck, 1998), (Green) Plants, and Animals (reviewed in Deininger and Roy-Engel, 2002). Some symbiotic obligate intracellular bacteria are known to lack TEs, apparently due to the evolutional pressure to genome streamlining (Andersson *et al.*, 2002, Bordenstein and Reznikoff, 2005).

1.1. Transposon classes

Transposons can be divided into two classes according to whether they have an RNA intermediate in their lifecycle or not. The terminology of these classes, that can be further divided into subclasses, has been conflicting and often confusing.

1.1.1. Class I transposons

Class I transposons have an RNA intermediate in their life cycle. The genomic element is transcribed into an RNA copy by an RNA polymerase and converted back to DNA by a reverse transcriptase (RT). These retroelements can be further divided into two classes according to whether DNA or RNA is the actual substrate for recombination.

Retrotransposons (or LTR retrotransposons, retroviral-like elements) have long terminal repeats (LTRs), directly repeated segments at their ends. In their life cycle, the RNA element is first reverse-transcribed into DNA and this DNA is then used as a substrate for transposition into the host genome. Their composition resembles animal retroviruses. Endogenous retroviruses residing in the genome and LTR retrotransposons are collectively referred to as LTR elements.

Retroposons (or non-LTR transposons) are mobile elements in which the RNA form acts as the transposition substrate by attaching directly to a nick in the DNA sequence that serves as a primer to the reverse transcription reaction. The DNA strand transcribed is thus covalently linked into the target DNA (Luan *et al.*, 1993).

1.1.2. Class II transposons (DNA transposons)

Class II transposons move in the genomes solely as discrete DNA elements with no RNA intermediates involved. Their movement is catalysed by one or several self-coded transposases that recognise specific end sequences at each end of the transposon. These terminal sequences are most commonly in inverse orientation (inverted repeats, IR). Class II elements are usually referred to as DNA transposons. In Bacteria, where class II elements of widely varied complexity are found, the simplest elements are called insertion sequences (ISs), whereas the term transposon traditionally refers to the more complex transposons (see chapter 1.2.1.).

In some, mostly eukaryotic genomes there are short (~100...500 bp), non-autonomous elements that are present in high copy numbers. These miniature inverted-repeat transposable elements (MITEs) have conserved terminal repeats and target-site specificity, but no coding potential (reviewed in Feschotte *et al.*, 2002). Because of their distinct properties they have sometimes been classified as class III transposons (used in e.g. Yu *et al.*, 2002) However, it is evident that these elements originated from a subset of existing DNA transposons and thus are classified as a subclass of class II transposons (Turcotte *et al.*, 2001, Feschotte *et al.*, 2002, Kidwell, 2002).

1.2. Transposons in different species

The common denominator in the discovery of transposable elements in model species – maize, bacteria, and fruit fly – was that they were found by accident, as a by-product of studying other phenomena (Shapiro, 1995). They were initially labelled "selfish DNA" with no evident benefits to the cell itself (Orgel and Crick, 1980, Doolittle and Sapienza, 1980, Doolittle *et al.*, 1984). As the idea that TEs provide evolutionarily important fluidity to the genomes has gained more ground (for example Blot, 1994) and extensive sequencing projects have provided genomic data from a variety of species, the occurrence of transposons has been studied as part of whole-genome sequencing projects (for example Lander *et al.*, 2001, Wood *et al.*, 2002, Hillier *et al.*, 2004) or as a separate analysis from whole-genome sequence data (for example Kim *et al.*, 1998, Kaminker *et al.*, 2002).

Generally, small genomes have few TEs while in large genomes there is an abundance of mobile elements. The proportion of transposable elements in the genomes of different species is represented in Table 1.

1.2.1. Bacterial transposons

There is a remarkable structural variety in bacterial DNA transposons ranging from simple insertion sequences (IS) and composite transposons to more complex elements and bacteriophages that use transposition as a lifestyle (Figure 1). In addition, the study of bacterial genomes has revealed a new class of elements, retrons, that are potentially proliferating though RNA-mediated transposition (for a recent review see Lampson *et al.*, 2005).

1.2.1.1. Bacteriophage Mu and other transposable viruses

The most complex transposons identified are bacteriophages that use transposition as a lifestyle. The temperate bacteriophage Mu (short for Mutator), isolated from the city sewage of Denver in Colorado, USA, was originally identified on the basis of its ability to cause mutations in a variety of Gram-negative bacteria including *Escherichia coli* (Taylor, 1963). The 36.7-kb Mu genome (Figure 1D) consists of two transposase genes and terminal inverted end sequences in addition to 53 genes involved in other functions of the Mu life cycle (Howe, 1987, Morgan *et al.*, 2002).

Bacteriophage Mu is exceptional in that it utilises DNA transposition in two different ways (see chapter 2.1). During the initial infection it integrates into the bacterial host genome by conservative transposition (Harshey, 1984). However, during lytic growth Mu replicates itself by multiple rounds of replicative transposition (Chaconas *et al.*, 1981) (see chapter 2.2).

Table 1. Transposable element contents of some model organisms

species	common name	genome size (Mb)	genomic proportion of		reference
			TEs (%)	DNA transposons (%)	
Escherichia coli	(eubacterium)	4.6	2.0	2.0	Blattner *et al.*, 1997
Burkholderia mallei	(eubacterium)	5.8	3.1	3.1	Nierman *et al.*, 2004
Methanobacterium thermoautotropicum	(archaeon)	1.8	0	0	Smith *et al.*, 1997
Sulfolobus solfataricus	(archaeon)	3.0	10.5	10.5	She *et al.*, 2001
Saccharomyces cerevisiae	budding yeast	12	3.1	0	Kim *et al.*, 1998
Schizosaccharomyces pombe	fission yeast	14	0.35	0	Wood *et al.*, 2002
Aspergillus oryzae	(filamentous fungus)	36	1.4	0.7	Galagan *et al.*, 2005
Magnaporthe grisea	(fungal plant pathogen)	40	9.7	2.0	Dean *et al.*, 2005
Arabidopsis thaliana	mustard weed	125	10	3	Arabidopsis Genome Initiative., 2000, Haas *et al.*, 2005
Oryza sativa	rice	430	13*	7*	Yu *et al.*, 2005
Aegilops tauschii	Tausch's goatgrass	4000	>90*	13*	Li *et al.*, 2004
Caenorhabditis elegans	(nematode worm)	100	6.5	5.3	Lander *et al.*, 2001
Drosophila melanogaster	fruit fly	121	3.9†	0.3†	Biemont and Vieira, 2005
Anopheles gambiae	malaria mosquito	280	16†	?	Holt *et al.*, 2002
Fugu rubripes	torafugu (bufferfish)	370	<5	?	Aparicio *et al.*, 2002
Gallus gallus	chicken	1100	9	0.8	Hillier *et al.*, 2004
Canis familiaris	dog	2400	35	2.0	Lindblad-Toh *et al.*, 2005
Mus musculus	mouse	2500	39	0.9	Waterston *et al.*, 2002, Lindblad-Toh *et al.*, 2005
Homo sapiens	man	2900	46	3.0	Lander *et al.*, 2001, Lindblad-Toh *et al.*, 2005

* estimate
† euchromatin only

The Mu transposition, extensively studied for years (for reviews see Mizuuchi and Craigie, 1986, Mizuuchi, 1992, Chaconas *et al.*, 1996, Chaconas and Harshey, 2002), was the first transposition reaction for which a cell-free *in vitro* system was developed (Mizuuchi, 1983). Bacteriophage Mu transposition has been used extensively both *in vivo* and *in vitro* in molecular biology applications (see chapter 3).

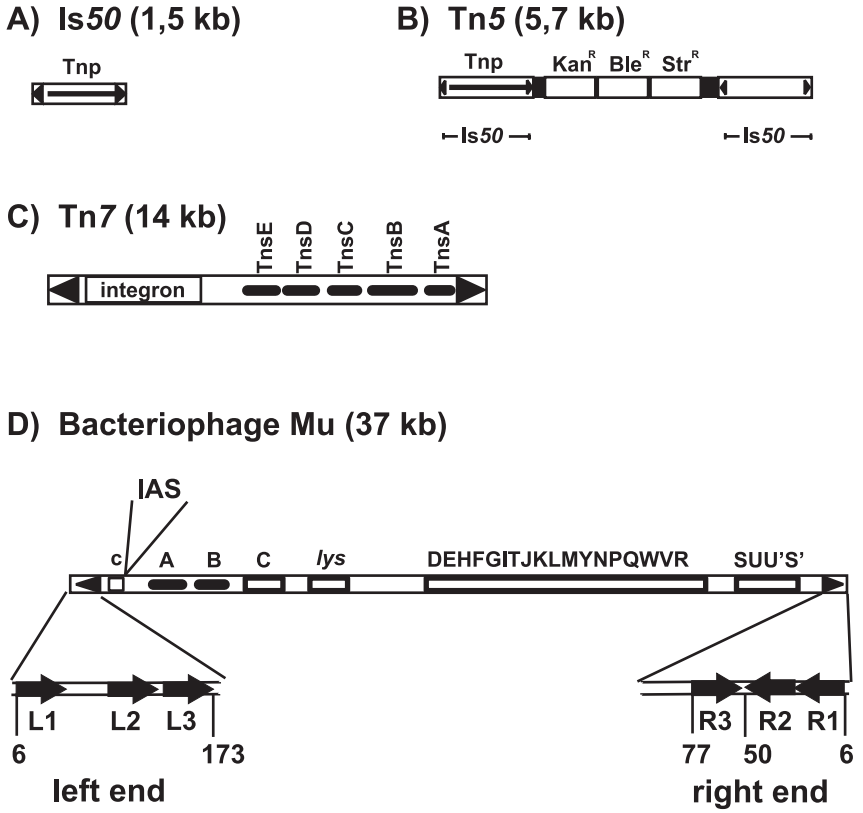

Figure 1. Examples of different bacterial transposons (not to scale). A) Insertion sequences (IS) typically only have transposase gene(s) (Tnp). Inverted repeats (IR, black triangles) are essential to mobility (drawn according to Reznikoff, 2003). B) Composite transposons consist of two ISs separated by one or several genes unrelated to transposition but often conferring a selective advantage to the host. In Tn*5* only one of the ISs encodes functional transposase. Kan^R, Ble^R, Str^R = genes encoding resistance to kanamycin, bleomycin, and streptomycin, respectively (Reznikoff, 2003). C) Complex transposons have more complicated structure consisting of several genes involved in transposition (TnsA-E in Tn*7*) and no ISs at their ends. Tn*7* has an integron containing several antibiotic resistance cassettes (Peters and Craig, 2001). D) Bacteriophage Mu has two genes involved in transposition (black). Other genes (white) are related to non-transposition functions. IAS (Internal Activator Sequence), ~100 bp-long enhancer region needed for efficient transposition in vivo is situated ~1000 bp from the left end (Howe, 1987, Morgan et al., 2002). Regions involved in transposition at each end are magnified and MuA transposase protein binding sites and their orientations are depicted with black arrows. Numbers indicate distance (in bp) from proximal transposon end (Craigie *et al.*, 1984).

Apart from Mu, only a few other transposable bacteriophages have been isolated. Bacteriophage D108, a closely related virus of *E. coli*, was discovered in Japan in 1971 (Mise, 1971). A large number of transposable bacteriophages of *Pseudomonas* genus were isolated in the former Soviet Union in the 1970s and 80s (Akhverdian *et al.*, 1984 – reviewed in DuBow, 1987). Moreover, bacterial genome-wide sequencing projects have revealed Mu-like prophages at least in *Haemophilus*, *Neisseria*, and *Deinococcus* (Morgan *et al.*, 2002).

1.2.1.2. Transposon content in bacterial genomes

There are large variations in the transposon content between related bacterial species and even between different strains of the same species. In the 4.6-Mb genome of *Escherichia coli* strain K-12 mobile elements, including transposable elements, bacteriophages, and plasmids, comprise 2.0 %. At least 42 IS copies were identified in the genome representing five different families. The distribution of IS elements was notably non-random as two multicomponent clusters were detected (Blattner *et al.*, 1997). There are notable differences in the occurrence and abundance of ISs between different *E. coli* strains (Sawyer *et al.*, 1987).

No insertion sequences were detected in the 4.2-Mb genome of Gram-positive *Bacillus subtilis*, whereas there were 93 and 10 such elements in closely related *B. halodurans* and *B. licheniformis*, respectively. These differences indicate that horizontal transfer of ISs might play an important role in the bacterial speciation (Kunst *et al.*, 1997, Takami *et al.*, 2000, Rey *et al.*, 2004).

Some symbiotic obligate intracellular bacteria that cannot survive outside their host cells and transmit themselves vertically from mother to offspring are known to lack TEs, apparently because of their constrained access to novel gene pools and the evolutional pressure to genome streamlining (Andersson *et al.*, 2002, Bordenstein and Reznikoff, 2005). Nevertheless, even parasitic obligate intracellular bacteria that have the ability to switch hosts and transmit themselves horizontally seem to have transposable elements in their genomes (Nierman *et al.*, 2004, Bordenstein and Reznikoff, 2005). The causative agent of glanders, *Burkholderia mallei* is a highly evolved obligate parasite of horses. In its 5.8 Mb genome ISs comprise 3.1 % in 171 complete or partial elements (Nierman *et al.*, 2004).

1.3. Transposon-like phenomena in cells

Apart from transposable elements themselves, other cellular phenomena that use mechanisms closely related to transposition range from parasitic viruses to functions essential to the host cells and organisms.

1.3.1. Retroviruses

Retroviruses have their genome in the form of RNA in viral particles. Following infection they are able to reverse-transcribe their genome into dsDNA and insert it into the host genome. DNA is processed and joined into the genome by a specific enzyme, the integrase, that resembles transposases of DNA transposons (Polard and Chandler, 1995). Mechanistically retroviral integration is very similar to replicative transposition (reviewed in Craig, 1995, see chapter 2). It has been proposed that retroviruses originated from an LTR retrotransposon that acquired the

env gene, a transmembrane host receptor-binding protein responsible for virus transmission (Boeke and Stoye, 1997, Eickbush and Malik, 2002).

1.3.2. V(D)J recombination

In most vertebrates the diversity of immune system is mediated by V(D)J recombination, the rearrangement of gene segments during the maturation of B and T cell lymphocytes. The parallels between V(D)J recombination and transposition are remarkable (Lewis and Wu, 1997, Agrawal *et al.*, 1998, Gellert, 2002, Brandt and Roth, 2004). V(D)J recombination is mediated by RAG1 and RAG2 proteins that break both strands of DNA precisely at the border between protein-coding and neighbouring recombination signal segments. In contrast to transposition, the ends of remaining flanking DNA are processed and joined in V(D)J, whereas the ends that contain the recombination signal segments are simply circularised and released (reviewed in Brandt and Roth, 2004). RAG1 apparently belongs to the DDE transposase family that have a core of three catalytic acidic residues essential to DNA cleavage and strand transfer (Haren *et al.*, 1999b, Landree *et al.*, 1999, Zhou *et al.*, 2004) and RAG system has the ability to perform transposition *in vitro* (Agrawal *et al.*, 1998, Hiom *et al.*, 1998). Intronless *RAG* genes lie side-to-side in the genome, resembling more genes of bacteria than those of higher eukaryotes (Oettinger *et al.*, 1990, Thompson, 1995). Moreover, housefly DNA transposon *Hermes* processes the DNA ends flanking the element analogously to V(D)J (Zhou *et al.*, 2004), reinforcing the view that V(D)J recombination arose when an ancient transposon was harnessed by the lymphocytes (Lewis and Wu, 1997, Agrawal *et al.*, 1998, Gellert, 2002, Oettinger, 2004, Zhou *et al.*, 2004).

1.3.3. Telomere-associated retrotransposons

The linear chromosomes of most eukaryotic cells are protected at their ends by telomeres, simple repeats that are periodically extended by the enzyme telomerase that synthesises new repeat sequence by reverse transcription (reviewed in Pardue and DeBaryshe, 1999). Because of the similarity in their catalytic mechanisms and the phylogenetic relation of their sequences (Lingner *et al.*, 1997), it has been suggested that telomerase enzymes might have evolved by cellular recruitment of a retroposon RT gene (Zimmerly *et al.*, 1995, Eickbush, 1997). This view is reinforced by the fact that in *Drosophila* telomeres are maintained by specialised retroposons instead of the telomerase enzyme (Biessmann *et al.*, 1992, Levis *et al.*, 1993, Abad *et al.*, 2004).

1.4. Why to study transposons?

Transposons are ubiquitous components of virtually all cells. How transposition occurs and what factors determine its frequency is of considerable interest as transposons can play a profound role in genome evolution and in a variety of genetic diseases. Transposition offers a potent mechanism to introduce a variety of mutations and alter the expression of genes to analyse their functions. Moreover, other important phenomena in the cells exploit mechanisms similar to transposition. Consequently, understanding mechanisms and control of transposition can help to define processes involved, for instance, in retrovirus infections or vertebrate

immune reactions. Moreover, transposons can be utilised in molecular biology to carry out such basic tasks as cloning and sequencing.

2. Mechanisms of DNA-mediated transposition

Most DNA-mediated transposition reactions described, including LTR retrotransposition and retroviral integration, occur by similar DNA breakage and joining reactions, demonstrating remarkable unity (Figure 2). Transposition proceeds by endonucleolytic cleavage of the phosphodiester bonds at the exact ends of the TE and transfer of these ends into a target DNA molecule. Breakage events precisely expose 3' tips of TEs that are joined to the exposed 5' ends of cleaved target DNA. These functions are catalysed by one or several TE-encoded transposase proteins. The last steps of a transposition reaction, repair of remaining gaps and possible replication of the transposon, are carried out by the host replication mechanism (reviewed in Craig, 1995).

2.1. Conservative vs. replicative transposition

According to whether a TE is copied in the process, transposition reactions can be divided into conservative and replicative events (Figure 2).

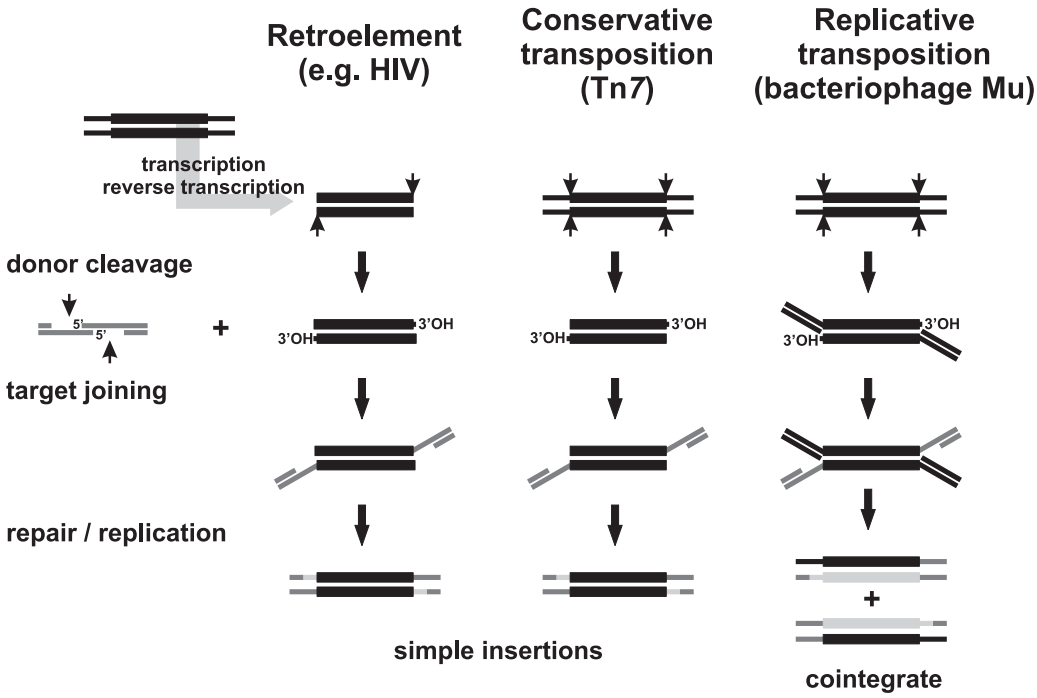

Figure 2. Unity in transposition reactions. Mobile elements are depicted by bold black double line, target DNA by dark grey double line, and DNA synthesised during transposition by light grey. Small arrows indicate DNA cleavage sites. See text for details. (Drawn according to Craig, 1995)

In conservative (also called non-replicative or cut-and-paste) transposition, a TE is excised from its previous location and moved into a new site. Even though the element is not duplicated in the process, TEs that employ conservative transposition are still capable of propagating in the genome by moving from replicated areas of the genome into unreplicated regions during cell division (Brookfield, 1995). In the conservative replication pathway the TE is cleaved at both its 3' and 5' ends, releasing the element that is then joined to the target DNA molecule (Craig, 1995, Haren *et al.*, 1999a). This is the most common mode of transposition, observed in a wide variety of elements including bacterial Tn*7* (reviewed in Craig, 2002), IS*10* (Kleckner *et al.*, 1996), and Tn*5* & IS*50* (Goryshin and Reznikoff, 1998), as well as eukaryotic DNA transposons of the *mariner*/Tc family (reviewed in Plasterk *et al.*, 1999) and the Drosophila P element (Engels *et al.*, 1990). In each case the transposase only cleaves one DNA strand, releasing the 3' ends of the element. Several different strategies have been adopted for the cleavage of the other strand including a passage through a hairpin intermediate, the use of two endonucleases, or transposition via a circular intermediate (reviewed in Turlan and Chandler, 2000).

In replicative transposition a copy of the TE remains in the original site while another clone of the transposon is copied into a new location. This type of transposition requires that the element is replicated in the transposition process. In replicative transposition the transposon is only cleaved at its 3' ends. If the transposition is intermolecular, replication of the element generates cointegrates in which donor and target replicons are joined but separated by a directly repeated copy of the TE at each junction. Resolution of this cointegrate by recombination between the two elements regenerates the donor and target molecules each carrying a copy of the TE (Craig, 1995, Haren *et al.*, 1999a). A few bacterial systems including bacteriophage Mu (Chaconas *et al.*, 1981), the IS*6* family (Mahillon and Chandler, 1998), and likely also the members of the Tn*3* transposon family (Craig, 2002) use replicative transposition. Bacterial IS*903* uses mostly the conservative transposition pathway but products of replicative transposition occur at a low frequency (<0.1 %). Apparently these are the result of a delay in the transposon 5' end cleavage, further implicating the similarity of conservative and replicative transposition pathways (Tavakoli and Derbyshire, 2001; also see Figure 2).

