Microbiology Spectrum
American Society for Microbiology Press

Everyman's Guide to Bacterial Insertion Sequences

PATRICIA SIGUIER,[1] EDITH GOURBEYRE,[1] ALESSANDRO VARANI,[2] BAO TON-HOANG,[1] and MICK CHANDLER[1]

[1]Laboratoire de Microbiologie et Génétique Moléculaires, CNRS, Toulouse, France; [2]Departamento de Tecnologia, Faculdade de Ciências Agrárias e Veterinárias de Jaboticabal, UNESP - Univ. Estadual Paulista, Jaboticabal, SP, Brazil

ABSTRACT The number and diversity of known prokaryotic insertion sequences (IS) have increased enormously since their discovery in the late 1960s. At present the sequences of more than 4000 different IS have been deposited in the specialized ISfinder database. Over time it has become increasingly apparent that they are important actors in the evolution of their host genomes and are involved in sequestering, transmitting, mutating and activating genes, and in the rearrangement of both plasmids and chromosomes. This review presents an overview of our current understanding of these transposable elements (TE), their organization and their transposition mechanism as well as their distribution and genomic impact. In spite of their diversity, they share only a very limited number of transposition mechanisms which we outline here. Prokaryotic IS are but one example of a variety of diverse TE which are being revealed due to the advent of extensive genome sequencing projects. A major conclusion from sequence comparisons of various TE is that frontiers between the different types are becoming less clear. We detail these receding frontiers between different IS-related TE. Several, more specialized chapters in this volume include additional detailed information concerning a number of these.

In a second section of the review, we provide a detailed description of the expanding variety of IS, which we have divided into families for convenience. Our perception of these families continues to evolve and families emerge regularly as more IS are identified. This section is designed as an aid and a source of information for consultation by interested specialist readers.

INTRODUCTION

We have divided this review into two major sections. In one, we have attempted to present an overview of our current understanding of prokaryotic insertion sequences (IS), their diversity in sequence, in organization and in mechanism, their distribution and impact on their host genome, and their relation to their eukaryotic cousins. We discuss several IS-related transposable elements (TE) which have been identified since the previous edition of *Mobile DNA*. These include IS that use single-strand DNA intermediates and their related "domesticated" relations, insertion sequences with a common region (IS*CR*), and integrative conjugative elements (ICE), which use IS-related transposases (Tpases) for excision and integration. Several more specialized chapters in this volume include additional detailed information concerning a number of these topics. One of the major conclusions from this section is that the frontiers between the different types of TE are becoming less clear as more are identified. In the second part, we have provided a detailed description of the expanding variety of IS, which we have divided into families for convenience. We emphasize that there is no "quantitative" measure of the weight of each of the criteria we use to define a family. Our perception of these families continues to evolve and families emerge regularly as more IS are added. This section is designed as an aid and a source of information for consultation by interested specialist readers.

Received: 6 June 2014, **Accepted:** 29 January 2015, **Published:** 2 April 2015

Editors: Mick Chandler, Université Paul Sabatier, Toulouse, France, and Nancy Craig, Johns Hopkins University, Baltimore, MD

Citation: Siguier P, Gourbeyre E, Varani A, Ton-Hoang B, Chandler M. 2015. Everyman's guide to bacterial insertion sequences. *Microbiol Spectrum* 3(2):MDNA3-0030-2014. doi:10.1128/microbiolspec.MDNA3-0030-2014.

Correspondence: Mick Chandler, mike.chandler@ibcg.biotoul.fr

HISTORY

It is now over 40 years since the first IS were described. They were identified as short DNA segments found repeatedly associated with mutations in the *gal* operon and bacteriophage λ (1–3). Shortly after, it was established that IS were normal residents of the *Escherichia coli* chromosome (4) sometimes present in multiple copies. They were shown to be involved in generating deletions (5) and in activating gene expression (6). They were also identified as constituents of bacterial plasmids (7). At about the same time, it was observed that antibiotic resistance genes could also be transferred or "transposed" from one plasmid to another (8–10) and it was recognized that IS and "transposons" were both members of a group of genetic entities: transposable or mobile genetic elements (TE or MGE). This relationship between IS and transposons was reinforced by the observation that different DNA segments carrying different genes could be translocated by two flanking IS (11, 12). It was also realized (13) that they might be related to the controlling elements discovered by genetic analysis of maize several decades previously (14).