Some groups of DNA transposons use mechanisms fundamentally different from both conservative and replicative transposition. The bacterial IS*91* (Mendiola *et al.*, 1994) and eukaryotic *Helitrons* (Kapitonov and Jurka, 2001) employ a rolling-circle mechanism of transposition and resemble more prokaryotic plasmids than other transposable elements (Mendiola and de la Cruz, 1992, Garcillán-Barcia *et al.*, 2002).

2.2. Replicative transposition of bacteriophage Mu

The replicative transposition of bacteriophage Mu is well characterised on the biochemical level (see Mizuuchi, 1992, Chaconas *et al.*, 1996, Chaconas and Harshey, 2002 for reviews). The recombination mechanism used by Mu is similar to that of many other transposable elements (Craig, 1995). There are also mechanistic similarities to retroviral integration (Fujiwara and Mizuuchi, 1988, Mizuuchi, 1992, Craig, 1995, Andrake

and Skalka, 1996) and the early stages of V(D)J recombination (Craig, 1996, van Gent *et al.*, 1996, Roth and Craig, 1998).

2.2.1. Transposition requirements *in vivo*

A nucleoprotein complex that consists of specific DNA sites at the ends of the bacteriophage and a tetramer of viral-encoded transposase protein MuA is essential to the Mu transposition reaction. This nucleoprotein complex is called a transpososome (Surette *et al.*, 1987, Baker and Mizuuchi, 1992, Yuan *et al.*, 2005). Another Mu-encoded protein, MuB, is also directly involved in Mu transposition (Faelen *et al.*, 1978). In addition, several host-encoded factors assist the transposition reaction. In *E. coli*, the most important host accessory proteins are DNA binding proteins IHF (Integration Host Factor) (Surette and Chaconas, 1989) and histone-like protein HU (Craigie *et al.*, 1985). Moreover, molecular chaperone ClpX is essential in the DNA-transposase complex disassembly (Levchenko *et al.*, 1995, Kruklitis *et al.*, 1996).

Of the enzymatic cofactors, divalent cations are required in the assembly of the nucleoprotein complex and later in DNA cleavage and strand transfer reactions. Assembly of the nucleoprotein complex is possible using Ca^{2+}, Mg^{2+}, or Mn^{2+} but calcium does not support the subsequent DNA cleavage step. Mg^{2+} is likely the biologically relevant cation (Craigie and Mizuuchi, 1987, Surette *et al.*, 1987, Baker *et al.*, 1991, Mizuuchi *et al.*, 1992, Savilahti *et al.*, 1995). ATP is only needed for the stimulatory action of MuB protein (Craigie and Mizuuchi, 1987).

At the ends of the 37-kb Mu genome there are six binding sites for the specific DNA binding protein MuA, three at the left end (named L1-L3 inwards from the terminus) and three at the right end (R1-R3) (Figure 1D). These sites share a 22-bp consensus sequence (Craigie *et al.*, 1984). Sites L1, R1, and R2 are essential in the formation of the transpososome complex (Lavoie *et al.*, 1991), and L1 has a central role in the catalytic commitment of the transpososome (Kobryn *et al.*, 2002). Another enhancer site, internal activation sequence (IAS – see Figure 1D) is also essential for an efficient *in vivo* transposition (Mizuuchi and Mizuuchi, 1989). MuA binds Mu ends and the IAS with different subdomains (Leung *et al.*, 1989). In order to the transposition reaction to occur in normal cellular conditions, Mu ends and the IAS are required to be in the same molecule in a proper orientation relative to each other. Moreover, the Mu transposon DNA has to be negatively supercoiled (Craigie *et al.*, 1985).

The MuA transposase is a 663-amino acid (75 kDa) protein (Harshey *et al.*, 1985, Yuan *et al.*, 2005) divided into three major globular domains (Nakayama *et al.*, 1987). The N-terminal domain I binds transposon DNA: Iα interacts with the IAS enhancer whereas Iβ and Iγ recognise and bind transposon ends (Nakayama *et al.*, 1987, Leung *et al.*, 1989, Kim and Harshey, 1995). The central domain II is the catalytic core: IIα contains the DDE motif, a triad of catalytic acidic residues essential to DNA cleavage and strand transfer (Baker and Luo, 1994, Krementsova *et al.*, 1998) that is conserved in transposases of prokaryotic and eukaryotic mobile elements as well as retroviral integrases (Rice and Mizuuchi, 1995, reviewed in Haren *et al.*, 1999a, Chandler and Mahillon, 2002). The C-terminal domain III can be divided into two regions: domain IIIα is thought to act together with IIβ in the Mu-host junction interactions, assembly of the transpososome complex and structural transitions of that complex

(Krementsova *et al.*, 1998, Naigamwalla *et al.*, 1998, Namgoong *et al.*, 1998). Domain IIIβ interacts with MuB and CplX proteins (Baker *et al.*, 1991, Levchenko *et al.*, 1997).

2.2.2. Reaction intermediates

In the Mu transposition reaction the transpososome nucleoprotein complex evolves through several intermediate steps (Figure 3). Even though transposition is traditionally depicted as a linear pathway, several steps such as donor DNA cleavage, MuB target binding, and MuB-mediated target DNA entry can in reality occur in different orders or simultaneously, forming a network of reactions rather than an orderly linear pathway (Baker *et al.*, 1991, Naigamwalla and Chaconas, 1997, Yamauchi and Baker, 1998).

MuA binds the Mu ends initially as catalytically inactive monomers (Craigie *et al.*, 1984, Kuo *et al.*, 1991, Baker and Mizuuchi, 1992). At an early stage, an interaction of the Mu right end with IAS enhancer results in a two-site complex (enhancer - right end, ER) required for the HU-assisted capture of left end (Pathania *et al.*, 2003) forming the three-site synaptic complex (left end - enhancer - right end, LER) (Watson and Chaconas, 1996). MuA monomers are then transformed into the active MuA tetramer, forming the stable synaptic complex (SSC) (Mizuuchi *et al.*, 1992), also called type 0 transpososome (Chaconas *et al.*, 1996).

In the SSC, L1 and R1-bound MuA monomers catalyse the cleavage of 3' transposon ends *in trans*: reaction in one end of DNA is catalysed by the MuA monomer bound to the other end (Savilahti and Mizuuchi, 1996, Namgoong and Harshey, 1998, Williams *et al.*, 1999), forming the cleaved donor complex (CDC) (Craigie and Mizuuchi, 1987), also called type 1 transpososome (Surette *et al.*, 1987).

The next stage is the strand transfer reaction. MuB, an ATP-dependent DNA binding protein (Maxwell *et al.*, 1987) that forms large polymers on DNA (Greene and Mizuuchi, 2002a, Greene and Mizuuchi, 2002b), stimulates the strand transfer by recruiting a target molecule and delivering it to the transpososome (Baker *et al.*, 1991, Yamauchi and Baker, 1998). MuA cleaves the target DNA in a staggered manner, generating two target ends with 5 bp 5' overhangs. Free transposon 3'-OH groups are joined into these target 5' ends, converting CDC into the strand transfer complex (STC) or type 2 transpososome (Craigie and Mizuuchi, 1985, Surette *et al.*, 1987).

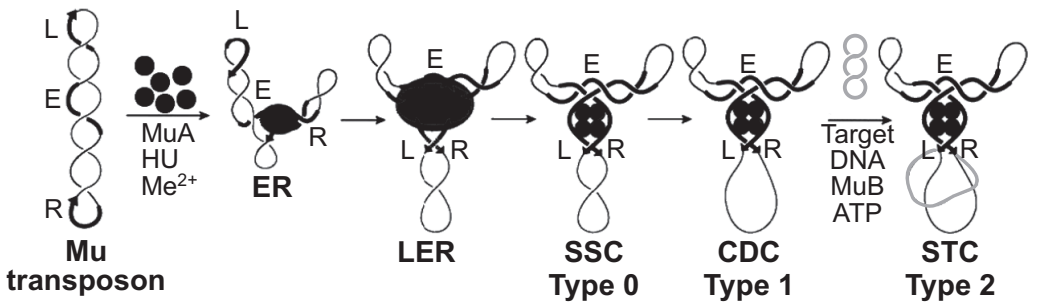

Figure 3. The replicative pathway of Mu transposition. L = Mu left end, R = Mu right end, E = IAS enhancer, Me^{2+} = divalent cation. MuA transposase monomers are depicted as black circles, target DNA as grey line. See text for details. (Drawn according to Pathania *et al*., 2003)

STC disassembly starts with the activity of ClpX protein that alters the conformation of DNA-bound MuA and converts the STC into STC2 (Kruklitis *et al.*, 1996) or type 3 transpososome (Chaconas and Harshey, 2002). MuA, still present in this complex, is removed through the action of other host proteins (Kruklitis *et al.*, 1996). Host-encoded factors then initiate DNA replication, carried out by the replication machinery of the cell, including DnaB helicase, DnaC, and polymerase III holoenzyme (Kruklitis and Nakai, 1994, Nakai and Kruklitis, 1995). The host replication machinery also repairs the single-stranded gaps produced by transposition. This process creates a 5-bp target site duplication, a hallmark of the Mu transposition (Allet, 1979).

Much less is known about the conservative transposition of Mu into the host genome following infection. Apparently the mechanism of transpososome assembly is different from replicative transposition, with less emphasis on DNA topology (Sokolsky and Baker, 2003).

2.2.3. Control of transposition

Uncontrolled proliferation of any transposable element would be deleterious to the host cell and thus very likely to the element itself as well. Moreover, an intramolecular transposition would most likely render the element inactive, thus destroying it. Therefore all TEs have to be able to restrict and control their transposition (Plasterk, 1995). Bacteriophage Mu codes for a repressor protein that negatively regulates early transcription including transposase genes. In addition, the repressor also directly inhibits transposition by binding into Mu operator sequences that overlap the IAS (Craigie *et al.*, 1984), thus preventing the transpososome formation (Mizuuchi and Mizuuchi, 1989).

The requirements for target DNA are almost minimal in the Mu transposition reaction. Even though the target site selection is non-random (Castilho and Casadaban, 1991, Manna *et al.*, 2001), there is no consensus sequence. Rather, some sites are statistically preferred (5' C-Py-G/C-Pu-G 3') (Mizuuchi and Mizuuchi, 1993, Butterfield *et al.*, 2002, Haapa-Paananen *et al.*, 2002). In regional level there is a preference associated with DNA sequences that have high affinity to MuB protein, which demonstrates a weak but recognisable binding site preference (Mizuuchi and Mizuuchi, 1993, Greene and Mizuuchi, 2002a). Finally, non-Mu end sequences are strongly preferred as targets through a phenomenon called target (or transposon) immunity, in which one Mu sequence exerts an inhibitory effect on the probability of a second Mu inserting in nearby sequence (Adzuma and Mizuuchi, 1988, Manna and Higgins, 1999). This is caused by MuB protein that binds DNA in the presence of ATP. MuA that specifically binds Mu end DNA can catalyse the dissociation of MuB from DNA in a process that requires ATP hydrolysis (Adzuma and Mizuuchi, 1988, Adzuma and Mizuuchi, 1989, Greene and Mizuuchi, 2002c). Contrastingly, in the transpososome complex, MuB-bound DNA is the preferred transposition target because MuB stimulates the recombinase activities of MuA (Baker *et al.*, 1991, Yamauchi and Baker, 1998). Target immunity confers a gradually decaying immunity up to 25 kb from the transposon ends. This phenomenon probably evolved as a mechanism to prevent the bacteriophage from intramolecular transposition into itself (Adzuma and Mizuuchi, 1988, Adzuma and Mizuuchi, 1989, Manna and Higgins, 1999).

2.3. Mu transposition reactions *in vitro*

Bacteriophage Mu was the first transposon for which a cell-free *in vitro* transposition reaction was developed. The original reaction used conditions suitable for *in vitro* DNA replication, a supercoiled donor plasmid carrying Mu end DNAs in proper orientation and *E. coli* cell extracts that contained MuA and MuB proteins (Mizuuchi, 1983). After that initial experiment Mu transposition has been studied extensively and transposition requirements for an *in vitro* reaction have been defined and minimised.

The topological constraints on transposon donor can be relaxed by adding dimethyl sulfoxide (DMSO) to the reaction, enabling the use of linear donors (Craigie and Mizuuchi, 1986). Moreover, DMSO makes IAS enhancer and HU protein obsolete. IHF, normally involved in IAS binding, is not required in these conditions (Mizuuchi and Mizuuchi, 1989).

The initial stages of transposition until CDC can be bypassed altogether by using pre-cut donor molecules. In this type of reaction the innermost MuA binding sites (L3 and R3) are not required; in fact, the most efficient transposition is achieved using donors with two minimal right end (R1 & R2) sequences in inverted orientation (Craigie and Mizuuchi, 1987, Namgoong *et al.*, 1994). In reduced salt concentrations glycerol can be used instead of DMSO to stimulate strand transfer reactions (Mizuuchi and Mizuuchi, 1989).

In relaxed *in vitro* conditions the MuB protein is not necessary for the reaction (Baker *et al.*, 1991, Savilahti *et al.*, 1995), even though MuB has been shown to stimulate the cleavage and strand transfer reactions also in altered (DMSO, glycerol) conditions (Lee and Harshey, 2001, Lee and Harshey, 2003). Without MuB, the reaction loses its target immunity (Baker *et al.*, 1991).

Taken together, the MuA transposase, pre-cleaved donor DNA with (R1,R2) Mu ends, and target DNA are the only macromolecular components required for the minimal Mu *in vitro* transposition reaction (Savilahti *et al.*, 1995, Haapa *et al.*, 1999a). It is possible to insert marker sequences of variable length between the two Mu ends (Haapa *et al.*, 1999a). In addition, the Mu ends themselves can also be modified moderately (Taira *et al.*, 1999, Laurent *et al.*, 2000, Goldhaber-Gordon *et al.*, 2002a, Goldhaber-Gordon *et al.*, 2002b, Poussu *et al.*, 2005). Minimal reaction conditions have been used in a variety of Mu *in vitro* applications (see chapter 3.4).

3. Transposons as tools in molecular biology

The ability of transposons to move discrete segments of DNA to new locations where they disrupt target DNA, effectively combining the activities of a restriction endonuclease and a ligase, make them valuable tools in molecular biology (Berg and Berg, 1995). Apart from obvious loss-of-function mutations associated with transposon insertion, TEs can also inflict chromosomal rearrangements – deletions, duplications, inversions – by homologous recombination between two repeats. These rearrangements can be very complex, especially if several intertwined elements, possibly located in different DNA molecules, are involved (for a review see Gray, 2000).

The actual transposition reaction can occur either *in vivo*, within the cells, or *in vitro*, i.e. in the test tube. A wide range of different TEs and transposases from various organisms has been used

in vivo and for an increasing number of transposons an efficient *in vitro* reaction has also been developed (reviewed in Boeke, 2002).

3.1. *In vivo* vs. *in vitro*

In the first generation of transposition applications the reactions were carried out *in vivo*, within the cells. *In vivo* transposition requires either the use of an endogenous transposon residing in the host genome or that the main reaction components, the transposon and the transposase protein, are introduced into the host cell. Typically, the transposon is in a suicide vector that does not proliferate in the host cell (for example Gonzales *et al.*, 1996, Lee and Henk, 1996). The transposase is usually expressed in subsequent generations resulting in potential genetic instability (Kleckner, 1990, Goryshin *et al.*, 2000). Moreover, the use of heterologous transposons derived from another organism is not always possible due to limitations in the host range of transposons. To circumvent this problem, *E. coli* can be used as a surrogate host where transposition is carried out and the resultant products are subsequently transferred back to the original host (reviewed in Hamer *et al.*, 2001). Despite its limitations *in vivo* transposition is a practical mutagenesis method that is frequently used, for example, in the functional genomics of bacteria (recent examples include Banh *et al.*, 2005, McCarren and Brahamsha, 2005, Glass *et al.*, 2006).

Performing transposition reactions *in vitro* in a cell-free system offers several advantages over traditional *in vivo* systems. Host-range limitations can be avoided if transposition reactions are carried out *in vitro* and the reaction products transformed into host cells. The reactions can be carried out using fewer components, allowing better control over reactions and a wider range of reaction conditions. Thus, *in vitro* reactions are generally much more efficient (Goryshin and Reznikoff, 1998, Haapa *et al.*, 1999a, Biery *et al.*, 2000). If components that are critical for a specific step of the process are left out, the reaction can be arrested at a suitable stage to study (for example Craigie and Mizuuchi, 1985, Mizuuchi *et al.*, 1992) or otherwise exploit reaction intermediates (Goryshin *et al.*, 2000, Lamberg *et al.*, 2002). Moreover, transposable elements often display target site preferences *in vivo*; transposon integration can be markedly more random using fewer reaction components *in vitro* (Boeke and Devine, 1998, Biery *et al.*, 2000, Boeke, 2002). Finally, *in vitro* reactions also allow the subsequent manipulation of transposons and reaction products (for example Taira *et al.*, 1999, Laurent *et al.*, 2000, Kekarainen *et al.*, 2002, Poussu *et al.*, 2004, Poussu *et al.*, 2005). Simple *in vitro* transposition systems have been developed for several transposons including bacteriophage Mu (Haapa *et al.*, 1999a), Tn*5* (Goryshin and Reznikoff, 1998), Tn*10* (Chalmers and Kleckner, 1994), Tn*552* (Griffin *et al.*, 1999), Tn*7* (Biery *et al.*, 2000), *mariner* (Tosi and Beverley, 2000), and Ty*1* (Devine and Boeke, 1994).

One disadvantage of *in vitro* transposition is that transposition does not occur in the natural environment of genes. In order to have biological selection, *in vitro* transposition products have to be delivered into the host cells. Generally, plasmids or small viral genomes mutagenised can be transferred into cells, even though some transposition products can be lost in transfer. However, the mutagenesis of chromosomal DNA this way is cumbersome. The region studied has to be cloned in a vector, limiting its size, and even

then replacing the original DNA segment with a mutagenised allele usually relies on comparatively inefficient homological recombination and subsequent screening for mutants. Preparing transpososome complexes *in vitro*, transferring them to target cells and letting transposition occur *in vivo* circumvents these problems (Goryshin *et al.*, 2000, Lamberg *et al.*, 2002). Even though not as controllable and efficient as *in vitro* reactions, the use of preassembled transpososomes enables transposition to occur *in vivo* with no need to transfer the reaction products. Unlike traditional *in vivo* methods, transposition occurs at a fixed time point and there is no risk of remobilisation (Goryshin *et al.*, 2000, Lamberg *et al.*, 2002). Preassembled transpososomes have been used with Tn*5* (Goryshin *et al.*, 2000, Reznikoff *et al.*, 2004) and bacteriophage Mu systems (Lamberg *et al.*, 2002, Pajunen *et al.*, 2005).

3.2. Applications of transposition technology

Transposons have been exploited as research tools since the late 1970s, almost as soon as their true nature became evident. Bacteriophage Mu was used to induce homologous recombination between two regions containing inserted elements (Casadaban, 1975) and in functional gene analysis using a derivative transposon containing a promoterless reporter gene (Casadaban and Cohen, 1979). The study of eukaryotes with transposable elements started with the P element in *Drosophila* (Spradling and Rubin, 1982). Methods based on *in vivo* transposition of several transposons and transposon derivatives have been used in identification, mapping, regulation and sequencing of genes in prokaryotes and eukaryotes (reviewed in Berg and Berg, 1995, Kaiser *et al.*, 1995).

3.3. Functional genomics applications

Functional genetics studies the function of genes, their parts, and their products (RNA, proteins) trying to elucidate their function and relationships behind these functions. Functional genomics expands this study to larger regions of DNA, even to whole genomes. Traditionally, these functional studies are conducted by inducing mutations that disrupt gene function.

The completion of extensive whole-genome sequencing projects, conducted on prokaryotes, eukaryotes and viruses at an increasing rate onwards from mid-1990s, has yielded a massive amount of sequence data. However, the functions of most putative genes and other genetic elements remain elusive. Transposable elements, with their ability to insert into a wide range of locations on a DNA molecule, are a powerful mutagenising agent that has been widely used to analyse the functions of genes and their products (reviewed in Berg and Berg, 1995, Kaiser *et al.*, 1995, Hayes and Hallet, 2000, Hamer *et al.*, 2001, Hayes, 2003), as well as entire genomes (for example Hutchison *et al.*, 1999, Wong and Mekalanos, 2000, Hare *et al.*, 2001, Gerdes *et al.*, 2002, Gerdes *et al.*, 2003, Banh *et al.*, 2005, McCarren and Brahamsha, 2005, Glass *et al.*, 2006).

3.3.1. Individual genes

Simple insertional transposon mutagenesis is an effective way to disrupt target gene(s) with both forward and reverse genetics approaches. Transposons have been engineered to contain specific regulatory and reporter sequences to study the function of target genes and proteins they encode. Transposition can generate random gene fusions between gene(s) of interest and a reporter gene in the transposon to facilitate

the analysis of transcription, translation, and cellular location of gene products. In addition, transposons with a marker gene but lacking an enhancer, a promoter or a polyadenylation signal are used to probe for those regulatory elements (reviewed in Berg and Berg, 1995, Kaiser *et al.*, 1995, Boeke, 2002, Hayes, 2003).

In targeted mutagenesis of specific genes the target gene is mutagenised, isolated and subsequently transferred back into host cell where it replaces the wild-type allele by homologous recombination. This mutagenesis step can be carried out with transposition to, for example, generate gene-targeting vectors for mouse (Westphal and Leder, 1997).

Transposition has also been used to generate different mutant libraries to study functional domains of proteins and protein-protein interactions. Deletion variants of proteins have been produced by intramolecular (Ahmed, 1984, Tomcsanyi *et al.*, 1990, Morita *et al.*, 1996, York *et al.*, 1998) or intermolecular transposition (Poussu *et al.*, 2005). In scanning linker mutagenesis, the insertion and imprecise excision of a transposon leaves only a short, in-frame insertion in the target site. This produces a library of mutations throughout the coding region of the gene studied, causing a short random insertion in the translated protein. By sampling a large amount of such mutants, functional or interacting domains of the protein studied can be defined. A variety of *in vivo* and *in vitro* scanning linker mutagenesis systems have been developed employing different transposons (Hallet *et al.*, 1997, Hayes and Hallet, 2000, Poussu *et al.*, 2004).