However, in spite of the observation that IS can be present in some bacterial species in extremely high copy numbers (15, 16), little at the time prepared us for the subsequent recognition of the preponderant role they play in shaping genomes, of their extreme diversity and their widespread distribution (see reference 17).

WHAT IS AN IS?

The original definition of an IS was: a short, generally phenotypically cryptic, DNA segment encoding only the enzymes necessary for its transposition and capable of repeated insertion into many different sites within a genome using mechanisms independent of large regions of DNA homology between the IS and target (18, 19). Classical IS are between 0.7 and 2.5 kb in length, genetically compact with one or two open reading frames (*orf*s) which occupy the entire length of the IS and terminate in flanking imperfect terminal repeat sequences (IR). The *orf*s include the Tpase that catalyzes the DNA cleavages and strand transfers leading to IS movement and, in some cases, regulatory proteins. Their highly compact nature is illustrated by the fact that some IS have developed "recoding" strategies such as Programmed Ribosomal Frameshifting (involving ribosome slippage) and Programmed Transcriptional Realignment (involving RNA polymerase slippage) (Chandler et al., this volume; (20, 21)). These permit assembly of different functional protein domains effectively encoding two proteins of different function in one DNA segment. IS also often generate a short flanking directly repeated duplication (DR) of the target on insertion. These characteristics are not limited to prokaryotic IS but are also shared with most eukaryotic DNA transposons. However, for prokaryotic IS, this strict definition has been broadened over the years with the discovery of an increasing number of noncanonical derivatives and variants, some of which are described below. Moreover, as we learn more about diversity from sequenced genomes, classification is becoming more problematic because the large degree of MGE diversity is obscuring the borders between certain types of TE (see *Fuzzy Borders* section) (20).

Despite their abundance and diversity, the number of different chemical mechanisms used in TE movement is surprisingly limited and many quite divergent TE share a similar mechanism.

ISFINDER AND THE GROWING NUMBERS OF IS

Since 1998, IS have been centralized in the ISfinder database (www-is.biotoul.fr). This provides a basic framework for nomenclature and IS classification into related groups or families (22). Initially IS were each assigned a simple number (23). However, to provide information about their provenance, IS nomenclature rules were changed and now resemble those used for restriction enzymes: with the first letter of the genus followed by the first two letters of the species and a number (24) (e.g., IS*Bce1* for *Bacillus cereus*).

In 1977 only five IS (IS*1*, IS2, IS*3*, IS*4*, and IS*5*) had been identified (13). At the time of publication of the first edition of *Mobile DNA* this had risen to 50 (25); at the time of the second edition, there were more than 700 (26). Currently, ISfinder includes more than 4,000 different IS. This represents only a fraction of IS present in the public databases. Not only has the number of IS identified increased dramatically with the advent of high-throughput genome sequencing but examination of the public databases has shown that genes annotated as Tpases, the enzymes that catalyze TE movement (or proteins with related functions), are by far the most abundant functional class (17).

MAJOR IS GROUPS ARE DEFINED BY THE TYPE OF TRANSPOSASE THEY USE

Insertion sequences can be grouped into families but, in the first instance, the principal division in IS classifica-

tion is based on the nature of their Tpases (Table 1). These can be divided into two major types based on the chemistry used in breaking and re-joining DNA during TE displacement: the DDE (and DEDD) and HUH enzymes.

DDE transposases

DDE enzymes, so-called because of a conserved Asp, Asp, Glu triad of amino acids that coordinate essential metal ions, use OH (e.g., H_2O) as a nucleophile in a transesterification reaction (27) (Hickman and Dyda, this volume). They do not form covalent Tpase–DNA intermediates during the transposition process.