3.3.2. Whole-genome level

Different transposon-based methods have been devised to study the essential genes needed for growth of an organism, including genetic footprinting (Smith *et al.*, 1995, Akerley *et al.*, 1998, Wong and Mekalanos, 2000), signature-tagged mutagenesis (Hensel *et al.*, 1995), and microarray-based techniques (Badarinarayana *et al.*, 2001, Sassetti *et al.*, 2001). These methods involve simultaneous generation of large number of insertion mutants either *in vivo* or *in vitro*, growing mutagenised organism under specific conditions and screening for mutations that affected survival. These global screens have been conducted in various bacteria and few eukaryotic model organisms (reviewed in Judson and Mekalanos, 2000, Hamer *et al.*, 2001, Hayes, 2003). Transposon-based functional genomics of viruses are discussed in chapter 3.3.2.5.

3.3.2.1. Genetic footprinting

Genetic footprinting is a transposon-based genomic strategy for determining the functions of sequenced genes (Figure 4). This method, originally devised *in vivo* in yeast with Ty*1* retrotransposon (Smith *et al.*, 1995), has also been applied to cloned genes *in vitro* with transposons or retroviral integrases (Singh *et al.*, 1997, Haapa *et al.*, 1999a, Rothenberg *et al.*, 2001, Auerbach *et al.*, 2003).

The use of genetic footprinting has been extended to identify candidate essential or important genes of bacteria by utilising *in vitro* mutagenesis of cloned / purified genomic DNA or PCR products followed by transformation into cells and homologous recombination into host genomes of naturally competent bacteria and subsequent growth selection under different conditions. This method is also called genomic analysis and mapping with *in vitro* transposition (GAMBIT) (Akerley *et al.*, 1998).

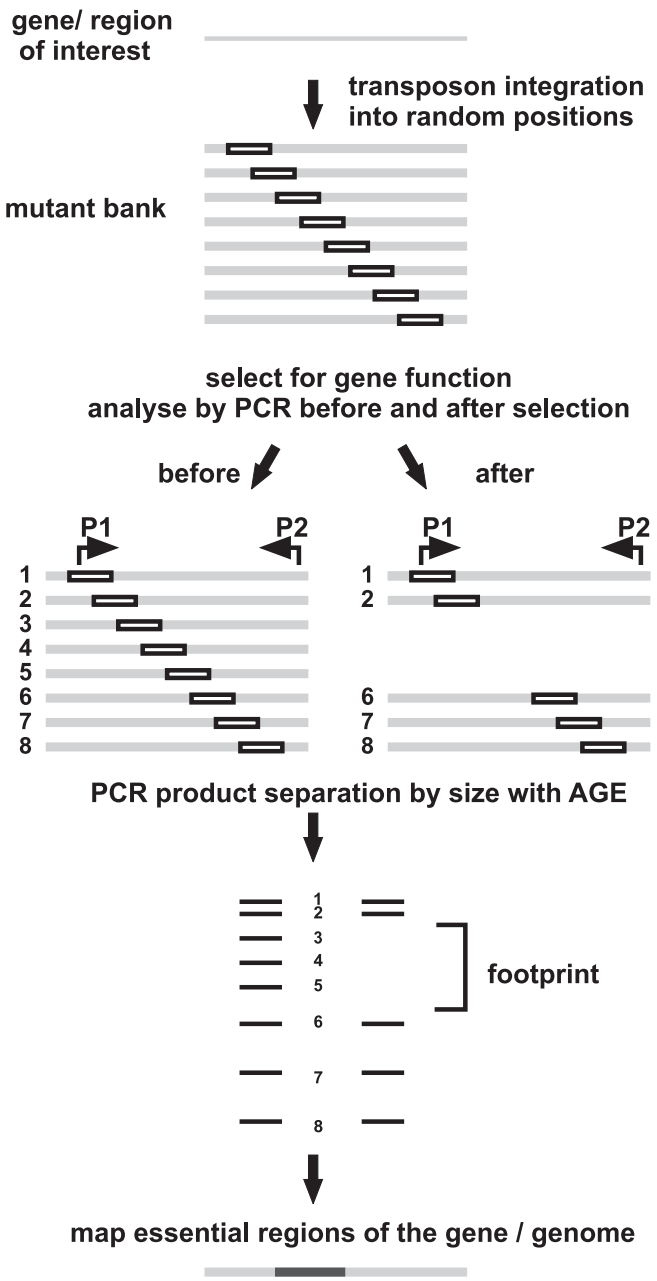

Figure 4. Genetic footprinting. A large number of clones are mutagenised with transposons to generate a mutant bank that is subsequently subjected to selection. Polymerase chain reaction (PCR) with one transposon-specific (P1) and one target-specific (P2) primer is used to determine the effects of the mutations under selection. PCR bands corresponding to clones that are not viable under selection are excluded from the selected pool, forming a footprint. AGE = agarose gel electrophoresis. (Drawn according to Singh et al., 1997)

Genetic footprinting allows the use of large mutant pools but only a relatively small number of genes can be screened in each PCR, making global whole-genome analysis laborious (Sassetti *et al.*, 2001). Variations of this theme use DNA hybridisation and/or direct sequencing to determine transposon integration sites (Akerley *et al.*, 1998, Reich *et al.*, 1999).

The whole-genome analyses of bacteria have been carried out with *in vivo* transposon mutagenesis and subsequent analysis of transposon integration sites in viable clones in several bacteria including *Mycoplasma genitalium* (Hutchison *et al.*, 1999, Glass *et al.*, 2006), *M. pneumoniae* (Hutchison *et al.*, 1999), *Pseudomonas aeruginosa* (Wong and Mekalanos, 2000), *Helicobacter pylori* (Jenks *et al.*, 2001), and *E. coli* (Badarinarayana *et al.*, 2001, Hare *et al.*, 2001, Gerdes *et al.*, 2002, Gerdes *et al.*, 2003).

3.3.2.2. Signature-tagged mutagenesis (STM)

Signature-tagged mutagenesis (STM) is a transposon-based method that uses negative selection to simultaneously screen multiple mutants affecting virulence of pathogenic organisms (Figure 5). For STM the pathogen in question has to be able to infect its host as a mixed population. In addition, only those mutant genes that cannot be trans-complemented by other virulent strains are expected to be identified (Chiang *et al.*, 1999). Signature-tagged mutagenesis has been used with various bacterial and fungal pathogens (reviewed in Chiang *et al.*, 1999, Hayes, 2003). While particularly useful for the study of pathogenesis, STM is limited by the relatively small size of each mutant pool and comprehensive screens are laborious (Sassetti *et al.*, 2001).

3.3.2.3. Microarray-based methods

Microarray technology allows convenient monitoring of genome-wide gene expression patterns (Schena *et al.*, 1995, Schena *et al.*, 1996). Microarrays have been combined with DNA hybridisation of transposon sites to study essential genes in *Mycobacterium* (Sassetti *et al.*, 2001, Sassetti *et al.*, 2003). This method is especially useful for identifying the set of genes required for survival in different conditions (Sassetti *et al.*, 2001). However, the ability to define growth requirements for individual genes are limited (Sassetti *et al.*, 2003). Microarray technology has also been used with genetic footprinting to study essential genes in *E. coli* under defined growth conditions (Badarinarayana *et al.*, 2001).

3.3.2.4. Functional studies in eukaryotes

Because of their smaller genomes, easier handling, and abundance of functional transposon systems, most genome-wide mutagenesis experiments have been conducted in various bacteria (reviewed in Judson and Mekalanos, 2000, Hamer *et al.*, 2001, Hayes, 2003). In eukaryotes, large genome size and the lack of endogenous active transposon systems have hindered the use of transposable elements in the functional studies. Budding yeast (*S. cerevisiae*) has been the subject of functional analysis with several methods including genetic footprinting of chromosomal regions with Ty*1* (Smith *et al.*, 1996) and shuttle mutagenesis in *E. coli* with bacterial Tn*3* (Ross-Macdonald *et al.*, 1999). *In vivo* transposition of preassembled Tn*5* transpososome complexes into budding yeast has also been reported (Goryshin *et al.*, 2000).

Genome-wide insertion libraries have also been generated with endogenous P element and heterologous *piggyBac* in *Drosophila* (Spradling *et al.*, 1999, Ryder and Russell, 2003, Bellen *et al.*, 2004), Tc1/*mariner* family in *C. elegans* (Plasterk *et al.*, 1999), and transposons belonging to hAT and CACTA superfamilies in *Arabidopsis*, maize, and other plants (Martienssen, 1998, Parinov and Sundaresan, 2000, Walbot, 2000). In vertebrates *Sleeping Beauty*, a Tc1/*mariner* family DNA transposon resurrected by comparing inactive elements in teleost fish (Ivics *et al.*, 1997) has been the most widely used TE in functional genomics studies (reviewed in Izsvak and Ivics, 2004, An and Boeke, 2005, Miskey *et al.*, 2005).

3.3.2.5. Viral functional genomics

Virus genomes have traditionally been studied by inducing conditional mutations gene-by-gene, a tedious process that requires complementing cell lines and homologous recombination followed by several rounds of selection. The use of transposon mutagenesis in viral functional genetics has long been hampered by the lack of suitable host systems. However, advances in transposon technology and development of whole-genome analysis methods also applicable with viruses have facilitated the study of transposon-based viral functional genomics, examples of which are summarised in Table 2.

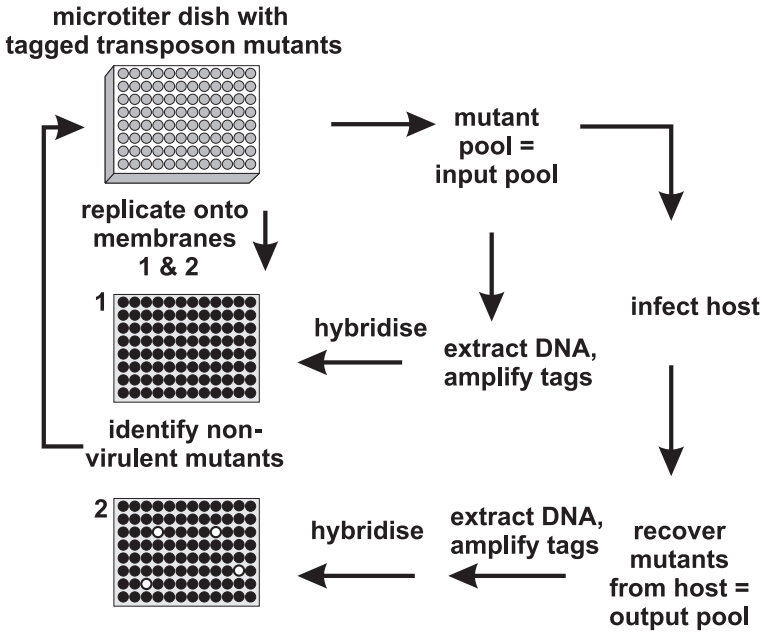

Figure 5. Signature-tagged mutagenesis (STM). Bacteria (or viruses) are mutagenised with transposons that are tagged with a short unique DNA sequence that permits the identification of individual mutants. Mutagenised bacteria are then combined into an input pool that is used to infect host. After infection an output pool comprising mutants capable of proliferating in host is collected. Mutants with attenuated virulence are the ones detected in input pool but absent in output pool. (Adapted from Hensel et al., 1995)

Table 2. Transposon-based functional virus genomics studies

virus name	abbreviation	genome (kb)	mutants described	transposon	*in vivo / in vitro*	method	reference
herpes simplex virus type 1	HSV-1	150	3	Tn*5*	*vivo*	short region cloned in a plasmid, insertional	Weber *et al.*, 1987
murine cytomegalo-virus	MCMV	230	8	Tn*1721*	*vivo*	complete BAC, insertional	Brune *et al.*, 1999
pseudorabies virus	PRV	142	23	Tn*5*	*vivo*	complete BAC, insertional	Smith and Enquist, 1999
murine cytomegalo-virus	MCMV	230	3	Tn*3*	*vivo*	plasmid library, insertional	Zhan *et al.*, 2000b
human cytomegalo-virus AD 169	HCMV AD169	230	26	Tn*1721*	*vivo*	complete BAC, insertional	Hobom *et al.*, 2000
human immuno-deficiency virus type 1	HIV-1	9.2	-	Mu ends	*vitro*	cloned segment, footprinting	Laurent *et al.*, 2000
potato virus A	PVA	9.6	1125	Mu ends	*vitro*	cloned BAC, footprinting	Kekarainen *et al.*, 2002
bacteriophage PRD1	PRD1	14.9	130	mini-Mu	*vitro*	isolated genome, insertional	II
human cytomegalo-virus AD169	HCMV AD169	230	413	Tn*1721*	*vivo*	cloned BAC footprinting	Yu *et al.*, 2003
γ-herpesvirus 68	γHV68	118	53	Tn*5*	*vivo*	Cloned BAC, STM	Moorman *et al.*, 2004
bacteriophage ΦYeO3-12	ΦYeO3-12	39.6	18	mini-Mu	*vitro*	isolated genome, insertional	III
γ-herpesvirus 68	γHV68	118	1152	mini-Mu	*vitro*	Cloned BAC, STM	Song *et al.*, 2005
bacteriophage PM2	PM2	10.1	101	mini-Mu	*vitro*	isolated genome, insertional	IV

Optimally, both transposition and viral replication machineries would be functional within the same host cell. Classic reports of transposon Tn*3* and Tn*10* insertions into genomes of bacteriophage M13 (Ray and Kook, 1978) and λ (Kleckner *et al.*, 1978), respectively, fill these criteria. If no such system exists, the viral genome or parts thereof could be cloned on a vector capable of replicating in *E. coli* and subsequent *in vivo* transpositional mutagenesis conducted with one of the well-characterised bacterial transposons. Parts of herpes simplex virus type 1 (HSV-1) cloned in a plasmid were mutagenised with bacteriophage Mu and transferred into viral genome by homologous recombination (Jenkins *et al.*, 1985). Similar technique was used with Tn*5* to conduct an assessment of nonessential genes in a short region of HSV-1 genome (Weber *et al.*, 1987). Low efficiency of transposition, multiple cloning steps required, and a laborious screening process limit the applicability of this method. Nevertheless, Tn*3* mutagenesis of randomly digested genomic fragments of murine cytomegalovirus permitted the analysis in viral genomic level and facilitated concomitant studies on several gene loci, (Lee *et al.*, 2000, Xiao *et al.*, 2000, Zhan *et al.*, 2000a, Zhan *et al.*, 2000b, Abenes *et al.*, 2004).

The cloning of virus genomes in an infective form in a plasmid, cosmid, or bacterial artificial chromosome (BAC) that is able to replicate in *E. coli* has enabled *in vivo* transposon mutagenesis of entire viral genomes. Such strategies, employing different transposons, have been used to identify genes essential for virus propagation in several cytomegaloviruses (Brune *et al.*, 1999, Smith and Enquist, 1999, Hobom *et al.*, 2000, Yu *et al.*, 2003). Moreover, the generation of large mutant banks allow whole-genome analyses using methods like genetic footprinting and signature-tagged mutagenesis.

Besides the analysis of viral proteins (Rothenberg *et al.*, 2001, Auerbach *et al.*, 2003), genetic footprinting has been used in viral genome-level studies. Parts and entire cloned viral genomes have been mutagenised *in vitro* with transposons to study genes essential to viral life cycles. A genomic segment of human immunodeficiency virus 1 (HIV) (Laurent *et al.*, 2000) and the whole genome of Potato virus A (PVA) (Kekarainen *et al.*, 2002) were mutagenised *in vitro* with short bacteriophage Mu ends, transferred to host for propagation and analysed with genetic footprinting.

Signature-tagged mutagenesis, originally developed for the identification of bacterial virulence genes, has also been used *in vivo* (Moorman *et al.*, 2004) and *in vitro* (Song *et al.*, 2005) to simultaneously identify multiple genes essential for replication of murine gammaherpesvirus 68.

3.4. Applications of Mu *in vitro* transposition

The minimal Mu *in vitro* reaction allows the transposition of almost any marker element(s) surrounded by Mu end DNA sequences into whatever target DNA deemed convenient enough to handle after insertional mutagenesis (Savilahti *et al.*, 1995, Haapa *et al.*, 1999a). Mu *in vitro* transposition reaction is highly efficient and has relatively low target site selectivity (Mizuuchi and Mizuuchi, 1993, Haapa *et al.*, 1999a, Butterfield *et al.*, 2002, Haapa-Paananen *et al.*, 2002). In addition, it is possible to have an efficient reaction with Mu ends that are modified to contain restriction enzyme cutting sites or translational stop codons (Laurent *et al.*, 2000, article II, Poussu *et al.*, 2005) further

broadening the scope of target sequence manipulation. These properties make Mu *in vitro* reaction useful for generating comprehensive libraries of mutant DNA molecules that have been utilised in a variety of molecular biology applications.

DNA sequencing was the first Mu *in vitro* application developed. This simple method is based on random transposon insertions throughout the target and subsequent sequencing to both directions using a pair of transposon-specific primers (Haapa *et al.*, 1999b). Because its efficiency and near-randomness of integration, this method has also been applied to high-throughput sequencing of cDNA clones using pools of different targets (Butterfield *et al.*, 2002).

Mu *in vitro* transposition has been used to study functional domains of proteins and protein-protein interactions with scanning linker mutagenesis. A transposon with modified Mu ends allows the bulk of the transposon to be removed after random transposon integration, leaving only a 15-bp insert. By sampling a large amount of such mutants, functional or interacting domains of the protein studied can be defined (Taira *et al.*, 1999, Poussu *et al.*, 2004). Similarly, a nested set of N-terminal deletions can be prepared if, in addition to the restriction sites in transposon ends, similar restriction site is situated in the vector at the start of the coding region. A random transposon integration and subsequent removal of the transposon and the intervening 5' part of the coding region produces a set of mutations with N-terminal deletions of different sizes. Generating a nested set of C-terminal mutations is even simpler: another modified-end transposon, with stop codons in all three reading frames, is randomly inserted in the coding region, producing a set of mutations with C-terminal deletions of different sizes. N- and C-terminal deletion libraries can be used to study functions of protein domains (Poussu *et al.*, 2005).

The function of genes, larger regions of genomic DNA/RNA, and even entire genomes can also be studied using Mu *in vitro* transposition. After random insertion of transposons the target DNA is transferred into host cells where it can propagate and then be subjected to selection. The principle of this genetic analysis was tested using plasmid pBC SK+. After the plasmid was insertionally mutagenised it was transformed into *E. coli* host cells and subjected to plasmid replication. The mutants having an insertion in the plasmid origin of replication, present in the unselected mutant pool, were absent in the pool of mutants that underwent selection (Haapa *et al.*, 1999a).

Mu *in vitro* transposition can also be used to determine whether a viral genomic region is *cis*-acting or can be complemented *in trans* from another molecule. A 1000-nucleotide 5' end segment of human immunodeficiency virus (HIV1) RNA genome was cloned as DNA and subjected to insertional mutagenesis. After the removal of transposon ends a library of 15-bp insertion mutants was generated. This mutation library was transfected into producer cells and the survival of cotransfected mutant virions having a defective 5' end was monitored through different stages of the viral life cycle. Regions with cis-acting functions in a specific stage of the life cycle do not have transposon insertions in viruses that underwent that stage (Laurent *et al.*, 2000).

It is even possible to perform a functional analysis of an entire viral genome with Mu *in vitro* transposition. The whole RNA genome of Potato virus A (PVA), cloned into a dsDNA vector, was subjected to transpositional mutagenesis to generate a library of 15-bp insertion

mutants. These dsDNA mutants were transfected into tobacco protoplasts, where they are capable of initiating infection. After two days of virus propagation the viruses were collected and the transposon insertion sites analysed. Genomic regions that did not tolerate transposon insertions were deemed essential (Kekarainen *et al.*, 2002).

The use of Mu *in vitro* transposition in functional genetics of mouse and whole-genome analysis of bacteriophages, the subject of this thesis, is described in sections C through E.

4. Examples of model organisms for functional genetics and genomics

4.1. Mouse

Mouse (*Mus musculus*) has historically been the most widely used model organism in mammal genetics and cell biology, primarily due to its physiological similarity to man, relatively short generation time, and prolific offspring production. Even though the 2.6 Gb mouse genome is 14 % smaller than its human counterpart, the number of protein-coding genes appears to be similar (~30000) and approximately 80 % of mouse genes have a single identifiable human ortholog. The proportion of mouse genes that have no detectable human homologue appears to be less than 1 % (Waterston *et al.*, 2002, updated in Gibbs *et al.*, 2004).

The development of powerful genome manipulation techniques has further emphasised the role of mouse as a model organism. Mouse embryonic stem cells (ES cells) offer an efficient means for mouse genomic alteration, especially gene targeting, systematic alteration of genome by homologous recombination (Thomas and Capecchi, 1987). Mouse ES cells can retain their pluripotency while cultured *in vitro* (Evans and Kaufman, 1981). If these cells are reintroduced into mouse blastocyst they can contribute to all the cell lines of the resulting chimeras, including the germ line (Bradley *et al.*, 1984, Thompson *et al.*, 1989). Mice that are homozygous for the altered allele are obtained by cross-breeding the offspring of these chimeras. It is also possible to completely derive a mouse from manipulated, cultured ES cells by using developmentally compromised tetraploid embryos (Nagy *et al.*, 1993). Today, transgenic mice models are pivotal in the study of mammalian biology and human disease (for reviews see Rossant and McKerlie, 2001, Sands, 2003).

4.2. Bacteriophages

4.2.1. PRD1

PRD1 belongs to *Tectiviridae*, a group of icosahedral bacteriophages containing an internal membrane. It is a lytic, broad-host-range virus that infects a variety of Gram-negative host bacteria harbouring a conjugative N-, P-, or W-type plasmid, including *Escherichia coli* and *Salmonella typhimurium* (Olsen *et al.*, 1974). The PRD1 genome is a linear, 14927 bp long double-stranded (ds) DNA molecule with covalently linked terminal proteins at the 5' ends of genomic DNA (Bamford *et al.*, 1983, Bamford *et al.*, 1991, Saren *et al.*, 2005). The organization of the PRD1 genome at the transcriptional level is relatively well known; there are 49 open reading frames (ORFs) divided into two early (OE1-2) and three late (OL1-3) operons (Grahn *et al.*, 1994, Bamford *et al.*, 2002). Because of its physical characteristics and broad host-range PRD1 has been a model organism for biological membranes, genomic organisation, and recently also structural studies.

4.2.2. ΦYeO3-12

ΦYeO3-12 belongs to *Podoviridae*, a group of icosahedral, short-tailed dsDNA bacteriophages (Ackermann *et al.*, 1997). It is a lytic, narrow-host-range virus that infects only *Yersinia enterocolitica* serotype O:3 strains (al-Hendy *et al.*, 1991); yet, it is closely related to *Escherichia coli* bacteriophages T3 and T7 (Pajunen *et al.*, 2000). The genome of ΦYeO3-12 is a linear, 39600 bp long double-stranded DNA molecule that harbours 54 putative genes, all transcribed from the same DNA strand and divided to early, middle, and late regions (Pajunen *et al.*, 2001). *Y. enterocolitica*, a Gram-negative bacterium that belongs to the family *Enterobacteriaceae*, is a major source of food-borne diseases. The virus-specific serotype O:3 is, along with O:9, one of the two most common pathogenic serotypes in Europe, Canada, Japan, and South Africa (Bottone, 1997). ΦYeO3-12 was isolated from the city sewage of Turku, Finland in the late 1980s (al-Hendy *et al.*, 1991). O:3 serotype specificity makes it a potential biotechnological tool; thus, it was the first yersiniophage characterised at molecular level (Pajunen *et al.*, 2001).

4.2.3. PM2

PM2 is the only characterised member of *Corticoviridae*, containing an icosahedral particle, an internal lipid bilayer, and a highly supercoiled circular dsDNA genome. It was the first detected lipid-containing virus when it was isolated off the coast of Chile in 1968 (Espejo and Canelo, 1968b). It is a narrow-host-range lytic bacteriophage that infects Gram-negative marine bacterium *Pseudoalteromonas espejiana* strain BAL-31 (Espejo and Canelo, 1968a, Chan *et al.*, 1978, Gauthier *et al.*, 1995) and *Pseudoalteromonas* sp. ER72M2 (Kivelä *et al.*, 1999). The genome of PM2 is a circular, 10097 bp long dsDNA molecule harbouring 21 putative genes (Männistö *et al.*, 1999) that are organised into three operons: two early (OEL and OER) and one late (OL)(Männistö *et al.*, 2003). The PM2 genome has the highest reported number of negative supercoils in a natural DNA molecule (Gray *et al.*, 1971). Because of its internal membrane PM2 has been the subject of studies on the membrane structure and biosynthesis. Additionally, its small, circular, and highly supercoiled genome has been of interest in the DNA topology studies.

C. AIMS OF THE STUDY

The aim of this project was to study the possibilities to expand the use of Mu *in vitro* transposition in functional genetics and genomics.

1. To prove that this methodology can be used in the functional genetics of higher eukaryotes by generating a rapid method for the construction of mouse gene targeting vectors essential in the production of transgenic mice.

2. To design a method for the whole-genome functional analysis of bacteriophages that have infective genomes and provide means for subsequent modification of those genomes.

3. To demonstrate that the whole-genome analysis method developed is applicable to bacteriophages infecting different hosts and can also be used to modify viruses that have been difficult to modify using traditional methods.

4. To further refine the method developed and to develop an efficient mutant selection scheme that is not dependent on properties of the organism studied or its host.

D. MATERIALS AND METHODS

The transposon donor DNA fragments and bacteriophages are described in detail in the original publications and summarised in Tables 3 and 4, respectively. The bacterial strains and plasmids used as well as the bacteriophage mutants generated in this study are described in the original publications. The experimental methods used in this study are described in the original publications and summarised in Table 5. References for published methods can be found in the articles.

Table 3. Transposon donors used in this study

Transposon	***Marker feature***	***Length (bp)***	***Reference***
SupF-Mu	amber suppressor	370	Haapa *et al.*, 1999a
LacZ'-Mu	lacZ' α fragment	460	II
LacZ'-Mu(NotI)	lacZ' α fragment	460	II, III, IV
Neoflox-Mu	neomycin resistance	1418	I

Table 4. Bacteriophages used in this study

Bacteriophage	***genotype***	***Reference***
PRD1	*wt*	Olsen *et al.*, 1974
ΦYeO3-12	*wt*	al-Hendy *et al.*, 1991
PM2	*wt*	Espejo and Canelo, 1968b

Table 5. Methods used in this study

Method	***Described and used in***			
3D matrix screening	I			
autoradiography		II		
blue/white screening		II	III	
Cre/*loxP* recombination test	I			
DNA sequencing and sequence analysis	I	II	III	IV
electroporation	I	II	III	IV
fitness analysis			III	
generation of transgenic mice	I			
HPLC	I	II	III	IV
in vitro transposition reactions	I	II	III	IV
in-gel DNA restriction				IV
luminescence assay			III	
mRNA analysis			III	
PCR	I	II	III	
phage growth and purification		II	III	IV
plaque hybridization		II		
plasmid DNA isolation	I	II	III	IV
restriction analysis	I	II		IV
self-ligation				IV
Southern blotting	I			
standard molecular cloning techniques	I	II	III	IV
viral DNA isolation		II	III	IV
Western blotting	I			

E. RESULTS AND DISCUSSION

1. Construction of novel mini-Mu transposons (I, II)

In the minimal Mu *in vitro* reaction (Savilahti *et al.*, 1995, Haapa *et al.*, 1999a), 50-bp fragments derived from the right end of bacteriophage Mu that contain R1 and R2 transposase binding sites are used as transposon donors. Artificial mini-Mu transposons contain these Mu end sequences at their termini in inverted-repeat orientation. Marker element(s) can be placed between these Mu ends in order to exploit mini-Mu transposons in various applications. Moreover, it is possible to modify the transposon end sequence somewhat without critically losing efficiency (Laurent *et al.*, 2000, Poussu *et al.*, 2005, see below).

For this study, three novel mini-Mu transposons were constructed by placing a marker sequence between two Mu ends in proper orientation. The 1418-bp Neoflox-Mu (Figure 2A in I) contains a neomycin phospotransferase (*neo*) gene, allowing selection in mouse ES cells (Thomas and Capecchi, 1987). Two *loxP* sites flank the marker gene, allowing its removal by site-specific recombination in cells expressing Cre recombinase from yeast (Sauer, 1998).

The 460-bp LacZ'-Mu (Figure 2A in II) contains the gene segment coding for the α fragment of *E. coli* β-galactosidase. The 460-bp LacZ'-Mu(NotI) transposon (Figure 2A in II) was generated from LacZ'-Mu by modifying the Mu end sequence to contain a *Not*I restriction site that partially overlaps R1 binding site (Figure 2B in II). Additionally, we used the 370-bp SupF-Mu that contains the *supF* amber suppressor tRNA gene from *E. coli* under its own promoter and has been described previously (Haapa *et al.*, 1999a; see Figure 2A in II). *supF* and *lacZ'* genes were selected as markers due to their small size and potential usefulness in gene expression analysis. The *Not*I sites of LacZ'-Mu(NotI) enable the removal of the marker gene entirely, leaving only a 15-bp insertion in the site of integration and allow further modification of viral genomes by traditional restriction-and-ligation -based methods.

Mini-Mu transposons were used in linear pre-cut form that ensures efficient *in vitro* assembly of stable transpososomes, a critical prerequisite of transposition (Craigie and Mizuuchi, 1987, (Savilahti *et al.*, 1995, Haapa *et al.*, 1999a).

The length of an intervening sequence between two Mu ends can be altered considerably. As the shortest mini-Mu transposon described, the 370-bp SupF-Mu transposes efficiently. The upper limit for the length of a mini-Mu transposon has not been defined; a 6.8-kb transposon functions efficiently (Lamberg *et al.*, 2002), but much longer constructs would probably be functional considering that the bacteriophage Mu genome is ~37 kb. However, several binding sites not present on mini-Mu transposons assist the assembly of native transpososome (see chapter 2.2.1. in Introduction).

2. Generation of mutant banks with transposition

In an *in vitro* transposition reaction the MuA transposase catalyses an efficient two-ended integration of linear mini-Mu transposons into target DNA at essentially random locations (Haapa *et al.*, 1999a). This is a highly efficient method of generating a large pool of insertion mutants having a random transposon insertion at

different locations on the target molecule. We performed such reactions on

A) linear 6688-bp fragment from the mouse *KCC2* locus (I)
B) linear 14927-bp dsDNA genome of bacteriophage PRD1 containing terminal proteins (II)
C) linear 39600-bp dsDNA genome of bacteriophage ΦYeO3-12 (III)
D) circular 10079-bp dsDNA genome of bacteriophage PM2 (IV)

The progression of reactions was monitored by collecting samples from different time points and analysing them in agarose gel electrophoresis (AGE). The resolution of transposition components, products, and by-products depends on comparative sizes of donor and target molecules as well as the absolute size and type (linear / circular) of target molecules. The transposons and reaction conditions used resulted in an efficient transposition reaction yielding mostly target molecules hit by a single transposon. Owing to the unspecific nature of *in vitro* transposition, some by-products (transposons hit by transposons, targets hit by multiple transposons) were also generated.

2.1. End-modified transposons enable further modifications of target (II, III)

While modified transposon ends containing a *Not*I restriction site have been used previously in reactions employing separate Mu ends (Laurent *et al.*, 2000, Kekarainen *et al.*, 2002), the impact of end-modification on two-ended mini-Mu transposons had not been tested previously. The transposition of end-modified LacZ'-Mu(NotI) progressed somewhat slower than with LacZ'-Mu (Figures 3A and 3B in II). However, this difference could be overcome by simply increasing the incubation time. These results indicate that end-modified mini-Mu transposons can be used to generate a large pool of insertion mutants.

After transposon integration, target molecules can be further modified. As the sequence between transposon ends is inconsequential to transposition, novel restriction sites can easily be added to mini-Mu transposons to allow post-integration modifications. However, in order to remove the bulk of the transposon, the Mu ends themselves must be modified.

Because each clone has a unique insertion site, a collection of such clones is a valuable source for further manipulation of the target genomes. This is especially useful with bacterial viruses as many bacteriophages are notably devoid of useful restriction sites. As an eight-cutter that has a restriction site on average every 65536 base pairs in a random double-stranded sequence, *Not*I is particularly useful in this regard. Novel *Not*I sites introduced to the ΦYeO3-12 genome by LacZ'-Mu(NotI) integration were used to generate four deletion mutants that were analysed to compare the effects of insertion and deletion mutations on same genes (II). Moreover, if the bulk of the transposon can be removed from the integration site, the impact of a transposon insertion can be minimised, allowing functional analysis on a much finer level (Hallet *et al.*, 1997, Hayes and Hallet, 2000, Laurent *et al.*, 2000, Kekarainen *et al.*, 2002, Poussu *et al.*, 2004).

3. Generation of mouse gene-targeting vectors (I)

Gene targeting into mammalian genomes by means of homologous recombination is a powerful technique for analysing gene function through disruption of genes

of interest and subsequent generation of transgenic animals. This technique was first applied on and has been most widely used with mouse ES cells (Thomas and Capecchi, 1987, Capecchi, 1989, Nagy and Rossant, 1996, Capecchi, 2001). The gene-targeting vector generally contains a selection marker flanked by two homology arms. Traditionally standard molecular cloning techniques involving multiple cloning steps and employing unique restriction sites have been used in the construction of such vectors, a time-consuming and laborious process (Torres and Kühn, 1997, Hasty *et al.*, 2001). A transposon insertion method based on yeast Ty*1* element (Westphal and Leder, 1997) allows quicker generation of gene targeting constructs. However, defining exact locations of transposon integrations and screening for suitable constructs remain slow and laborious with the strategy described.

We developed an easy strategy for generating a large pool of gene-targeting plasmid constructs and quick screening of the constructs desired based on the *in vitro* transposition of bacteriophage Mu. To test the strategy we targeted mouse *KCC2* gene that encodes a neuronal K^+/Cl^- cotransporter involved in the regulation of inhibitory neurotransmission in brain (Payne *et al.*, 1996, Rivera *et al.*, 1999). *KCC2* is lacking convenient restriction sites (Haapa *et al.*, 1999b), making generation of the constructs with standard recombinant DNA methods troublesome.

A transposon integration into an exon will produce a knock-out construction. Alternatively, an integration of *neo* cassette into an intron in the opposite direction relative to the target gene can produce a hypomorphic allele by reducing target gene expression (Meyers *et al.*, 1998, Nagy *et al.*, 1998, Partanen *et al.*, 1998). The mechanism involved is unclear and it is not known whether hypomorphic alleles can be produced from all genes.

A linear target fragment was mutagenised with an artificial mini-Mu transposon Neoflox (1418 bp) that contains a selection marker placed between *loxP* sites. Insertionally mutagenised target molecules were separated in AGE, isolated, cloned into a plasmid backbone and transferred into bacterial cells. When plasmids were amplified from randomly-picked colonies, 19 / 20 included target gene segment hit by a transposon, proving the efficiency of mutagenesis and separation methods.

3.1. A three-dimensional matrix is an effective way to sample the targeted clones

A total of 343 bacterial clones, each containing a plasmid that harbours a randomly inserted transposon in the target gene were grown and pooled according to a three-dimensional matrix system (Zwaal *et al.*, 1993, Koes *et al.*, 1995) with seven rows, seven columns, and seven blocks (Figure 3A in I); making the total number of pools 7 + 7 + 7 = 21. Each clone was pooled three times: once according to rows (Xnn, X = 1…7)), once according to columns (nYn, Y = 1…7), and once according to blocks (nnZ, Z = 1…7). This way, each clone is represented by a unique combination of a row, a column, and a block (XYZ). Individual clones harbouring a transposon at a specific site were identified by analysing the sizes of PCR products in all the pools and detecting which three pools contained a specific PCR product.

Multi-dimensional matrix systems are an efficient way to reduce the number of PCRs required to distinguish unique clones. In our system, 343 clones can be identified in 21 simultaneous reactions. Since exon 4 (147 bp) comprises 2.2 %

of the 6688-bp target segment, assuming random integration and 100 % transposition accuracy, the probability that in none of the 343 clones the transposon integration site would reside within exon 4 is almost negligible ($0.978^{343} \approx 0.05$ %).

To confirm the efficiency and reproducibility of our selection method, two sets of clones were pooled in the same manner and a clone harbouring a transposon close to the 5´ end in exon 4 of *KCC2* was singled out (Figure 3C in I and not shown) from each pool. Third mutant, harbouring a transposon in intron 3 was identified for the construction of a potentially hypomorphic allele.

Obviously, PCR screens can also be performed on single clones. However, if multiple integrants are to be screened, pooling systems are most likely less laborious, especially if the targeted region(s) comprise only a small proportion of the target molecule. The three-dimensional matrix system has been previously used to detect rare *in vivo* transposon insertions into defined target regions (Zwaal *et al.*, 1993, Koes *et al.*, 1995). Our results indicate that it is also very useful and reproducible method to screen multiple plasmid constructions in one set of simultaneous reactions.

3.2. A variety of different mutants can be screened and further generated

The Mu *in vitro* insertional mutagenesis strategy described yields a mutant bank in which each clone harbours a transposon at a unique location and integration sites are spread randomly throughout the target molecule. With three-dimensional matrix a variety of mutants that have different regions disrupted by a transposon can be screened.

In addition to two regular knock-out constructs, one potentially hypomorphic mutant was picked up and the subsequent Western blot analysis of a homozygous mouse brain tissue revealed substantially reduced KCC2 protein levels (Figure 5 in I), confirming that the allele produced was hypomorphic. In a follow-up study, hypomorphic mice produced using our constructs displayed several behavioural phenotypes including increased anxiety-like behaviour and seizure susceptibility, having potential implications for human neuropsychiatric disorders such as epilepsy and anxiety (Tornberg *et al.*, 2005).

Moreover, conditional allele -containing constructs can be generated from constructs harbouring a transposon in an intron by adding a third *loxP* site on the other side of the exon to be removed. This can be done with a *loxP*-containing transposon, or with a Neoflox transposon followed by removal of selection cassette and another transposition with Neoflox transposon, or by cloning a *loxP* site using traditional restriction-and-ligation -based methods (Figure 6 in I).

4. The whole-genome analysis of bacteriophages (II, III, IV)

To date, the use of *in vivo* transposon mutagenesis in the functional analysis of viruses has been difficult due to the incompatibility between transposition and viral replication machineries. The cloning of viral genomes in infective form in plasmids, cosmids, or BACs has been used with *in vivo* transposon mutagenesis and whole-genome analysis of various cytomegaloviruses (Brune *et al.*, 1999, Smith and Enquist, 1999, Hobom *et al.*, 2000, Yu *et al.*, 2003). *In vitro* transposition, generally more efficient than reactions *in vivo*, has also been used previously to mutagenise complete or partial genomes cloned in a suitable vector (Laurent *et al.*, 2000, Kekarainen *et al.*, 2002). These

methods, while effective if conducted on a suitable target, are not applicable to those viral genomes that, for example, contain terminal proteins that prohibit cloning into a standard vector (e.g. PRD1). Moreover, vector sequences attached to cloned genomes might make it difficult to analyse correlations between functions disrupted and phenotypes observed (Zhan *et al.*, 2000b).

We utilised bacteriophage Mu *in vitro* transposition to develop an efficient strategy for the whole-genome analysis of bacteriophages by mutagenesis. The method developed is rapid, efficient, and does not require the genome to be cloned. It is also directly applicable to all DNA and RNA viruses that have infective genomes in the form of DNA, whether cloned or uncloned. In our strategy, infective viral genomes are mutagenised *in vitro* to generate a large mutant pool that is subsequently transferred to host cells plated on suitable bacterial growth plates. Resulting mutant virus progeny are identified and transposon integration sites determined to assess genomic regions essential or nonessential for the viral life cycle.

4.1. Some mutagenised viral genomes remain infective (II, III, IV)

All the bacteriophage species in this study – PRD1 (Lyra *et al.*, 1991), ΦYeO3-12 (S. Kiljunen and M. Skurnik, unpublished), and PM2 (van der Schans *et al.*, 1971, Kivelä *et al.*, 2004) – have infective genomes. In other words, the transfer of viral genome into the host cytoplasm is sufficient to initiate productive infection. The viral genomes insertionally mutagenised in an *in vitro* transposition reaction were transferred to host cells by electroporation (II, III, IV). Incubation with host bacteria yielded plaques on growth plates using all three viruses (II, III, IV), reiterating the fact that these viral genomes are infective.

Mutant plaques were screened from growth plates by autoradiography (II) or visual inspection using blue/white screening (II, III). With PM2, mutant genomes were isolated in agarose gel electrophoresis prior to electroporation, yielding almost exclusively mutant plaques (IV).

The growth of wild-type-like plaques carrying a transposon proved that it is feasible to integrate a Mu-derived transposon into the viral genomes of all the bacteriophages studied with no evident effects on the viral life cycle. These results demonstrate that within the context of a bacteriophage genome, gapped and single-stranded transposition intermediates are repaired in host cells following electroporation. However, it is not known whether this repair is done by host- or virus-encoded factors.

Furthermore, the results indicate that it is possible to introduce at least an additional 460 bp of DNA into genomes of all the bacteriophages studied without affecting genome packaging into viral particles, indicating that all the available space in the virus particle is not taken up by the wild-type genome. This is consistent with previous experiments where additional 393 bp was inserted into the PRD1 genome by more traditional methods (Bamford and Bamford, 2000). With ΦYeO3-12 we detected one clone that had two separate LacZ'-Mu(NotI) integrations, demonstrating that an additional 920 bp can be packaged into ΦYeO3-12 virus particles.

4.2. Mutant bacteriophages can be rapidly identified with blue/white screening (II, III)

Mutant plaques can be identified from growth plates by autoradiography (II). However, for rapid visual screening of mutants we developed a system based on the blue/white screening method (Messing *et al.*, 1977, Vieira and Messing, 1982).

Transposons LacZ'-Mu and LacZ'-Mu(NotI) contain promoterless *lacZ*' gene that produces colour on indicator plates if the transposon integrates in a proper orientation into a genomic region that is under a functional promoter. PRD1 and ΦYeO3-12 genomes mutagenised with LacZ'-Mu / LacZ'-Mu(NotI) some plaques displayed distinctive blue colour either at or around them (Table 2 in II, Figure 5 in III) proving that blue/white screening is a rapid and effective method to screen mutant bacteriophages. By comparing the distribution and orientation of transposons in clones forming blue and colourless plaques it is possible to draw conclusions on the direction of gene expression at transposon integration sites (Figure 7 in II). Moreover, the intensity of colour varied among virus clones (II, III), evidently reflecting differences in the level of gene expression. However, colour intensities were not compared in our studies.

An initial attempt to use blue/white screening using SupF-Mu, PRD1, and host strain having an amber mutation of *lacZ* gene failed (II), possibly due to the shutdown of host gene expression upon bacteriophage infection or inadequate timing in the gene expression of the blue/white system components. Blue/white selection with promoter-containing *lacZ'* transposons was equally unsuccessful, implicating that foreign promoter elements can negatively influence viral gene expression (II, data not shown).

4.3. Circular mutant genomes can be separated with agarose gel electrophoresis (IV)

Since blue/white screening is not currently available in *P. espejiana*, the host of PM2, we set out to separate mutagenised PM2 genomes from wild-type genomes prior to their introduction into the host. The *wt* PM2 genome is circular, superhelical dsDNA, naturally occurring in the covalently closed circular (ccc) form. A transposon insertion into the genome results in two single-stranded gaps around the transposon. Consequently, target superhelicity is lost and the mutagenised genome takes the open circular (oc) form. These two topological forms can be separated in AGE. AGE separation proved very effective as 85...95 % of the clones obtained contained a transposon insertion (IV and data not shown). This method, applicable at least to all small circular, superhelical genomes, is not dependent on any marker gene activity.