Insertion sequences with DDE enzymes are the most abundant type in the public databases. This is partly because the definition of an IS became implicitly coupled to the presence of a DDE Tpase, an idea probably reinforced by the similarity between Tpases of IS (and other prokaryotic and eukaryotic TE) and the retroviral integrases (28–30) particularly in the region including the catalytic site. More precisely, for these TE, the triad is DD(35)E in which the second D and E are separated by 35 residues. As more DDE Tpases were identified, the distance separating the D and E residues was found to vary slightly. However, for certain IS, this distance was significantly larger. In these cases, the Tpases include an "insertion domain" between the second D and E residues (31) with either α-helical or β-strand configurations. Although in most cases this is a prediction, it has been confirmed by crystallographic studies for the IS*50* (β-strand) (32) and Hermes (α–helical) (33) Tpases. The function of these "insertion domains" is not entirely clear.

Although DDE-type transposons share basic transposition chemistry, different TE vary in the steps leading to formation of an insertion intermediate that has shed flanking donor DNA (Figure 1) (Hickman and Dyda, this volume). They catalyze cleavage of a single DNA strand to generate a 3′ OH at the TE ends, which is subsequently used as a nucleophile to attack the DNA target phosphate backbone. This is known as the transferred strand. The variations are a result of the way in which the second (nontransferred) strand is processed (34, 35). There are several ways in which second-strand processing can occur (Figure 1).

For certain IS, the second strand is not cleaved but replication following transfer of the first strand fuses donor and target molecules to generate cointegrates with a directly repeated copy at each donor/target junction. This is known as replicative transposition (e.g., IS*6*, Tn*3*) or more precisely, target primed replicative transposition (Figure 1, first column).

In the other pathways, the flanking donor DNA can be shed in several different ways.

The nontransferred strand may be cleaved initially several bases within the IS before cleavage of the transferred strand (e.g., IS*630* and *Tc1*) (36, 37) (Figure 1, second column).

The 3′ OH generated by first-strand cleavage may be used to attack the second strand to form a hairpin structure at the IS ends liberating the IS from flanking DNA and subsequently hydrolyzed to regenerate the 3′ OH (conservative or cut-and-paste transposition; e.g., IS*4*; Haniford and Ellis, this volume) (Figure 1, third column).

The 3′ OH of the transferred strand from one IS end may attack the other to generate a donor molecule with a single-strand bridge that is then replicated to produce a double-strand transposon circle intermediate and regenerating the original donor molecule (Copy out-Paste in or more precisely donor primed transposon replication, e.g., IS*3*; Chandler et al, this volume) (Figure 1, fourth column).

Or finally, the 3′ OH at the flank of the nontransferred strand may attack the second strand to form a hairpin on the flanking DNA and a 3′ OH on the transferred strand (at present this has only been demonstrated for eukaryotic TE of the hAT family and in V(D)J recombination (38).

Clearly, many families produce double-strand circular intermediates but this does not necessarily mean that they all use the Copy out-Paste in donor primed transposon replication mechanism because a circle could be generated by excision involving recombination of both strands (see Hickman and Dyda, this volume). These differences are reflected in the different IS families.

DEDD transposases

A similar type of Tpase, known as a DEDD Tpase, is related to the Holliday junction resolvase, RuvC (39, 40). These possess a similar predicted structural topology in their catalytic site and presumably have similar chemistry to the DDE enzymes. They are at present limited to a single IS family (IS*110*). They also exhibit a different order in their functional domains compared with most DDE Tpases: with the catalytic domain N-terminal to the DNA-binding domain. In addition, the associated IS do not have significant terminal IR and insertion does not necessarily generate direct target repeats. Hence, while the transposition chemistry may be similar to that of the DDE Tpases, the overall transposition mechanism is probably different.