4.4. Transposon integrations can be verified by restriction analysis (II, II, IV)

Viral DNA was isolated from purified mutant plaques and subjected to restriction analysis (II, III, IV) to determine 1) the number of transposons integrated 2) the approximate genomic location of such integrations, and 3) whether more complex genomic rearrangements had occurred.

In most clones, 145 of 167 mutants in PRD1 (II), 14 of 17 mutants in ΦYeO3-12 (III), and 99 of 103 mutants in PM2 (IV + unpublished data) the restriction pattern was consistent with a

single transposon integration. In all the bacteriophage genomes mutagenised, transposon integrations were observed on several different restriction fragments, indicating that transposons can integrate into several locations along the genome. Complex rearrangements were observed with low frequency (< 10 %) in all the genomes mutagenised. The most common rearrangements were deletions of a short genomic region (up to 1560 bp in III), probably arising from two transposon integrations close to each other and subsequent homologous recombination between these elements.

4.5. Transposon integration sites shed light on essential regions of the genome

Conceptually, there are two ways to identify essential genes or regions of the genome, negative and positive approaches (Judson and Mekalanos, 2000). The negative approach (Figure 6A) identifies many regions that are not essential and presumes that everything else is essential. The obvious problem with negative approach is that it requires a large number of independent mutations in order to be reliable and even then the essentiality of a given gene is just a probability. Contrastingly, the positive approach (Figure 6B) identifies genes that are essential by generating a conditional mutation and showing that it has a lethal phenotype. The problem with this approach is that producing conditional phenotypes is laborious; it is very inconvenient to try to cover the whole genome with the positive approach.

The transposon insertion method described in this study takes a fundamentally negative approach to identify essential and nonessential regions of the genome. The number of infective bacteriophages decreases gradually and substantially in an *in vitro* transposition reaction, indicating that at least a fraction of transposon integrations inactivate viruses by insertion into the essential regions of the genome (Figure 3B in II). Viral mutants able to form plaques are collected and their transposon integration sites, reflecting nonessential genomic regions, are determined by DNA sequencing outwards to both directions from the transposons used (Figure 2A in II).

4.5.1. Integration sites in infective mutagenised genomes are not random (II, III, IV)

Sequencing information was collected from 130 infective PRD1 mutants (Figures 6 and 7 in II), 17 infective ΦYeO3-12 mutants (Figure 1 and Table 2 in III) and 101 infective PM2 mutants (Figure 2 and Table 1 in IV).

Overall, in all the genomes mutagenised, the transposon integrations observed were concentrated on a few genomic regions. In PRD1 (II) the integration sites were concentrated mostly on the ends of the genome. Genomes mutagenised with SupF-Mu (370 bp) and LacZ'-Mu / LacZ'-Mu(NotI) (both 460 bp) displayed similar insertion patterns. In PM2 (IV) LacZ'-Mu(NotI) transposon integrations concentrated on a few regions, namely gene *XV* and promoter region boundaries. As transposition of Mu into target sites is near-random (Mizuuchi and Mizuuchi, 1993, Butterfield *et al.*, 2002, Haapa-Paananen *et al.*, 2002) and most transposon integrations into a coding region are expected to inactivate the gene in question, the distribution of integrations observed in functional viruses reflects the uneven distribution of essential and nonessential regions in the genome.

The whole-genome analysis using our method is obviously dependent on a

sufficient number of insertion mutants collected. In PRD1 98 / 130 mutants were situated in unique insertion sites, while the remaining 32 mutants mapped to 13 different positions (Figures 6 and 7 in II). With PM2, 82 / 101 insertions located at unique positions while remaining 19 were situated at 8 different sites (IV). These figures indicate that our conditions are near-saturating. Even though there are likely a few undetected genomic sites that can tolerate a transposon integration, it is probable that all the major nonessential regions have been detected in our analysis.

In the case of ΦYeO3-12 (III) the total number of mutants is too low to draw conclusions on essential regions of the genome, even though it is striking that all the integrations are located in the early and middle regions of the genome. Moreover, ΦYeO3-12 mutants were detected by blue/

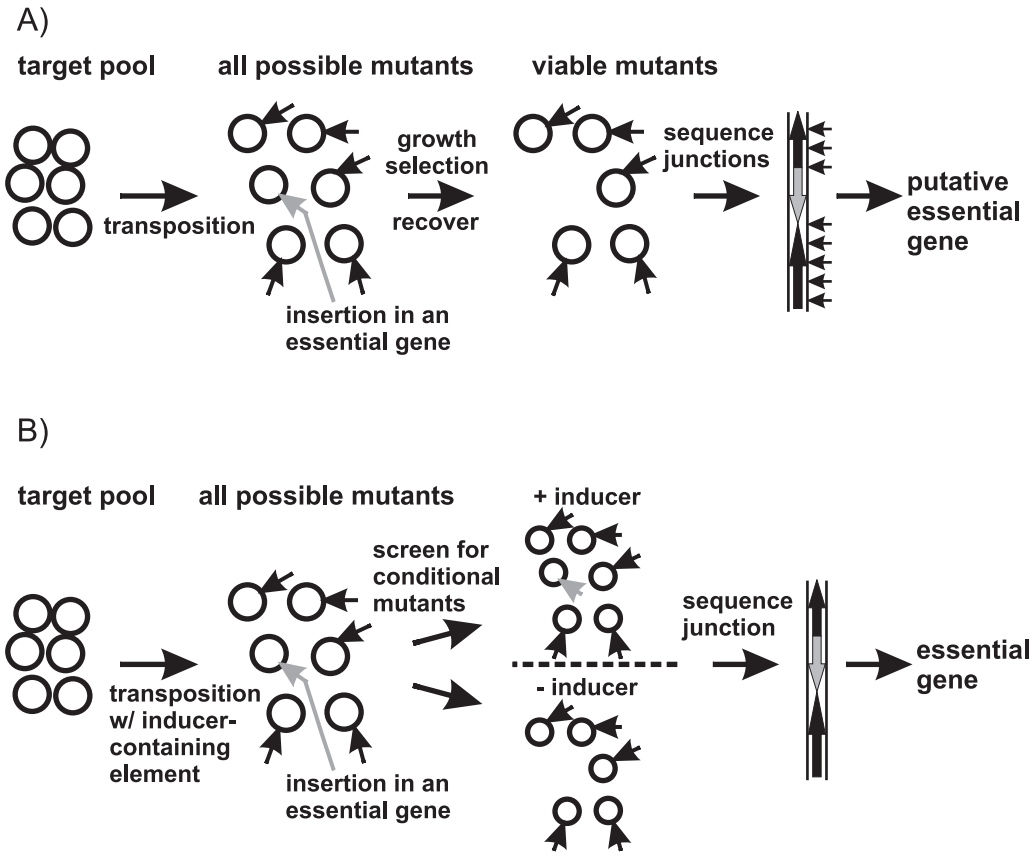

Figure 6. Two ways to identify essential genes or regions of the genome. A) The negative approach identifies non-essential regions. Global transposon mutagenesis generates a large number of mutants that are conducted to selection. Viable mutants are recovered and the transposon integration sites are identified. Regions where no transposon integrations are detected are assumed to be essential. B) The positive approach identifies essential regions directly. Mutagenesis with a transposon containing an inducible promoter into a promoter region of a gene generates conditional mutants dependent on inducer for viability. Transposon integration site is determined, allowing the identification of an essential gene downstream from promoter. (Adapted from Judson and Mekalanos, 2000)

white screening, meaning that they would all be within coding regions, thus excluding nonessential noncoding regions from the analysis. In PRD1 the distribution of integration sites of the mutants producing blue plaques was noticeably more limited than of those producing colourless plaques (Figure 7 in II), demonstrating the limitations of blue/white screening when used as the sole method of detecting mutants.

4.5.2. Genes and ORFs can be divided into three classes (II, IV)

According to our analysis of PRD1 and PM2 integration sites, the genes and ORFs studied can be divided into three classes:

1) **ORFs that do not tolerate transposon insertions** are most likely protein-coding genes that code for a function essential to the viral life cycle. However, a transposon integration into an upstream location can cause disturbances also in the expression of downstream genes transcribed from the same mRNA. In this regard internal-promoter-containing SupF-Mu might be more destructive than promoterless *lacZ'* transposons.

In the PRD1 genome most known genes did not tolerate transposon integrations, including those encoding viral DNA polymerase (gene *I* – Mindich *et al.*, 1982, Savilahti *et al.*, 1991), major structural capsid protein (gene *III* – Bamford *et al.*, 1983, Bamford and Bamford, 1990, Butcher *et al.*, 1995), proteins forming the vertex complex (genes *II*, *V*, and *XXXI* – Mindich *et al.*, 1982, Grahn *et al.*, 1999, Rydman *et al.*, 1999, Bamford and Bamford, 2000, Caldentey *et al.*, 2000), and many proteins associated with the viral membrane (Bamford *et al.*, 2002).

In the PM2 genome the genes encoding major structural proteins under late promoter (Kivelä *et al.*, 1999, Huiskonen *et al.*, 2004), as well as genes *XIII*, *XIV*, and *XVI* encoding repressors and activators involved in the switch from early to late promoter (Männistö *et al.*, 2003) did not tolerate any transposon integrations. ORF h was in this class as well, indicating that it probably encodes a protein product essential to the viral life cycle.

2) **ORFs that tolerate insertions but apparently only in restricted locations** are likely to be protein-coding genes the product of which is essential for the viral life cycle – but the C-terminal part of the protein is nonessential. In these cases, a truncated, yet still functional, protein was expressed (Figure 8 in II, not shown for IV).

Known PRD1 genes that tolerated some transposon insertions were *VIII* (encoding the genome terminal protein), *XII* (a ssDNA binding protein), and *XVII* (a non-structural assembly factor) while ORF t seemed to tolerate insertions only into its distal region, indicating that it is likely to contain a bona-fide protein-coding gene. ORF t (renamed gene *XXXV*) has later been identified as the gene encoding the PRD1 holin (Rydman and Bamford, 2003), thereby confirming our assessment.

In PM2, gene *XII* (replication initiation protein – Männistö *et al.*, 1999) tolerated transposon integrations to the distal 3' end of its coding region, indicating that the N-terminal part of the protein product is essential for function. Gene *IX* (a structural protein believed to be involved in genome packaging – Männistö *et al.*, 1999, Strömsten *et al.*, 2005) tolerated only the substitution of its final amino acid, indicating that the protein product is essential.

ORF q of PRD1 as well as ORFs b and d of PM2 tolerated a transposon integration at a single site, not enough to draw conclusions on their essentiality.

3) **ORFs that tolerate transposon insertions throughout their entire length** either code for proteins nonessential for the viral life cycle in conditions used or are phantom ORFs that do not code for a protein product.

In PRD1, genome gene *XIX* (a ssDNA binding protein), known to be nonessential (Mindich *et al.*, 1982, Pakula *et al.*, 1993), tolerated transposons throughout its coding region. Moreover, ORFs u and v are at most nonessential and, on the basis of the lack of colour forming with *lacZ'*, very likely do not code for a protein product.

In PM2, gene *XV* (repressor of early promoters – Männistö *et al.*, 2003) unexpectedly appeared to be nonessential as it tolerated 39 transposon integrations throughout its coding region. ORFs e, k, and l also tolerated integrations at multiple locations.

Noncoding regions that are transposon-intolerant are apparently involved with essential virus functions, e.g. virus replication, genome packaging, or regulation of gene expression. This was evident in PRD1 ITRs, where extreme 85 bp was transposon-intolerant at both ends. In PM2, transposon integrations near promoter regions clearly affected viral gene expression, delaying host cell lysis. In ΦYeO3-12, the mutants were exclusively detected using blue/white screening (III), effectively selecting for coding-region insertions. Thus, the evaluation of noncoding regions was not possible.

It must be noted that our assessments are only valid in the experimental conditions used. For instance in the case of ΦYeO3-12, several mutations had more severe effect on virus fitness in *Y. enterocolitica* than in *E. coli* (Table 4 and Figure 2 in III).

4.6. Additional experiments can complement integration site information

The generation of insertionally mutagenised bacteriophage genome pools, their subsequent introduction into host cells, and the analysis of integration sites in viable clones yields important information on essential and nonessential regions of the whole virus genome. Coupled to blue/white screening, this method reveals information on the direction and potentially also the intensity of gene expression along the genome. In addition to general genomic information directly derived from the integration sites, additional experiments can be conducted on the mutant viruses generated to obtain more specific information on the genes altered.

4.6.1. Fitness analysis (III)

Fitness analysis (Rokyta *et al.*, 2002) measures the number of doublings of infective bacteriophage particles per generation. These can be used to determine the effect of transposon integration on the overall fitness of virus. To study whether the changes caused by transposition integration are host-specific, viruses capable of infecting several host species can be analysed using alternative hosts.

4.6.2. Deletion mutants and complementation tests (III)

The introduction of rare *Not*I restriction sites enable further genome modifications with standard molecular cloning techniques. The bulk of the transposon can be removed, generating a library of short

insertions. Alternatively, other control elements can be inserted to the integration site.

Deletion mutants are useful in studying whether the phenotypes observed are due to the disruption of a particular target gene or the insertion of the transposon itself. Deletion mutants can be generated using two clones having transposon integrations at nearby sites, digesting genomes with *Not*I and ligating the 5' part of one clone to the 3' part of the other clone. To study whether the function disrupted can be complemented *in trans*, intact target gene can be cloned into a plasmid vector and transferred into the host cell. This type of analysis can also confirm that the phenotype detected is due to the mutation induced and not a random mutation elsewhere in the genome.

4.6.3. Level of gene expression (III)

Since bacteriophage genes are polycistronic, i.e. several genes are controlled by the same promoter, a phenotype observed can result not only from a mutation to the gene itself but also due to polar effects of mutations further upstream in the transcript. To study the polar effects of a mutation, an analysis of mRNA transcription timing and levels of a downstream gene can be studied by RT-PCR.

The activity of viral RNA polymerase can also be studied in general. In the luminescence assay a firefly luminescence gene is inserted under a bacteriophage-specific promoter in a plasmid that is then transferred into a host cell. During infection, luminescence levels are measured to monitor the activity of the bacteriophage RNA polymerase.

4.6.4. One-step growth assay (IV)

With all the bacteriophages mutagenised some clones displayed plaque morphologies deviating from wild-type viruses. In PM2 30 clones forming non-*wt* plaques were further analysed in the one-step growth assay in which the absorbance of host bacteria infected with mutant viruses is followed until lysis. This experiment yielded information on factors affecting virus-induced host cell lysis (IV).

4.7. The method developed yielded novel information on each bacteriophage studied

4.7.1. PRD1 (II)

PRD1 was ideally suited to be the target with which to develop the *in vitro* transpositional mutagenesis method. Its terminal protein -containing (Bamford *et al.*, 1983), infective (Lyra *et al.*, 1991) genome is completely sequenced (Bamford *et al.*, 1991) and transcriptional organisation has been established (Grahn *et al.*, 1994), providing comparative data for functional analysis. Moreover, the genome is almost totally devoid of useful restriction sites, limiting genome manipulation with traditional methods.

Experiments with PRD1 (I) proved that it is possible to use *in vitro* transposition to study an isolated but uncloned bacteriophage genome, even one with terminal proteins attached. Furthermore, the analysis was possible with multiple transposons and blue/white selection proved efficient. The end-modified lacZ'-Mu(NotI), that allows further modification with standard molecular cloning techniques, was almost as effective as unmodified transposons.

The information obtained on the functional organisation of the PRD1

genome was in accordance with previous studies (Grahn *et al.*, 1994, updated in Bamford *et al.*, 2002), further validating the method developed. Moreover, the mutants generated can be used to analyse functions of individual genes and proteins. Rydman and Bamford (2003) identified ORF t as gene encoding PRD1 holin using more traditional methods and also analysed the effect of six transposon integration mutants generated in our study on the holin function.

4.7.2. ΦYeO3-12 (III)

The results on ΦYeO3-12, the second bacteriophage mutagenised with mini-Mu transposons, proved that the methods developed are applicable to bacteriophages other than PRD1.

The emphasis of the analysis of ΦYeO3-12 was on the effects of individual genes and the possibility of genome manipulation. Hence, unlike with PRD1 and PM2, the amount and type of mutants collected is insufficient for a true whole-genome analysis. Nevertheless, the use of fitness, complementation, and gene expression analysis on individual mutants demonstrated how individual clones obtained by large-scale mutagenesis can be further analysed individually using the positive approach for the assessment of essentiality (Judson and Mekalanos, 2000 and Figure 6). Equally important, the deletion mutants generated demonstrate the ease of which transposons can be utilised to alter the genome of a bacteriophage that has been particularly troublesome to manipulate using standard molecular cloning techniques (S. Kiljunen, personal communication).

The results on ΦYeO3-12 also illustrate the limits of whole-genome analysis. Based on complementation analysis, transposon integrations upstream of gene *1* encoding RNA polymerase delay its expression and lead to reduced bacteriophage fitness. Thus, genes upstream of gene 1 appear more important to bacteriophage fitness with transposon analysis than they really are. On the other hand, integrations to genes *1.3* and *3.5* encoding bacteriophage ligase and lysozyme, respectively, reduce bacteriophage fitness in *Y. enterocolitica* but not in *E. coli*, indicating that ΦYeO3-12 has adapted to use the former as a host and further reinforce the fact that the essentiality of a given gene in a study applies only to conditions used. While the whole-genome mutagenisation gives a useful overview of the genome studied, functions of individual genes can only be reliably assessed with concomitant studies focused on individual genes – and in a biologically relevant system.

4.7.3. PM2 (IV)

The mutagenisation of PM2 proved that the method developed is also applicable to a bacteriophage and a host that and are from a completely different environment and belong to different phylogenetic groups than those used previously. Furthermore, transpositional mutagenesis was shown to be a useful method in a system for which no alternative genetic tools are available. The gel-electrophoresis -based separation method of mutant genomes proved very efficient, generating a pool of almost exclusively mutant clones.

Based on the one-step growth assay analysis of the mutants and analogous genes present in other viruses, ORFs k and l were identified to be involved in host cell lysis, suggesting that they form a lysis cassette. These genes merit further examination.

4.8. In PM2 / *P. espejiana* system not all integration sites are repaired in the same way (IV)

One of the hallmarks of Mu transposition, the 5-bp target site duplication (Allet, 1979, Kahmann and Kamp, 1979) has been evident in Mu *in vitro* transposition applications where transposon products are repaired in *E. coli*, (Haapa *et al.*, 1999a, Haapa *et al.*, 1999b, Laurent *et al.*, 2000, Haapa-Paananen *et al.*, 2002, Kekarainen *et al.*, 2002, Poussu *et al.*, 2004). The same pattern was observed in all the mutants generated for mouse gene-targeting (I), as well as in all the PRD1 (II) and ΦYeO3-12 (III) mutants apart from those that had undergone more complex rearrangements. However, in the case of PM2, the 5-bp target site duplication was not uniform as 37 of 99 (37%) mutants had one normal transposon / target junction, but retained the 4-nt single-stranded transposon flank in the other end. Seven of these mutants had 1-bp target site duplication, while in 29 mutants the target sequence had no duplications or deletions. One mutant displayed a one-base-pair deletion.

In Mu transposition the transposon / target junction single-stranded gaps are repaired by host DNA replication / repair machinery (Kruklitis and Nakai, 1994, Nakai and Kruklitis, 1995). In previous Mu *in vitro* applications, the processing of the transposition product ends has occurred in *E. coli*. With PM2, its natural host *P. espejiana* was used. Alternative end processing has also been observed in yeast using preassembled Mu transpososomes as donors (H. Turakainen and H. Savilahti, unpublished), indicating that non-*E. coli* hosts might at least occasionally repair the transposon ends differently. However, it is not entirely impossible that virus-encoded factors were involved in the repair of transposon / target junctions involving viral genomes. This possibility could be studied by transferring mutagenised genomes of a broad-host-range virus to a variety of hosts and detecting the transposon / target junction sequences.

4.8.1. A model for the alternative processing of transposon ends observed in *P. espejiana*

We present a feasible model for alternative processing of transposon / target junctions observed with PM2 in *P. espejiana*. Transposition generates a transposon / target intermediate that contains 5-nucleotide (nt) single-stranded regions and 4-nt non-complementary overhangs from transposon flanks. In standard processing (Figure 5A in IV) 4-nt overhangs are cleaved by a nuclease and 5-nt single-stranded region is filled in by a DNA polymerase. The nick is sealed by a ligase. Contrastingly, in the alternative end processing (Figure 5B in IV) the 5-nt single-stranded region is cleaved by an endonuclease and ensuing DNA ends are processed by an exonuclease and a DNA-polymerase, resulting in blunt ends that are joined by a ligase. Within the joint, the 4-nt transposon flank is retained. This model of alternative end processing would explain different outcomes observed at alternatively-processed ends (1-bp duplication / no change in target sequence / 1-bp deletion).

Interestingly, even though this alternative processing was relatively common (37 % of mutants had one alternatively processed end), no mutants with two alternatively processed transposon ends were detected. If the method of end processing is uncoupled

from processing at the other transposon / target junction, in a pool of 99 mutants one would statistically expect to encounter several mutants with two alternatively processed ends. Even though the sample size is too small to draw any definite conclusions, this might indicate that the end processing of one end restricts the processing of the other end. According to the model presented, the transposon / target bond is temporarily cleaved in the alternative processing (Figure 5B in IV). Evidently the repair machinery can at least occasionally ligate loose ends back together at one transposon / target junction, indicating that ends might be held together in a nucleoprotein complex. However, this type of end rescue might be too difficult if both junctions are cleaved, effectively releasing the transposon entirely from the target site.

4.8.2. Alternative end processing can affect Mu *in vitro*-based applications

In its natural host, *E. coli*, Mu transposon ends appear to be repaired in uniform fashion, making the transpositional mutagenesis system practical for a variety of applications. Alternative end processing is a phenomenon that potentially has consequences affecting applications based on the Mu *in vitro* transposition. On one hand, the fact that ends are not processed in the same way every time provides an opportunity to generate slightly different mutations on the same site, thus providing additional material for functional analysis. On the other hand, applications that rely on consistent 5-bp target site duplication, for example pentapeptide scanning mutagenesis (Hallet *et al.*, 1997, Poussu *et al.*, 2004), would be disrupted in some non-coli hosts.