TABLE 1 Characteristics of insertion sequence families

Families	Sub-groups	Typical size-range	DR (bp)	Ends	IRs	No. of ORF	Frameshift	Catalytic residues	Comments	Mechanism
IS*1*	-	740–1,180	8–9	GGnnnTG	Y	2	ORFAB	DDE		copy-and-paste and cointegrate
	single ORF	800–1,200	0–9			1				
	IS*Mhu11*	900–4,600	0–10		Y	2	ORFAB			
IS1595	IS*Pna2*	1,000–1,150	8	GGCnnTG	Y	1		DDNK		copy-and-paste ?
	ISPna2+pass	1,500–2,600	8							
	IS*H4*	1,000	8	CGCTCTT		1		DDNK		
	IS*1016*	700–745	7–9	GGGgctg		1		DDEK		
	IS*1595*	900–1,100	8	CcTGATT		1		DDNK + ER4R7		
	IS*Sod11*	1,000–1,100	8	nnnGcnTATC		1		DDHK + ER4R7		
	IS*Nwi1*	1,080–1,200	8	ggnnatTAT		1		DDEK + ER4		
	ISNwi1+pass	1,750–4,750	8			1				
	IS*Nha5*	3,450–7,900	8	CGGnnTT		1		DDER/K		
IS*3*	IS*150*	1,200–1,600	3–4	TG	Y	2	ORFAB	DDE		copy-paste
	IS*407*	1,100–1,400	4	TG						
	IS*51*	1,000–1,400	3–4	TG						
	IS*3*	1,150–1,750	3–4	TGa/g						
	IS*2*	1,300–1,400	5	TG						
IS*481*	-	950–1,300	4–15	TGT	Y	1		DDE		copy-paste ?
IS*NCY*	IS*1202*	1,400–1,700	5	TGT	Y	1		DDEQ		
IS*4*	IS*10*	1,200–1,350	9	CT	Y	1		DDE	Hairpin intermediate	cut-and-paste
	IS*50*	1,350–1,550	8–9	C					Hairpin intermediate	
	IS*Pepr1*	1,500–1,600	7–8	-T-AA					?	
	IS*4*	1,400–1,600	10–13	-AAT					?	
	IS*4Sa*	1,150–1,750	8–10	CA					?	
	IS*H8*	1,400–1,800	10	CAT					?	
	IS*231*	1,450–5,400	10–12	CAT		1 or + *			* Passenger genes	
IS*701*	-	1,400–1,550	4		Y	1		DDE		
	IS*Aba11*				Y	1		DDE		
IS*H3*	-	1,225–1,500	4–5	C-GT	Y	1		DDE		
IS*1634*	-	1,500–2,000	5–6	C	Y	1		DDE		

IS*5*	IS*903* IS*L2* IS*H1* IS*5* IS*1031* IS*427*	950–1,150 850–1,200 900–1,150 1,000–1,500 850–1,050 800–1,000	9 2–3 8 4 3 2–4	GG -GC Ga/g GAa/g Ga/g	Y	1 1 1 1 1 2	 ORFAB	DDE		
IS*1182*	-	1,330–1,950	0–60		Y	1		DDE		
IS*NCY*	IS*Dol1*	1,600–1,900	6–7		Y	1		DDE		
IS*6*	-	700–900	8	GG	Y	1		DDE		co-integrate
IS*21*	-	1,750–2,600	4–8	TG	Y	2 *		DDE	* istB: transposition helper	
IS*30*	-	1,000–1,700	2–3		Y	1		DDE		copy-and-paste
IS*66*	- IS*Bst12*	2,000–3,000 1,350–1,900	8–9 8–9	GTAA GTAA	Y Y	3 * 1		DDE * DDE	* TnpC has the DDE domain	
IS*256*	- IS*1249* IS*C1250*	1,200–1,500 1,300 1,250	8–9 0–10 0–9	Ga/g GG GG	Y Y Y	1 1 1		DDE DDE DDE		copy-paste
IS*H6*	-	1,450	8	GGT	Y	1		DDE		
IS*Lre2*	-	1,500–2,000	9		Y	1		DDE		
IS*Kra4*	IS*Azba1* IS*Mich2* IS*Kra4*	1,400–2,900 1,250–1,400 1,400–3,700	0 8 9	 GGG GGG	Y Y Y	1 or + * 1 or 2 1 or + *	 ORFAB 	DDE DDE DDE	* Passenger genes * Passenger genes	
IS*630*	-	1,000–1,400	2 *		Y	1 or 2	ORFAB	DDE	* Target site : often NTAN with duplication of the TA	cut-and-paste
IS*982*	-	1,000	3–9	AC	Y	1		DDE		
IS*1380*	-	1,550–2,000	4–5	CC	Y	1		DDE		
IS*As1*	-	1,200–1,500	8–10	CAGGG	Y	1				
IS*L3*	-	1,300–2,300	8	GG	Y	1				
Tn*3*	-	>3,000	0	GGGG	Y	>1		DDE		co-integrate