F. CONCLUSIONS AND FUTURE PERSPECTIVES

This study expanded the use of bacteriophage Mu *in vitro* transposition methodology in functional genetics and genomics by describing novel methods applicable to the targeted transgenesis of mouse and the whole-genome analysis of bacteriophages. The methods described in this study are rapid, efficient, and easily applicable to a wide variety of organisms, demonstrating the potential of the Mu *in vitro* transposition technology in the functional analysis of genes and genomes.

An easy-to-use, rapid strategy to generate constructs needed for targeted transgenesis of mouse and to sample desired constructs from a large mutant pool was developed (I). This method was shown to generate a large pool of mutants. Desirable mutants with transposon integrated in the region of interest were conveniently picked up with the PCR-based three-dimensional matrix screening technique adopted. Transpositional mutagenesis was also shown to produce potentially hypomorphic alleles and one such allele was shown to cause a hypomorphic phenotype in a transgenic mouse. In addition, the constructs generated can easily be used to produce conditional alleles with an additional round of transposon mutagenesis and mutant screening. The strategy developed is suitable for generating gene-targeting constructs of any gene from mouse or any other organism for which analogous targeted gene disruption techniques based on mutagenesis and subsequent homologous recombination are used. The hypomorphic alleles produced in this study were later used to study behavioural phenotypes in KCC2-deficient mice (Tornberg *et al.*, 2005). Moreover, other selection methods could easily be developed for this system. Placed under a prokaryotic promoter, neomycin phosphotransferase gene confers resistance to prokaryotic antibiotic kanamycin. If the *neo* cassette used was placed under a dual prokaryotic/eukaryotic promoter, this kanamycin selection could be used to generate a mutant construct pool. More complex mini-Mu transposons can also be designed to allow additional marker genes and sophisticated selection schemes, as was done in a later study (Jukkola *et al.*, 2005).

Similarly, an efficient strategy to study the functional organisation of bacteriophages was developed. Bacteriophage genomes were mutagenised with marker transposons and selected for their ability to form plaques to distinguish regions essential and nonessential for the viral life cycle. The strategy was shown to function with three bacteriophages from different genera and hosts: PRD1 (II), ΦYeO3-12 (III), and PM2 (IV). All three viruses were able to survive transposon integration into their genomes, demonstrating additional packaging capability of their respective capsids. Blue/white screening was shown to be an efficient way to screen for mutant viruses with *E. coli* host, providing additional information on direction and intensity of gene expression at the integration sites. Furthermore, mutagenised PM2 genomes were efficiently isolated with agarose gel electrophoresis, bypassing the need for screening. End-modified transposons were shown to be efficient donors of transposition and their capability to further modify virus genomes was demonstrated by generating deletion mutants of ΦYeO3-12. Additional experiments were conducted on selected mutants to determine their fitness, growth curves, and/or whether the

mutation could be complemented *in trans*. New biological information was gathered from all three bacteriophages. Essential and nonessential regions of PRD1 and PM2 were defined, revealing several nonessential genes and ORFs. In PRD1 the essential region of ITRs were also defined. In PRD1, where information on the genomic organisation was available, the results were in accordance with previous and subsequent studies conducted with traditional methods (Grahn *et al.*, 1994, Bamford *et al.*, 2002, Rydman and Bamford, 2003). In ΦYeO3-12 ligase and lysozyme genes were identified to function host-specifically and transposon insertions upstream of gene *1* were detected to cause growth defects due to delayed expression of gene *1*. In PM2 lysis-associated genes were identified. With PM2, the processing of transposon ends was not uniform and a feasible model was presented to explain this phenomenon. The effect of different hosts to end processing could be studied by mutagenisation of a broad-host range virus and subsequent infection of different host bacteria.

Rapid and efficient Mu *in vitro* transpositional mutagenesis strategy is directly applicable to all prokaryotic and eukaryotic dsDNA viruses with infective genomes. It can equally be applied to RNA and ssDNA viruses that can be cloned as dsDNA in an infective form on a replicon and, using a helper virus, probably to most other viruses that do not pass aforementioned criteria. Moreover, the limits of packageable genome for each bacteriophage could easily be studied with a set of different-sized transposons. The genomes of independent organisms might be too large to handle and mutagenise *in vitro* with the strategy developed and the sequencing costs to determine integration sites in much larger genomes might currently prove prohibitive. However, it could be worthwhile to test this type of mutagenesis with the genomes of smallest intracellular parasite bacteria. With very large genomes the use of *in vivo* mutagenesis with preassembled transpososomes (Goryshin *et al.*, 2000, Lamberg *et al.*, 2002) is likely to be the method of choice.

G. ACKNOWLEDGEMENTS

This work was carried out at the DNA Recombination Laboratory of the Institute of Biotechnology, Research Program in Cellular Biotechnology and at the Faculty of Biosciences, Division of Genetics, University of Helsinki. I wish to express my sincere gratitude to professor Mart Saarma, the Director of the Institute of Biotechnology, for providing excellent facilities to work and for always finding time to discuss when I needed career advice. The current and former operative heads of the Division of Genetics, Tapio Palva, Hannu Saarilahti, and Pekka Heino, are also thanked for their help with the PhD bureaucracy.

I wish to express my gratitude to my supervisor, docent Harri Savilahti who provided an interesting subject to study and was willing to renew our working relationship when I expressed wish to complete this once-discontinued thesis.

I thank reviewers professor Jari Valkonen and docent Päivi Ojala for the time they spent reading this text as well as for their insightful suggestions for improvement.

Additional financial support to complete this thesis from the Finnish Cultural Foundation is gratefully acknowledged. Helsinki Graduate School in Biotechnology and Molecular Biology is acknowledged for organising interesting courses. Big thanks also go to Tinde Päivärinta for going way out of his way to get this thesis published.

The current members of our lab, Hilkka T, Pirjo R, Anja P, Maria P, Sari T, Anna-Helena S, Eini P, Tiina L, and Lotta H, as well as all the former members of the group, are thanked for creating an enjoyable working environment. Especially big thanks go to Eini and Anja for always making me, often the only male, feel part of an otherwise all-female gang. I also wish to acknowledge all the co-authors for their contribution to our mutual papers.

Special warm hugs go to my friends in the crime of studying in Viikki: Mari, Tuula, Elina, Hanna, and, most of all, Tarja. Thank you for all the fun times.

Especially huge thanks go to my oldest friends, Kaitsu and Rami, for always providing the most carefree – and competitive – moments of my life. I also wish to thank Pekka for sharing all the ups and downs in life and Riku as well as Sami and Piia for our lasting friendships. Warm thanks also go to all my friends in Wartti Basket.

I am deeply grateful for all the love, support, and encouragement of my mother Sirkka. I also thank my late father Reijo, who is mostly responsible for my interest in natural sciences. My brother Jaakko and his wife Tarja are also warmly thanked for all the family moments.

Finally, the biggest thanks go to Anja, whose relentless affection and support, often at the cost of putting her own concerns to the background, carried me through the final stages of this thesis. A gentle kiss and a hug, since the words are not nearly enough to express my appreciation.

Helsinki, April 2006

H. REFERENCES

Abad JP, De Pablos B, Osoegawa K, De Jong PJ, Martin-Gallardo A, Villasante A. (2004). TAHRE, a novel telomeric retrotransposon from Drosophila melanogaster, reveals the origin of Drosophila telomeres. *Mol Biol Evol* 21(9):1620-1624.

Abenes G, Chan K, Lee M, Haghjoo E, Zhu J, Zhou T*, et al*. (2004). Murine cytomegalovirus with a transposon insertional mutation at open reading frame m155 is deficient in growth and virulence in mice. *J Virol* 78(13):6891-6899.

Ackermann HW, DuBow MS, Gershman M, Karska-Wysocki B, Kasatiya SS, Loessner MJ*, et al*. (1997). Taxonomic changes in tailed phages of enterobacteria. *Arch Virol* 142(7):1381-1390.

Adzuma K, Mizuuchi K. (1989). Interaction of proteins located at a distance along DNA: mechanism of target immunity in the Mu DNA strand-transfer reaction. *Cell* 57(1):41-47.

Adzuma K, Mizuuchi K. (1988). Target immunity of Mu transposition reflects a differential distribution of Mu B protein. *Cell* 53(2):257-266.

Agrawal A, Eastman QM, Schatz DG. (1998). Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system. *Nature* 394(6695):744-751.

Ahmed A. (1984). Use of transposon-promoted deletions in DNA sequence analysis. *J Mol Biol* 178(4):941-948.

Akerley BJ, Rubin EJ, Camilli A, Lampe DJ, Robertson HM, Mekalanos JJ. (1998). Systematic identification of essential genes by in vitro mariner mutagenesis. *Proc Natl Acad Sci U S A* 95(15):8927-8932.

Akhverdian VZ, Khrenova EA, Bogush VG, Gerasimova TV, Kirsanov NB. (1984). Wide distribution of transposable phages in natural Pseudomonas aeruginosa populations. *Genetika* 20(10):1612-1619.

al-Hendy A, Toivanen P, Skurnik M. (1991). Expression cloning of Yersinia enterocolitica O:3 rfb gene cluster in Escherichia coli K12. *Microb Pathog* 10(1):47-59.

Allet B. (1979). Mu insertion duplicates a 5 base pair sequence at the host inserted site. *Cell* 16(1):123-129.

An W, Boeke JD. (2005). Transposon technology and vertebrate functional genomics. *Genome Biol* 6(12):361.

Andersson SG, Alsmark C, Canback B, Davids W, Frank C, Karlberg O*, et al*. (2002). Comparative genomics of microbial pathogens and symbionts. *Bioinformatics* 18 Suppl 2: S17.

Andrake MD, Skalka AM. (1996). Retroviral integrase, putting the pieces together. *J Biol Chem* 271(33):19633-19636.

Aparicio S, Chapman J, Stupka E, Putnam N, Chia JM, Dehal P*, et al*. (2002). Whole-genome shotgun assembly and analysis of the genome of Fugu rubripes. *Science* 297(5585):1301-1310.

Arabidopsis Genome Initiative. (2000). Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature* 408(6814):796-815.

Auerbach MR, Shu C, Kaplan A, Singh IR. (2003). Functional characterization of a portion of the Moloney murine leukemia virus gag gene by genetic footprinting. *Proc Natl Acad Sci U S A* 100(20):11678-11683.

Badarinarayana V, Estep PW,3rd, Shendure J, Edwards J, Tavazoie S, Lam F*, et al*. (2001). Selection analyses of insertional mutants using subgenic-resolution arrays. *Nat Biotechnol* 19(11):1060-1065.

Bailey JA, Carrel L, Chakravarti A, Eichler EE. (2000). Molecular evidence for a relationship between LINE-1 elements and X chromosome inactivation: the Lyon repeat hypothesis. *Proc Natl Acad Sci U S A* 97(12):6634-6639.

Baker TA, Luo L. (1994). Identification of residues in the Mu transposase essential

for catalysis. *Proc Natl Acad Sci U S A* 91(14):6654-6658.

Baker TA, Mizuuchi K. (1992). DNA-promoted assembly of the active tetramer of the Mu transposase. *Genes Dev* 6(11):2221-2232.

Baker TA, Mizuuchi M, Mizuuchi K. (1991). MuB protein allosterically activates strand transfer by the transposase of phage Mu. *Cell* 65(6):1003-1013.

Bamford D, McGraw T, MacKenzie G, Mindich L. (1983). Identification of a protein bound to the termini of bacteriophage PRD1 DNA. *J Virol* 47(2):311-316.

Bamford JK, Bamford DH. (2000). A new mutant class, made by targeted mutagenesis, of phage PRD1 reveals that protein P5 connects the receptor binding protein to the vertex. *J Virol* 74(17):7781-7786.

Bamford JK, Bamford DH. (1990). Capsomer proteins of bacteriophage PRD1, a bacterial virus with a membrane. *Virology* 177(2):445-451.

Bamford JK, Cockburn JJ, Diprose J, Grimes JM, Sutton G, Stuart DI*, et al*. (2002). Diffraction quality crystals of PRD1, a 66-MDa dsDNA virus with an internal membrane. *J Struct Biol* 139(2):103-112.

Bamford JK, Hanninen AL, Pakula TM, Ojala PM, Kalkkinen N, Frilander M*, et al*. (1991). Genome organization of membrane-containing bacteriophage PRD1. *Virology* 183(2):658-676.

Banh Q, Arenskotter M, Steinbuchel A. (2005). Establishment of Tn5096-based transposon mutagenesis in Gordonia polyisoprenivorans. *Appl Environ Microbiol* 71(9):5077-5084.

Bellen HJ, Levis RW, Liao G, He Y, Carlson JW, Tsang G*, et al*. (2004). The BDGP gene disruption project: single transposon insertions associated with 40% of Drosophila genes. *Genetics* 167(2):761-781.

Berg CM, Berg DE (1995) Transposable elements as tools for molecular analyses in bacteria. In: Mobile Genetic Elements. Sherratt DJ, editor. Oxford, UK: Oxford University Press, pp. 38-68.

Biemont C, Vieira C. (2005). What transposable elements tell us about genome organization and evolution: the case of Drosophila. *Cytogenet Genome Res* 110(1-4):25-34.

Biery MC, Stewart FJ, Stellwagen AE, Raleigh EA, Craig NL. (2000). A simple in vitro Tn7-based transposition system with low target site selectivity for genome and gene analysis. *Nucleic Acids Res* 28(5):1067-1077.

Biessmann H, Valgeirsdottir K, Lofsky A, Chin C, Ginther B, Levis RW*, et al*. (1992). HeT-A, a transposable element specifically involved in "healing" broken chromosome ends in Drosophila melanogaster. *Mol Cell Biol* 12(9):3910-3918.

Blattner FR, Plunkett G,3rd, Bloch CA, Perna NT, Burland V, Riley M*, et al*. (1997). The complete genome sequence of Escherichia coli K-12. *Science* 277(5331):1453-1474.

Blot M. (1994). Transposable elements and adaptation of host bacteria. *Genetica* 93(1-3):5-12.

Blot M, Hauer B, Monnet G. (1994). The Tn5 bleomycin resistance gene confers improved survival and growth advantage on Escherichia coli. *Mol Gen Genet* 242(5):595-601.

Boeke JD, Stoye JP (1997) Retrotransposons, Endogenous Retroviruses, and the Evolution of Retroelements. In: Retroviruses. Coffin JM, Hughes SH, Varmus HE, editors. 1st ed. New York, NY, USA: Cold Spring Harbor Laboratory Press, pp. 343-435.

Boeke JD (2002) Putting mobile DNA to work: the toolbox. In: Mobile DNA II. Craig N, Craigie R, Gellert M, Lambowitz AM, editors. 2nd ed. Washington D. C.: ASM Press, pp. 24-37.

Boeke JD, Devine SE. (1998). Yeast retrotransposons: finding a nice quiet neighborhood. *Cell* 93(7):1087-1089.

Bordenstein SR, Reznikoff WS. (2005). Mobile DNA in obligate intracellular bacteria. *Nat Rev Microbiol* 3(9):688-699.

Bottone EJ. (1997). Yersinia enterocolitica: the charisma continues. *Clin Microbiol Rev* 10(2):257-276.

Bradley A, Evans M, Kaufman MH, Robertson E. (1984). Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. *Nature* 309(5965):255-256.

Brandt VL, Roth DB. (2004). V(D)J recombination: how to tame a transposase. *Immunol Rev* 200:249-260.

Brookfield JFY (1995) Transposable elements as selfish DNA. In: Mobile Genetic Elements. Sherratt DJ, editor. Oxford, UK: Oxford University Press, pp. 130-153.

Brugger K, Redder P, She Q, Confalonieri F, Zivanovic Y, Garrett RA. (2002). Mobile elements in archaeal genomes. *FEMS Microbiol Lett* 206(2):131-141.

Brune W, Menard C, Hobom U, Odenbreit S, Messerle M, Koszinowski UH. (1999). Rapid identification of essential and nonessential herpesvirus genes by direct transposon mutagenesis. *Nat Biotechnol* 17(4):360-364.

Butcher SJ, Bamford DH, Fuller SD. (1995). DNA packaging orders the membrane of bacteriophage PRD1. *EMBO J* 14(24):6078-6086.

Butterfield YS, Marra MA, Asano JK, Chan SY, Guin R, Krzywinski MI*, et al*. (2002). An efficient strategy for large-scale high-throughput transposon-mediated sequencing of cDNA clones. *Nucleic Acids Res* 30(11):2460-2468.

Caldentey J, Tuma R, Bamford DH. (2000). Assembly of bacteriophage PRD1 spike complex: role of the multidomain protein P5. *Biochemistry* 39(34):10566-10573.

Campbell A (2002) Eubacterial Genomes. In: Mobile DNA II. Craig NL, Craigie R, Gellert M, Lambowitz AM, editors. Washington, DC, USA: ASM Press, pp. 1024-1039.

Capecchi MR. (2001). Generating mice with targeted mutations. *Nat Med* 7(10):1086-1090.

Capecchi MR. (1989). Altering the genome by homologous recombination. *Science* 244(4910):1288-1292.

Casadaban MJ. (1975). Fusion of the Escherichia coli lac genes to the ara promoter: a general technique using bacteriophage Mu-1 insertions. *Proc Natl Acad Sci U S A* 72(3):809-813.

Casadaban MJ, Cohen SN. (1979). Lactose genes fused to exogenous promoters in one step using a Mu-lac bacteriophage: in vivo probe for transcriptional control sequences. *Proc Natl Acad Sci U S A* 76(9):4530-4533.

Castilho BA, Casadaban MJ. (1991). Specificity of mini-Mu bacteriophage insertions in a small plasmid. *J Bacteriol* 173(3):1339-1343.

Chaconas G, Harshey RM (2002) Transposition of phage Mu DNA. In: Mobile DNA II. Craig NL, Craigie R, Gellert M, Lambowitz AM, editors. Washington DC, USA: ASM Press, pp. 384-402.

Chaconas G, Lavoie BD, Watson MA. (1996). DNA transposition: jumping gene machine, some assembly required. *Curr Biol* 6(7):817-820.

Chaconas G, Harshey RM, Sarvetnick N, Bukhari AI. (1981). Predominant end-products of prophage Mu DNA transposition during the lytic cycle are replicon fusions. *J Mol Biol* 150(3):341-359.

Chalmers RM, Kleckner N. (1994). Tn10/IS10 transposase purification, activation, and in vitro reaction. *J Biol Chem* 269(11):8029-8035.

Chan KY, Baumann L, Garza MM, Baumann P. (1978). Two new species of *Alteromonas*: *Alteromonas espejiana* and *Alteromonas undida*. *Int J Syst Bacteriol* 28:217-222.

Chandler M, Mahillon J (2002) Insertion Sequences Revisited. In: Mobile DNA II. Craig NL, Craigie R, Gellert M, Lambowitz AM, editors. Washington, DC, USA: ASM Press, pp. 305-366.

Chiang SL, Mekalanos JJ, Holden DW. (1999). In vivo genetic analysis of bacterial virulence. *Annu Rev Microbiol* 53:129-154.

Craig NL (2002) Tn*7*. In: Mobile DNA II. Craig NL, Craigie R, Gellert M, Lambowitz AM, editors. Washington, DC, USA: ASM Press, pp. 423-456.

Craig NL. (1996). V(D)J recombination and transposition: closer than expected. *Science* 271(5255):1512.

Craig NL. (1995). Unity in transposition reactions. *Science* 270(5234):253-254.

Craigie R, Mizuuchi K. (1987). Transposition of Mu DNA: joining of Mu to target DNA can be uncoupled from cleavage at the ends of Mu. *Cell* 51(3):493-501.

Craigie R, Mizuuchi K. (1986). Role of DNA topology in Mu transposition: mechanism of sensing the relative orientation of two DNA segments. *Cell* 45(6):793-800.

Craigie R, Mizuuchi K. (1985). Mechanism of transposition of bacteriophage Mu: structure of a transposition intermediate. *Cell* 41(3):867-876.

Craigie R, Arndt-Jovin DJ, Mizuuchi K. (1985). A defined system for the DNA strand-transfer reaction at the initiation of bacteriophage Mu transposition: protein and DNA substrate requirements. *Proc Natl Acad Sci U S A* 82(22):7570-7574.

Craigie R, Mizuuchi M, Mizuuchi K. (1984). Site-specific recognition of the bacteriophage Mu ends by the Mu A protein. *Cell* 39(2 Pt 1):387-394.

Dean RA, Talbot NJ, Ebbole DJ, Farman ML, Mitchell TK, Orbach MJ*, et al*. (2005). The genome sequence of the rice blast fungus Magnaporthe grisea. *Nature* 434(7036):980-986.

Deininger PL, Roy-Engel AM (2002) Mobile Elements in Animal and Plant Genomes. In: Mobile DNA II. Craig NL, Craigie R, Gellert M, Lambowitz AM, editors. Washington, DC, USA: ASM Press, pp. 1074-1092.

Devine SE, Boeke JD. (1994). Efficient integration of artificial transposons into plasmid targets in vitro: a useful tool for DNA mapping, sequencing and genetic analysis. *Nucleic Acids Res* 22(18):3765-3772.

Doolittle WF, Sapienza C. (1980). Selfish genes, the phenotype paradigm and genome evolution. *Nature* 284(5757):601-603.

Doolittle WF, Kirkwood TB, Dempster MA. (1984). Selfish DNAs with self-restraint. *Nature* 307(5951):501-502.

DuBow MS (1987) Transposable Mu-like phages. In: Phage Mu. Symonds N, Toussaint A, van de Putte, Pieter, Howe MM, editors. New York, NY, USA: Cold Spring Harbor Laboratory, pp. 201-213.

Eickbush TH, Malik HS (2002) Origins and Evolution of Retrotransposons. In: Mobile DNA II. Craig NL, Craigie R, Gellert M, Lambowitz AM, editors. Washington, DC, USA: ASM Press, pp. 1111-1144.

Eickbush TH. (1997). Telomerase and retrotransposons: which came first? *Science* 277(5328):911-912.

Engels WR, Johnson-Schlitz DM, Eggleston WB, Sved J. (1990). High-frequency P element loss in Drosophila is homolog dependent. *Cell* 62(3):515-525.

Espejo RT, Canelo ES. (1968a). Properties and characterization of the host bacterium of bacteriophage PM2. *J Bacteriol* 95(5):1887-1891.

Espejo RT, Canelo ES. (1968b). Properties of bacteriophage PM2: a lipid-containing bacterial virus. *Virology* 34(4):738-747.

Evans MJ, Kaufman MH. (1981). Establishment in culture of pluripotential cells from mouse embryos. *Nature* 292(5819):154-156.

Faelen M, Huisman O, Toussaint A. (1978). Involvement of phage Mu-1 early functions in Mu-mediated chromosomal rearrangements. *Nature* 271(5645):580-582.

Feschotte C, Zhang X, Wessler SR (2002) Miniature Inverted-Repeat Transposable Elements and Their Relationship to Established DNA Transposons. In: Mobile DNA II. Craig NL, Craigie R, Gellert M, Lambowitz AM, editors. Washington, DC, USA: ASM Press, pp. 1147-1158.

Fujiwara T, Mizuuchi K. (1988). Retroviral DNA integration: structure of an integration intermediate. *Cell* 54(4):497-504.

Galagan JE, Calvo SE, Cuomo C, Ma LJ, Wortman JR, Batzoglou S*, et al*. (2005). Sequencing of Aspergillus nidulans and comparative analysis with A. fumigatus and A. oryzae. *Nature* 438(7071):1105-1115.

Garcillán-Barcia MP, Bernales I, Mendiola MV, de la Cruz, F. (2002) IS*91* Rolling-Circle Transposition. In: Mobile DNA II. Craig NL, Craigie R, Gellert M, Lambowitz AM, editors. Washington, DC, USA: ASM Press, pp. 891-904.

Gauthier G, Gauthier M, Christen R. (1995). Phylogenetic analysis of the genera Alteromonas, Shewanella, and Moritella using genes coding for small-subunit rRNA sequences and division of the genus Alteromonas into two genera, Alteromonas (emended) and Pseudoalteromonas gen. nov., and proposal of twelve new species combinations. *Int J Syst Bacteriol* 45(4):755-761.

Gellert M. (2002). V(D)J recombination: RAG proteins, repair factors, and regulation. *Annu Rev Biochem* 71:101-132.

Gerdes SY, Scholle MD, D'Souza M, Bernal A, Baev MV, Farrell M*, et al*. (2002). From genetic footprinting to antimicrobial drug targets: examples in cofactor biosynthetic pathways. *J Bacteriol* 184(16):4555-4572.

Gerdes SY, Scholle MD, Campbell JW, Balazsi G, Ravasz E, Daugherty MD*, et al*. (2003). Experimental determination and system level analysis of essential genes in Escherichia coli MG1655. *J Bacteriol* 185(19):5673-5684.

Gibbs RA, Weinstock GM, Metzker ML, Muzny DM, Sodergren EJ, Scherer S*, et al*. (2004). Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* 428(6982):493-521.

Glass JI, Assad-Garcia N, Alperovich N, Yooseph S, Lewis MR, Maruf M*, et al*. (2006). Essential genes of a minimal bacterium. *Proc Natl Acad Sci U S A* 103(2):425-430.

Goldhaber-Gordon I, Williams TL, Baker TA. (2002a). DNA recognition sites activate MuA transposase to perform transposition of non-Mu DNA. *J Biol Chem* 277(10):7694-7702.

Goldhaber-Gordon I, Early MH, Gray MK, Baker TA. (2002b). Sequence and positional requirements for DNA sites in a Mu transpososome. *J Biol Chem* 277(10):7703-7712.

Gonzales AE, Glisson JR, Jackwood MW. (1996). Transposon mutagenesis of Haemophilus paragallinarum with Tn916. *Vet Microbiol* 48(3-4):283-291.

Goryshin IY, Reznikoff WS. (1998). Tn5 in vitro transposition. *J Biol Chem* 273(13):7367-7374.

Goryshin IY, Jendrisak J, Hoffman LM, Meis R, Reznikoff WS. (2000). Insertional transposon mutagenesis by electroporation of released Tn5 transposition complexes. *Nat Biotechnol* 18(1):97-100.

Grahn AM, Caldentey J, Bamford JK, Bamford DH. (1999). Stable packaging of phage PRD1 DNA requires adsorption protein P2, which binds to the IncP plasmid-encoded conjugative transfer complex. *J Bacteriol* 181(21):6689-6696.

Grahn AM, Bamford JK, O'Neill MC, Bamford DH. (1994). Functional organization of the bacteriophage PRD1 genome. *J Bacteriol* 176(10):3062-3068.

Gray HB,Jr, Upholt WB, Vinograd J. (1971). A buoyant method for the determination of the superhelix density of closed circular DNA. *J Mol Biol* 62(1):1-19.

Gray YH. (2000). It takes two transposons to tango: transposable-element-mediated chromosomal rearrangements. *Trends Genet* 16(10):461-468.

Greene EC, Mizuuchi K. (2002a). Direct observation of single MuB polymers: evidence for a DNA-dependent conformational change for generating an active target complex. *Mol Cell* 9(5):1079-1089.

Greene EC, Mizuuchi K. (2002b). Dynamics of a protein polymer: the assembly and disassembly pathways of the MuB transposition target complex. *EMBO J* 21(6):1477-1486.

Greene EC, Mizuuchi K. (2002c). Target Immunity during Mu DNA Transposition. Transpososome Assembly and DNA Looping Enhance MuA-Mediated Disassembly of the MuB Target Complex. *Mol Cell* 10(6):1367-1378.

Griffin TJ,4th, Parsons L, Leschziner AE, DeVost J, Derbyshire KM, Grindley ND. (1999). In vitro transposition of Tn552: a tool for DNA sequencing and mutagenesis. *Nucleic Acids Res* 27(19):3859-3865.

Haapa S, Taira S, Heikkinen E, Savilahti H. (1999a). An efficient and accurate integration of mini-Mu transposons in vitro: a general methodology for functional genetic analysis and molecular biology applications. *Nucleic Acids Res* 27(13):2777-2784.

Haapa S, Suomalainen S, Eerikäinen S, Airaksinen M, Paulin L, Savilahti H. (1999b). An efficient DNA sequencing strategy based on the bacteriophage mu in vitro DNA transposition reaction. *Genome Res* 9(3):308-315.

Haapa-Paananen S, Rita H, Savilahti H. (2002). DNA transposition of bacteriophage Mu. A quantitative analysis of target site selection in vitro. *J Biol Chem* 277(4):2843-2851.

Haas BJ, Wortman JR, Ronning CM, Hannick LI, Smith RK,Jr, Maiti R*, et al*. (2005). Complete reannotation of the Arabidopsis genome: methods, tools, protocols and the final release. *BMC Biol* 3(1):7.

Hallet B, Sherratt DJ, Hayes F. (1997). Pentapeptide scanning mutagenesis: random insertion of a variable five amino acid cassette in a target protein. *Nucleic Acids Res* 25(9):1866-1867.

Hamer L, DeZwaan TM, Montenegro-Chamorro MV, Frank SA, Hamer JE. (2001). Recent advances in large-scale transposon mutagenesis. *Curr Opin Chem Biol* 5(1):67-73.

Hare RS, Walker SS, Dorman TE, Greene JR, Guzman LM, Kenney TJ*, et al*. (2001). Genetic footprinting in bacteria. *J Bacteriol* 183(5):1694-1706.

Haren L, Ton-Hoang B, Chandler M. (1999a). Integrating DNA: transposases and retroviral integrases. *Annu Rev Microbiol* 53:245-281.

Haren L, Ton-Hoang B, Chandler M. (1999b). Integrating DNA: transposases and retroviral integrases. *Annu Rev Microbiol* 53:245-281.

Harshey RM. (1984). Transposition without duplication of infecting bacteriophage Mu DNA. *Nature* 311(5986):580-581.

Harshey RM, Getzoff ED, Baldwin DL, Miller JL, Chaconas G. (1985). Primary structure of phage mu transposase: homology to mu repressor. *Proc Natl Acad Sci U S A* 82(22):7676-7680.

Hasty P, Abuin A, Bradley A (2001) Gene targeting, principles, and practice in mammalian cells. In: Gene Targeting, a Practical Approach. Joyner AL, editor. Oxford, UK: Oxford University Press, pp. 1-35.

Hayes F. (2003). Transposon-based strategies for microbial functional genomics and proteomics. *Annu Rev Genet* 37:3-29.

Hayes F, Hallet B. (2000). Pentapeptide scanning mutagenesis: encouraging old proteins to execute unusual tricks. *Trends Microbiol* 8(12):571-577.

Hensel M, Shea JE, Gleeson C, Jones MD, Dalton E, Holden DW. (1995). Simultaneous identification of bacterial virulence genes by negative selection. *Science* 269(5222):400-403.

Hillier LW, Miller W, Birney E, Warren W, Hardison RC, Ponting CP*, et al*. (2004).

Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 432(7018):695-716.

Hiom K, Melek M, Gellert M. (1998). DNA transposition by the RAG1 and RAG2 proteins: a possible source of oncogenic translocations. *Cell* 94(4):463-470.

Hobom U, Brune W, Messerle M, Hahn G, Koszinowski UH. (2000). Fast screening procedures for random transposon libraries of cloned herpesvirus genomes: mutational analysis of human cytomegalovirus envelope glycoprotein genes. *J Virol* 74(17):7720-7729.

Holt RA, Subramanian GM, Halpern A, Sutton GG, Charlab R, Nusskern DR*, et al*. (2002). The genome sequence of the malaria mosquito Anopheles gambiae. *Science* 298(5591):129-149.

Howe MM (1987) Phage Mu: an overview. In: Phage Mu. Symonds N, Toussaint A, van de Putte, Pieter, Howe MM, editors. New York, NY, USA: Cold Spring Harbor Laboratory, pp. 25-39.

Huiskonen JT, Kivelä HM, Bamford DH, Butcher SJ. (2004). The PM2 virion has a novel organization with an internal membrane and pentameric receptor binding spikes. *Nat Struct Mol Biol* 11(9):850-856.

Hutchison CA, Peterson SN, Gill SR, Cline RT, White O, Fraser CM*, et al*. (1999). Global transposon mutagenesis and a minimal Mycoplasma genome. *Science* 286(5447):2165-2169.

Ivics Z, Hackett PB, Plasterk RH, Izsvak Z. (1997). Molecular reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells. *Cell* 91(4):501-510.

Izsvak Z, Ivics Z. (2004). Sleeping beauty transposition: biology and applications for molecular therapy. *Mol Ther* 9(2):147-156.

Jenkins FJ, Casadaban MJ, Roizman B. (1985). Application of the mini-Mu-phage for target-sequence-specific insertional mutagenesis of the herpes simplex virus genome. *Proc Natl Acad Sci U S A* 82(14):4773-4777.

Jenks PJ, Chevalier C, Ecobichon C, Labigne A. (2001). Identification of nonessential Helicobacter pylori genes using random mutagenesis and loop amplification. *Res Microbiol* 152(8):725-734.

Judson N, Mekalanos JJ. (2000). Transposon-based approaches to identify essential bacterial genes. *Trends Microbiol* 8(11):521-526.

Kahmann R, Kamp D. (1979). Nucleotide sequences of the attachment sites of bacteriophage Mu DNA. *Nature* 280(5719):247-250.

Kaiser K, Sentry JW, Finnegan DJ (1995) Eukaryotic transposable elements as tools to study gene structure and function. In: Mobile Genetic Elements. Sherratt DJ, editor. Oxford, UK: Oxford University Press, pp. 69-100.

Kaminker JS, Bergman CM, Kronmiller B, Carlson J, Svirskas R, Patel S*, et al*. (2002). The transposable elements of the Drosophila melanogaster euchromatin: a genomics perspective. *Genome Biol* 3(12): RESEARCH0084.

Kapitonov VV, Jurka J. (2001). Rolling-circle transposons in eukaryotes. *Proc Natl Acad Sci U S A* 98(15):8714-8719.

Kekarainen T, Savilahti H, Valkonen JP. (2002). Functional genomics on potato virus A: virus genome-wide map of sites essential for virus propagation. *Genome Res* 12(4):584-594.

Kempken F, Kuck U. (1998). Transposons in filamentous fungi--facts and perspectives. *Bioessays* 20(8):652-659.

Kidwell MG. (2002). Transposable elements and the evolution of genome size in eukaryotes. *Genetica* 115(1):49-63.

Kim C, Rubin CM, Schmid CW. (2001). Genome-wide chromatin remodeling modulates the Alu heat shock response. *Gene* 276(1-2):127-133.

Kim JM, Vanguri S, Boeke JD, Gabriel A, Voytas DF. (1998). Transposable elements and genome organization: a comprehensive survey of retrotransposons revealed by the complete Saccharomyces cerevisiae genome sequence. *Genome Res* 8(5):464-478.

Kim K, Harshey RM. (1995). Mutational analysis of the att DNA-binding domain of phage Mu transposase. *Nucleic Acids Res* 23(19):3937-3943.

Kivelä HM, Männistö RH, Kalkkinen N, Bamford DH. (1999). Purification and protein composition of PM2, the first lipid-containing bacterial virus to be isolated. *Virology* 262(2):364-374.

Kivelä HM, Daugelavicius R, Hankkio RH, Bamford JK, Bamford DH. (2004). Penetration ofmembrane-containingdouble-stranded-DNA bacteriophage PM2 into Pseudoalteromonas hosts. *J Bacteriol* 186(16):5342-5354.

Kleckner N. (1990). Regulation of transposition in bacteria. *Annu Rev Cell Biol* 6:297-327.

Kleckner N, Barker DF, Ross DG, Botstein D. (1978). Properties of the translocatable tetracycline-resistance element Tn10 in Escherichia coli and bacteriophage lambda. *Genetics* 90(3):427-461.

Kleckner N, Chalmers RM, Kwon D, Sakai J, Bolland S. (1996). Tn10 and IS10 transposition and chromosome rearrangements: mechanism and regulation in vivo and in vitro. *Curr Top Microbiol Immunol* 204:49-82.

Kobryn K, Watson MA, Allison RG, Chaconas G. (2002). The Mu three-site synapse: a strained assembly platform in which delivery of the L1 transposase binding site triggers catalytic commitment. *Mol Cell* 10(3):659-669.

Koes R, Souer E, van Houwelingen A, Mur L, Spelt C, Quattrocchio F*, et al*. (1995). Targeted gene inactivation in petunia by PCR-based selection of transposon insertion mutants. *Proc Natl Acad Sci U S A* 92(18):8149-8153.

Koonin EV, Mushegian AR, Ryabov EV, Dolja VV. (1991). Diverse groups of plant RNA and DNA viruses share related movement proteins that may possess chaperone-like activity. *J Gen Virol* 72 (Pt 12)(Pt 12):2895-2903.

Krementsova E, Giffin MJ, Pincus D, Baker TA. (1998). Mutational analysis of the Mu transposase. Contributions of two distinct regions of domain II to recombination. *J Biol Chem* 273(47):31358-31365.

Kruklitis R, Nakai H. (1994). Participation of the bacteriophage Mu A protein and host factors in the initiation of Mu DNA synthesis in vitro. *J Biol Chem* 269(23):16469-16477.

Kruklitis R, Welty DJ, Nakai H. (1996). ClpX protein of Escherichia coli activates bacteriophage Mu transposase in the strand transfer complex for initiation of Mu DNA synthesis. *EMBO J* 15(4):935-944.

Kunst F, Ogasawara N, Moszer I, Albertini AM, Alloni G, Azevedo V*, et al*. (1997). The complete genome sequence of the gram-positive bacterium Bacillus subtilis. *Nature* 390(6657):249-256.

Kuo CF, Zou AH, Jayaram M, Getzoff E, Harshey R. (1991). DNA-protein complexes during attachment-site synapsis in Mu DNA transposition. *EMBO J* 10(6):1585-1591.

Lamberg A, Nieminen S, Qiao M, Savilahti H. (2002). Efficient insertion mutagenesis strategy for bacterial genomes involving electroporation of in vitro-assembled DNA transposition complexes of bacteriophage mu. *Appl Environ Microbiol* 68(2):705-712.

Lampson BC, Inouye M, Inouye S. (2005). Retrons, msDNA, and the bacterial genome. *Cytogenet Genome Res* 110(1-4):491-499.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J*, et al*. (2001). Initial sequencing and analysis of the human genome. *Nature* 409(6822):860-921.

Landree MA, Wibbenmeyer JA, Roth DB. (1999). Mutational analysis of RAG1 and RAG2 identifies three catalytic amino acids in RAG1 critical for both cleavage steps of V(D)J recombination. *Genes Dev* 13(23):3059-3069.

Laurent LC, Olsen MN, Crowley RA, Savilahti H, Brown PO. (2000). Functional characterization of the human immunodeficiency virus type 1 genome by genetic footprinting. *J Virol* 74(6):2760-2769.

Lavoie BD, Chan BS, Allison RG, Chaconas G. (1991). Structural aspects of a higher order nucleoprotein complex: induction of an altered DNA structure at the Mu-host junction of the Mu type 1 transpososome. *EMBO J* 10(10):3051-3059.

Lee I, Harshey RM. (2003). The conserved CA/TG motif at Mu termini: T specifies stable transpososome assembly. *J Mol Biol* 330(2):261-275.

Lee I, Harshey RM. (2001). Importance of the conserved CA dinucleotide at Mu termini. *J Mol Biol* 314(3):433-444.

Lee M, Xiao J, Haghjoo E, Zhan X, Abenes G, Tuong T*, et al*. (2000). Murine cytomegalovirus containing a mutation at open reading frame M37 is severely attenuated in growth and virulence in vivo. *J Virol* 74(23):11099-11107.

Lee MD, Henk AD. (1996). Tn10 insertional mutagenesis in Pasteurella multocida. *Vet Microbiol* 50(1-2):143-148.

Leung PC, Teplow DB, Harshey RM. (1989). Interaction of distinct domains in Mu transposase with Mu DNA ends and an internal transpositional enhancer. *Nature* 338(6217):656-658.

Levchenko I, Yamauchi M, Baker TA. (1997). ClpX and MuB interact with overlapping regions of Mu transposase: implications for control of the transposition pathway. *Genes Dev* 11(12):1561-1572.

Levchenko I, Luo L, Baker TA. (1995). Disassembly of the Mu transposase tetramer by the ClpX chaperone. *Genes Dev* 9(19):2399-2408.

Levis RW, Ganesan R, Houtchens K, Tolar LA, Sheen FM. (1993). Transposons in place of telomeric repeats at a Drosophila telomere. *Cell* 75(6):1083-1093.

Lewis SM, Wu GE. (1997). The origins of V(D)J recombination. *Cell* 88(2):159-162.

Li T, Spearow J, Rubin CM, Schmid CW. (1999). Physiological stresses increase mouse short interspersed element (SINE) RNA expression in vivo. *Gene* 239(2):367-372.

Li W, Zhang P, Fellers JP, Friebe B, Gill BS. (2004). Sequence composition, organization, and evolution of the core Triticeae genome. *Plant J* 40(4):500-511.

Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DB, Kamal M*, et al*. (2005). Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* 438(7069):803-819.

Lingner J, Hughes TR, Shevchenko A, Mann M, Lundblad V, Cech TR. (1997). Reverse transcriptase motifs in the catalytic subunit of telomerase. *Science* 276(5312):561-567.

Luan DD, Korman MH, Jakubczak JL, Eickbush TH. (1993). Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell* 72(4):595-605.

Lyon MF. (1998). X-chromosome inactivation: a repeat hypothesis. *Cytogenet Cell Genet* 80(1-4):133-137.

Lyra C, Savilahti H, Bamford DH. (1991). High-frequency transfer of linear DNA containing 5'-covalently linked terminal proteins: electroporation of bacteriophage PRD1 genome into Escherichia coli. *Mol Gen Genet* 228(1-2):65-69.

Mahillon J, Chandler M. (1998). Insertion sequences. *Microbiol Mol Biol Rev* 62(3):725-774.

Manna D, Higgins NP. (1999). Phage Mu transposition immunity reflects supercoil domain structure of the chromosome. *Mol Microbiol* 32(3):595-606.

Manna D, Wang X, Higgins NP. (2001). Mu and IS1 transpositions exhibit strong orientation bias at the Escherichia coli bgl locus. *J Bacteriol* 183(11):3328-3335.

Männistö RH, Grahn AM, Bamford DH, Bamford JK. (2003). Transcription of bacteriophage PM2 involves phage-encoded regulators of heterologous origin. *J Bacteriol* 185(11):3278-3287.

Männistö RH, Kivelä HM, Paulin L, Bamford DH, Bamford JK. (1999). The complete genome sequence of PM2, the first lipid-containing bacterial virus To Be isolated. *Virology* 262(2):355-363.

Martienssen RA. (1998). Functional genomics: probing plant gene function and expression with transposons. *Proc Natl Acad Sci U S A* 95(5):2021-2026.

Maxwell A, Craigie R, Mizuuchi K. (1987). B protein of bacteriophage mu is an ATPase that preferentially stimulates intermolecular DNA strand transfer. *Proc Natl Acad Sci U S A* 84(3):699-703.

McCarren J, Brahamsha B. (2005). Transposon mutagenesis in a marine synechococcus strain: isolation of swimming motility mutants. *J Bacteriol* 187(13):4457-4462.

Mendiola MV, de la Cruz F. (1992). IS91 transposase is related to the rolling-circle-type replication proteins of the pUB110 family of plasmids. *Nucleic Acids Res* 20(13):3521.

Mendiola MV, Bernales I, de la Cruz F. (1994). Differential roles of the transposon termini in IS91 transposition. *Proc Natl Acad Sci U S A* 91(5):1922-1926.

Messing J, Gronenborn B, Muller-Hill B, Hans Hopschneider P. (1977). Filamentous coliphage M13 as a cloning vehicle: insertion of a HindII fragment of the lac regulatory region in M13 replicative form in vitro. *Proc Natl Acad Sci U S A* 74(9):3642-3646.

Meyers EN, Lewandoski M, Martin GR. (1998). An Fgf8 mutant allelic series generated by Cre- and Flp-mediated recombination. *Nat Genet* 18(2):136-141.

Mindich L, Bamford D, Goldthwaite C, Laverty M, Mackenzie G. (1982). Isolation of nonsense mutants of lipid-containing bacteriophage PRD1. *J Virol* 44(3):1013-1020.

Mise K. (1971). Isolation and characterization of a new generalized transducing bacteriophage different from P1 in Escherichia coli. *J Virol* 7(1):168-175.

Miskey C, Izsvak Z, Kawakami K, Ivics Z. (2005). DNA transposons in vertebrate functional genomics. *Cell Mol Life Sci* 62(6):629-641.

Mizuuchi K. (1992). Transpositional recombination: mechanistic insights from studies of Mu and other elements. *Annu Rev Biochem* 61:1011-1051.

Mizuuchi K. (1983). In vitro transposition of bacteriophage Mu: a biochemical approach to a novel replication reaction. *Cell* 35(3 Pt 2):785-794.

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

Mizuuchi M, Mizuuchi K. (1993). Target site selection in transposition of phage Mu. *Cold Spring Harb Symp Quant Biol* 58:515-523.

Mizuuchi M, Mizuuchi K. (1989). Efficient Mu transposition requires interaction of transposase with a DNA sequence at the Mu operator: implications for regulation. *Cell* 58(2):399-408.

Mizuuchi M, Baker TA, Mizuuchi K. (1992). Assembly of the active form of the transposase-Mu DNA complex: a critical control point in Mu transposition. *Cell* 70(2):303-311.

Moorman NJ, Lin CY, Speck SH. (2004). Identification of candidate gammaherpesvirus 68 genes required for virus replication by signature-tagged transposon mutagenesis. *J Virol* 78(19):10282-10290.

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(3):337-359.

Morita M, Umemoto A, Li ZX, Nakazono N, Sugino Y. (1996). Nested deletions from a fixed site as an aid to nucleotide sequencing: an in vitro system using Tn3 transposase. *DNA Res* 3(6):431-433.

Nagy A, Rossant J. (1996). Targeted mutagenesis: analysis of phenotype without germ line transmission. *J Clin Invest* 97(6):1360-1365.

Nagy A, Rossant J, Nagy R, Abramow-Newerly W, Roder JC. (1993). Derivation of completely cell culture-derived mice from early-passage embryonic stem cells. *Proc Natl Acad Sci U S A* 90(18):8424-8428.

Nagy A, Moens C, Ivanyi E, Pawling J, Gertsenstein M, Hadjantonakis AK*, et al*. (1998). Dissecting the role of N-myc in development using a single targeting vector to generate a series of alleles. *Curr Biol* 8(11):661-664.

Naigamwalla DZ, Chaconas G. (1997). A new set of Mu DNA transposition intermediates: alternate pathways of target capture preceding strand transfer. *EMBO J* 16(17):5227-5234.

Naigamwalla DZ, Coros CJ, Wu Z, Chaconas G. (1998). Mutations in domain III alpha of the Mu transposase: evidence suggesting an active site component which interacts with the Mu-host junction. *J Mol Biol* 282(2):265-274.

Nakai H, Kruklitis R. (1995). Disassembly of the bacteriophage Mu transposase for the initiation of Mu DNA replication. *J Biol Chem* 270(33):19591-19598.

Nakayama C, Teplow DB, Harshey RM. (1987). Structural domains in phage Mu transposase: identification of the site-specific DNA-binding domain. *Proc Natl Acad Sci U S A* 84(7):1809-1813.

Namgoong SY, Harshey RM. (1998). The same two monomers within a MuA tetramer provide the DDE domains for the strand cleavage and strand transfer steps of transposition. *EMBO J* 17(13):3775-3785.

Namgoong SY, Jayaram M, Kim K, Harshey RM. (1994). DNA-protein cooperativity in the assembly and stabilization of mu strand transfer complex. Relevance of DNA phasing and att site cleavage. *J Mol Biol* 238(4):514-527.

Namgoong SY, Kim K, Saxena P, Yang JY, Jayaram M, Giedroc DP*, et al*. (1998). Mutational analysis of domain II beta of bacteriophage Mu transposase: domains II alpha and II beta belong to different catalytic complementation groups. *J Mol Biol* 275(2):221-232.

Nierman WC, DeShazer D, Kim HS, Tettelin H, Nelson KE, Feldblyum T*, et al*. (2004). Structural flexibility in the Burkholderia mallei genome. *Proc Natl Acad Sci U S A* 101(39):14246-14251.

Oettinger MA. (2004). Molecular biology: hairpins at split ends in DNA. *Nature* 432(7020):960-961.

Oettinger MA, Schatz DG, Gorka C, Baltimore D. (1990). RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination. *Science* 248(4962):1517-1523.

Olsen RH, Siak JS, Gray RH. (1974). Characteristics of PRD1, a plasmid-dependent broad host range DNA bacteriophage. *J Virol* 14(3):689-699.

Orgel LE, Crick FH. (1980). Selfish DNA: the ultimate parasite. *Nature* 284(5757):604-607.

Pajunen M, Kiljunen S, Skurnik M. (2000). Bacteriophage phiYeO3-12, specific for Yersinia enterocolitica serotype O:3, is related to coliphages T3 and T7. *J Bacteriol* 182(18):5114-5120.

Pajunen MI, Kiljunen SJ, Söderholm ME, Skurnik M. (2001). Complete genomic sequence of the lytic bacteriophage phiYeO3-12 of Yersinia enterocolitica serotype O:3. *J Bacteriol* 183(6):1928-1937.

Pajunen MI, Pulliainen AT, Finne J, Savilahti H. (2005). Generation of transposon insertion mutant libraries for Gram-positive bacteria by electroporation of phage Mu DNA transposition complexes. *Microbiology* 151(Pt 4):1209-1218.

Pakula TM, Caldentey J, Gutierrez C, Olkkonen VM, Salas M, Bamford DH. (1993). Overproduction, purification, and characterization of DNA-binding protein P19 of bacteriophage PRD1. *Gene* 126(1):99-104.

Pardue ML, DeBaryshe PG. (1999). Telomeres and telomerase: more than the end of the line. *Chromosoma* 108(2):73-82.

Parinov S, Sundaresan V. (2000). Functional genomics in Arabidopsis: large-scale insertional mutagenesis complements the genome sequencing project. *Curr Opin Biotechnol* 11(2):157-161.

Partanen J, Schwartz L, Rossant J. (1998). Opposite phenotypes of hypomorphic and Y766 phosphorylation site mutations reveal a function for Fgfr1 in anteroposterior patterning of mouse embryos. *Genes Dev* 12(15):2332-2344.

Pathania S, Jayaram M, Harshey RM. (2003). A unique right end-enhancer complex precedes synapsis of Mu ends: the enhancer is sequestered within the transpososome throughout transposition. *EMBO J* 22(14):3725-3736.

Payne JA, Stevenson TJ, Donaldson LF. (1996). Molecular characterization of a putative K-Cl cotransporter in rat brain. A neuronal-specific isoform. *J Biol Chem* 271(27):16245-16252.

Peters JE, Craig NL. (2001). Tn7: smarter than we thought. *Nat Rev Mol Cell Biol* 2(11):806-814.

Plasterk RH (1995) Mechanisms of DNA transposition. In: Mobile Genetic Elements. Sherratt DJ, editor. Oxford, UK: Oxford University Press, pp. 18-37.

Plasterk RH, Izsvak Z, Ivics Z. (1999). Resident aliens: the Tc1/mariner superfamily of transposable elements. *Trends Genet* 15(8):326-332.

Polard P, Chandler M. (1995). Bacterial transposases and retroviral integrases. *Mol Microbiol* 15(1):13-23.

Poussu E, Jäntti J, Savilahti H. (2005). A gene truncation strategy generating N- and C-terminal deletion variants of proteins for functional studies: mapping of the Sec1p binding domain in yeast Mso1p by a Mu in vitro transposition-based approach. *Nucleic Acids Res* 33(12):e104.

Poussu E, Vihinen M, Paulin L, Savilahti H. (2004). Probing the alpha-complementing domain of E. coli beta-galactosidase with use of an insertional pentapeptide mutagenesis strategy based on Mu in vitro DNA transposition. *Proteins* 54(4):681-692.

Ray DS, Kook K. (1978). Insertion of the Tn3 transposon into the genome of the single-stranded DNA phage M13. *Gene* 4(2):109-119.

Reich KA, Chovan L, Hessler P. (1999). Genome scanning in Haemophilus influenzae for identification of essential genes. *J Bacteriol* 181(16):4961-4968.

Rey MW, Ramaiya P, Nelson BA, Brody-Karpin SD, Zaretsky EJ, Tang M*, et al*. (2004). Complete genome sequence of the industrial bacterium Bacillus licheniformis and comparisons with closely related Bacillus species. *Genome Biol* 5(10):R77.

Reznikoff WS. (2003). Tn5 as a model for understanding DNA transposition. *Mol Microbiol* 47(5):1199-1206.

Reznikoff WS, Goryshin IY, Jendrisak JJ. (2004). Tn5 as a molecular genetics tool: In vitro transposition and the coupling of in vitro technologies with in vivo transposition. *Methods Mol Biol* 260:83-96.

Rice P, Mizuuchi K. (1995). Structure of the bacteriophage Mu transposase core: a common structural motif for DNA transposition and retroviral integration. *Cell* 82(2):209-220.

Rivera C, Voipio J, Payne JA, Ruusuvuori E, Lahtinen H, Lämsä K*, et al*. (1999). The K+/Cl- co-transporter KCC2 renders GABA hyperpolarizing during neuronal maturation. *Nature* 397(6716):251-255.

Rokyta D, Badgett MR, Molineux IJ, Bull JJ. (2002). Experimental genomic evolution: extensive compensation for loss of DNA ligase activity in a virus. *Mol Biol Evol* 19(3):230-238.

Rossant J, McKerlie C. (2001). Mouse-based phenogenomics for modelling human disease. *Trends Mol Med* 7(11):502-507.

Ross-Macdonald P, Coelho PS, Roemer T, Agarwal S, Kumar A, Jansen R*, et al*. (1999). Large-scale analysis of the yeast genome by transposon tagging and gene disruption. *Nature* 402(6760):413-418.

Roth DB, Craig NL. (1998). VDJ recombination: a transposase goes to work. *Cell* 94(4):411-414.

Rothenberg SM, Olsen MN, Laurent LC, Crowley RA, Brown PO. (2001). Comprehensive mutational analysis of the Moloney murine leukemia virus envelope protein. *J Virol* 75(23):11851-11862.

Ryder E, Russell S. (2003). Transposable elements as tools for genomics and genetics in Drosophila. *Brief Funct Genomic Proteomic* 2(1):57-71.

Rydman PS, Bamford DH. (2003). Identification and mutational analysis of bacteriophage PRD1 holin protein P35. *J Bacteriol* 185(13):3795-3803.

Rydman PS, Caldentey J, Butcher SJ, Fuller SD, Rutten T, Bamford DH. (1999). Bacteriophage PRD1 contains a labile receptor-binding structure at each vertex. *J Mol Biol* 291(3):575-587.

Sands AT. (2003). The master mammal. *Nat Biotechnol* 21(1):31-32.

Saren AM, Ravantti JJ, Benson SD, Burnett RM, Paulin L, Bamford DH*, et al*. (2005). A snapshot of viral evolution from genome analysis of the tectiviridae family. *J Mol Biol* 350(3):427-440.

Sassetti CM, Boyd DH, Rubin EJ. (2003). Genes required for mycobacterial growth defined by high density mutagenesis. *Mol Microbiol* 48(1):77-84.

Sassetti CM, Boyd DH, Rubin EJ. (2001). Comprehensive identification of conditionally essential genes in mycobacteria. *Proc Natl Acad Sci U S A* 98(22):12712-12717.

Sauer B. (1998). Inducible gene targeting in mice using the Cre/lox system. *Methods* 14(4):381-392.

Savilahti H, Mizuuchi K. (1996). Mu transpositional recombination: donor DNA cleavage and strand transfer in trans by the Mu transposase. *Cell* 85(2):271-280.

Savilahti H, Rice PA, Mizuuchi K. (1995). The phage Mu transpososome core: DNA requirements for assembly and function. *EMBO J* 14(19):4893-4903.

Savilahti H, Caldentey J, Lundstrom K, Syväoja JE, Bamford DH. (1991). Overexpression, purification, and characterization of Escherichia coli bacteriophage PRD1 DNA polymerase. In vitro synthesis of full-length PRD1 DNA with purified proteins. *J Biol Chem* 266(28):18737-18744.

Sawyer SA, Dykhuizen DE, DuBose RF, Green L, Mutangadura-Mhlanga T, Wolczyk DF*, et al*. (1987). Distribution and abundance of insertion sequences among natural isolates of Escherichia coli. *Genetics* 115(1):51-63.

Schena M, Shalon D, Davis RW, Brown PO. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270(5235):467-470.

Schena M, Shalon D, Heller R, Chai A, Brown PO, Davis RW. (1996). Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. *Proc Natl Acad Sci U S A* 93(20):10614-10619.

Shapiro JA (1995) The discovery and significance of mobile genetic elements. In: Mobile Genetic Elements. Sherratt DJ, editor. Oxford, UK: Oxford University Press, pp. 1-17.

She Q, Singh RK, Confalonieri F, Zivanovic Y, Allard G, Awayez MJ*, et al*. (2001). The complete genome of the crenarchaeon Sulfolobus solfataricus P2. *Proc Natl Acad Sci U S A* 98(14):7835-7840.

Singh IR, Crowley RA, Brown PO. (1997). High-resolution functional mapping of a cloned gene by genetic footprinting. *Proc Natl Acad Sci U S A* 94(4):1304-1309.

Smith DR, Doucette-Stamm LA, Deloughery C, Lee H, Dubois J, Aldredge T*, et al*.

(1997). Complete genome sequence of Methanobacterium thermoautotrophicum deltaH: functional analysis and comparative genomics. *J Bacteriol* 179(22):7135-7155.

Smith GA, Enquist LW. (1999). Construction and transposon mutagenesis in Escherichia coli of a full-length infectious clone of pseudorabies virus, an alphaherpesvirus. *J Virol* 73(8):6405-6414.

Smith V, Botstein D, Brown PO. (1995). Genetic footprinting: a genomic strategy for determining a gene's function given its sequence. *Proc Natl Acad Sci U S A* 92(14):6479-6483.

Smith V, Chou KN, Lashkari D, Botstein D, Brown PO. (1996). Functional analysis of the genes of yeast chromosome V by genetic footprinting. *Science* 274(5295):2069-2074.

Sokolsky TD, Baker TA. (2003). DNA gyrase requirements distinguish the alternate pathways of Mu transposition. *Mol Microbiol* 47(2):397-409.

Song MJ, Hwang S, Wong WH, Wu TT, Lee S, Liao HI*, et al*. (2005). Identification of viral genes essential for replication of murine gamma-herpesvirus 68 using signature-tagged mutagenesis. *Proc Natl Acad Sci U S A* 102(10):3805-3810.

Spradling AC, Rubin GM. (1982). Transposition of cloned P elements into Drosophila germ line chromosomes. *Science* 218(4570):341-347.

Spradling AC, Stern D, Beaton A, Rhem EJ, Laverty T, Mozden N*, et al*. (1999). The Berkeley Drosophila Genome Project gene disruption project: Single P-element insertions mutating 25% of vital Drosophila genes. *Genetics* 153(1):135-177.

Strömsten NJ, Bamford DH, Bamford JK. (2005). In vitro DNA packaging of PRD1: a common mechanism for internal-membrane viruses. *J Mol Biol* 348(3):617-629.

Surette MG, Chaconas G. (1989). A protein factor which reduces the negative supercoiling requirement in the Mu DNA strand transfer reaction is Escherichia coli integration host factor. *J Biol Chem* 264(5):3028-3034.

Surette MG, Buch SJ, Chaconas G. (1987). Transpososomes: stable protein-DNA complexes involved in the in vitro transposition of bacteriophage Mu DNA. *Cell* 49(2):253-262.

Taira S, Tuimala J, Roine E, Nurmiaho-Lassila EL, Savilahti H, Romantschuk M. (1999). Mutational analysis of the Pseudomonas syringae pv. tomato hrpA gene encoding Hrp pilus subunit. *Mol Microbiol* 34(4):737-744.

Takami H, Nakasone K, Takaki Y, Maeno G, Sasaki R, Masui N*, et al*. (2000). Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and genomic sequence comparison with Bacillus subtilis. *Nucleic Acids Res* 28(21):4317-4331.

Tavakoli NP, Derbyshire KM. (2001). Tipping the balance between replicative and simple transposition. *EMBO J* 20(11):2923-2930.

Taylor AL. (1963). Bacteriophage-induced mutation in Escherichia coli. *Proc Natl Acad Sci U S A* 50:1043-1051.

Thomas KR, Capecchi MR. (1987). Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell* 51(3):503-512.

Thompson CB. (1995). New insights into V(D)J recombination and its role in the evolution of the immune system. *Immunity* 3(5):531-539.

Thompson S, Clarke AR, Pow AM, Hooper ML, Melton DW. (1989). Germ line transmission and expression of a corrected HPRT gene produced by gene targeting in embryonic stem cells. *Cell* 56(2):313-321.

Tomcsanyi T, Berg CM, Phadnis SH, Berg DE. (1990). Intramolecular transposition by a synthetic IS50 (Tn5) derivative. *J Bacteriol* 172(11):6348-6354.

Tornberg J, Voikar V, Savilahti H, Rauvala H, Airaksinen MS. (2005). Behavioural phenotypes of hypomorphic KCC2-deficient mice. *Eur J Neurosci* 21(5):1327-1337.

Torres RM, Kühn R (1997) Laboratory protocols for conditional gene targeting. Oxford, UK: Oxford University Press.

Tosi LR, Beverley SM. (2000). cis and trans factors affecting Mos1 mariner evolution and transposition in vitro, and its potential for functional genomics. *Nucleic Acids Res* 28(3):784-790.

Turcotte K, Srinivasan S, Bureau T. (2001). Survey of transposable elements from rice genomic sequences. *Plant J* 25(2):169-179.

Turlan C, Chandler M. (2000). Playing second fiddle: second-strand processing and liberation of transposable elements from donor DNA. *Trends Microbiol* 8(6):268-274.

van der Schans GP, Weyermans JP, Bleichrodt JF. (1971). Infection of spheroplasts of Pseudomonas with DNA of bacteriophage PM2. *Mol Gen Genet* 110(3):263-271.

van Gent DC, Mizuuchi K, Gellert M. (1996). Similarities between initiation of V(D)J recombination and retroviral integration. *Science* 271(5255):1592-1594.

Vieira J, Messing J. (1982). The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. *Gene* 19(3):259-268.

Walbot V. (2000). Saturation mutagenesis using maize transposons. *Curr Opin Plant Biol* 3(2):103-107.

Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P*, et al*. (2002). Initial sequencing and comparative analysis of the mouse genome. *Nature* 420(6915):520-562.

Watson MA, Chaconas G. (1996). Three-site synapsis during Mu DNA transposition: a critical intermediate preceding engagement of the active site. *Cell* 85(3):435-445.

Weber PC, Levine M, Glorioso JC. (1987). Rapid identification of nonessential genes of herpes simplex virus type 1 by Tn5 mutagenesis. *Science* 236(4801):576-579.

Westphal CH, Leder P. (1997). Transposon-generated 'knock-out' and 'knock-in' gene-targeting constructs for use in mice. *Curr Biol* 7(7):530-533.

Williams TL, Jackson EL, Carritte A, Baker TA. (1999). Organization and dynamics of the Mu transpososome: recombination by communication between two active sites. *Genes Dev* 13(20):2725-2737.

Wong SM, Mekalanos JJ. (2000). Genetic footprinting with mariner-based transposition in Pseudomonas aeruginosa. *Proc Natl Acad Sci U S A* 97(18):10191-10196.

Wood V, Gwilliam R, Rajandream MA, Lyne M, Lyne R, Stewart A*, et al*. (2002). The genome sequence of Schizosaccharomyces pombe. *Nature* 415(6874):871-880.

Xiao J, Tong T, Zhan X, Haghjoo E, Liu F. (2000). In vitro and in vivo characterization of a murine cytomegalovirus with a transposon insertional mutation at open reading frame M43. *J Virol* 74(20):9488-9497.

Yamauchi M, Baker TA. (1998). An ATP-ADP switch in MuB controls progression of the Mu transposition pathway. *EMBO J* 17(18):5509-5518.

York D, Welch K, Goryshin IY, Reznikoff WS. (1998). Simple and efficient generation in vitro of nested deletions and inversions: Tn5 intramolecular transposition. *Nucleic Acids Res* 26(8):1927-1933.

Yu D, Silva MC, Shenk T. (2003). Functional map of human cytomegalovirus AD169 defined by global mutational analysis. *Proc Natl Acad Sci U S A* 100(21):12396-12401.

Yu J, Wang J, Lin W, Li S, Li H, Zhou J*, et al*. (2005). The Genomes of Oryza sativa: a history of duplications. *PLoS Biol* 3(2):e38.

Yu J, Hu S, Wang J, Wong GK, Li S, Liu B*, et al*. (2002). A draft sequence of the rice genome (Oryza sativa L. ssp. indica). *Science* 296(5565):79-92.

Yuan JF, Beniac DR, Chaconas G, Ottensmeyer FP. (2005). 3D reconstruction of the Mu transposase and the Type 1 transpososome: a structural framework for Mu DNA transposition. *Genes Dev* 19(7):840-852.

Zhan X, Lee M, Xiao J, Liu F. (2000a). Construction and characterization of murine cytomegaloviruses that contain transposon insertions at open reading frames m09 and M83. *J Virol* 74(16):7411-7421.

Zhan X, Lee M, Abenes G, Von Reis I, Kittinunvorakoon C, Ross-Macdonald P*, et al*. (2000b). Mutagenesis of murine cytomegalovirus using a Tn3-based transposon. *Virology* 266(2):264-274.

Zhou L, Mitra R, Atkinson PW, Hickman AB, Dyda F, Craig NL. (2004). Transposition of hAT elements links transposable elements and V(D)J recombination. *Nature* 432(7020):995-1001.

Zimmerly S, Guo H, Perlman PS, Lambowitz AM. (1995). Group II intron mobility occurs by target DNA-primed reverse transcription. *Cell* 82(4):545-554.

Zwaal RR, Broeks A, van Meurs J, Groenen JT, Plasterk RH. (1993). Target-selected gene inactivation in Caenorhabditis elegans by using a frozen transposon insertion mutant bank. *Proc Natl Acad Sci U S A* 90(16):7431-7435.