(continued)

TABLE 1 Characteristics of insertion sequence families *(continued)*

Families	Sub-groups	Typical size-range	DR (bp)	Ends	IRs	No. of ORF	Frameshift	Catalytic residues	Comments	Mechanism
IS*Azo13*	-	1,250–2,200	0–4	Ga/g	Y	1				
IS*110*	- IS*1111*	1,200–1,550	0		N Y *	1		DEDD	 * IRs not at the termini of the IS	
IS*91*	-	1,500–2,000	0		N	1		HUH/Y2	Target site GAAC ( ) CAAG	rolling circle
IS*200*/IS*605*	IS*200* IS*605* IS*1341*	600–750 1,300–2,000 1,200–1,500	0		N	1 * 2 * 1 *		HUH/Y1 HUH/Y1 ** 	* TnpA * TnpA + TnpB ; ** Y1 on TnpA * TnpB	peel-and-paste
IS*607*	-	1,700–2,500	0		N	2 *		Serine **	* TnpA + TnpB; ** TnpA	
IS*NCY*	IS*892* IS*Lbi1* IS*Mae2* IS*Plu15* IS*A1214* IS*C1217* IS*M1*	1,600 1,400–1,500 1,400–2,400 800–1,000 1,000–1,200 1,200 1,300–1,600	0–8 5 9 0 8–12 6–8 8–9	CTAG CAG TAG 	Y Y Y N Y Y Y	2 1 1 1 2 1 1	ORFAB			

Abbreviations: DR, duplication repeat; IS, insertion sequence; ORF, open reading frame.

HUH transposases

Named after a conserved pair of His residues separated by a large hydrophobic amino acid (U), the HUH Tpases use tyrosine as a nucleophile and generate a transitory covalent 5′ tyrosine–DNA transposition intermediate—for review see (41) (Hickman and Dyda, this volume; He et al., this volume).

The TE encoding the second major type of Tpase, called HUH, have been identified more recently. HUH enzymes are widespread single-strand nucleases. They include Rep proteins involved in bacteriophage and plasmid rolling circle replication and relaxases or Mob proteins involved in conjugative plasmid transfer (41). They are limited to two prokaryotic (IS*91* and IS*200*/IS*605*; He et al., this volume) and one eukaryotic (helitron; Thomas and Pritham this volume) TE family. As Tpases, they are involved in presumed rolling circle transposition and also in single-strand transposition (see Hickman and Dyda, this volume; He et al., this volume). Their transposition chemistry is radically different to that of DDE group elements. It involves DNA cleavage using a tyrosine residue and transient formation of a 5′ phospho-tyrosine bond between the enzyme and its substrate DNA. In addition, the associated transposons have an entirely different organization and include subterminal secondary structures instead of IR (see *IS families* section below and He et al., this volume). There are two major HUH Tpase families: Y1 and Y2 enzymes depending on whether there is a single or two catalytic Y residues (41) (Dyda and Hickman, this volume). Although these enzymes use the same Y-mediated cleavage mechanism, IS*200*/IS*605* family Y1 transposases and IS*91* transposases appear to carry out the transposition process in quite different ways.

Serine transposases

A third but minor type of Tpase resembles a site-specific serine recombinase and is at present limited to a single IS family, IS*607*. These presumably use the catalytic serine to generate a 5′ phospho-serine bond between the enzyme and its substrate DNA in a similar way to serine recombinases such as the Tn*3* resolvase (see chapters by: Stark; Rice; Smith; and Johnson, this volume).

Tyrosine transposases

Finally, tyrosine site-specific recombinases of the bacteriophage integrase (Int) type are often associated with conjugative transposons (ICE) and are considered to be Tpases. However, at present there are no known IS that use this type of enzyme. These transposases presumably use a catalytic Y to generate a transitory intermediate with a 3′ phospho-tyrosine bond between the enzyme and its substrate DNA as do the site-specific recombinases (see chapters by Jayaram et al.; van Duyne; Landy; Carraro and Burrus; and Wood and Gardner, this volume) as suggested by early studies with Tn*916* (42).

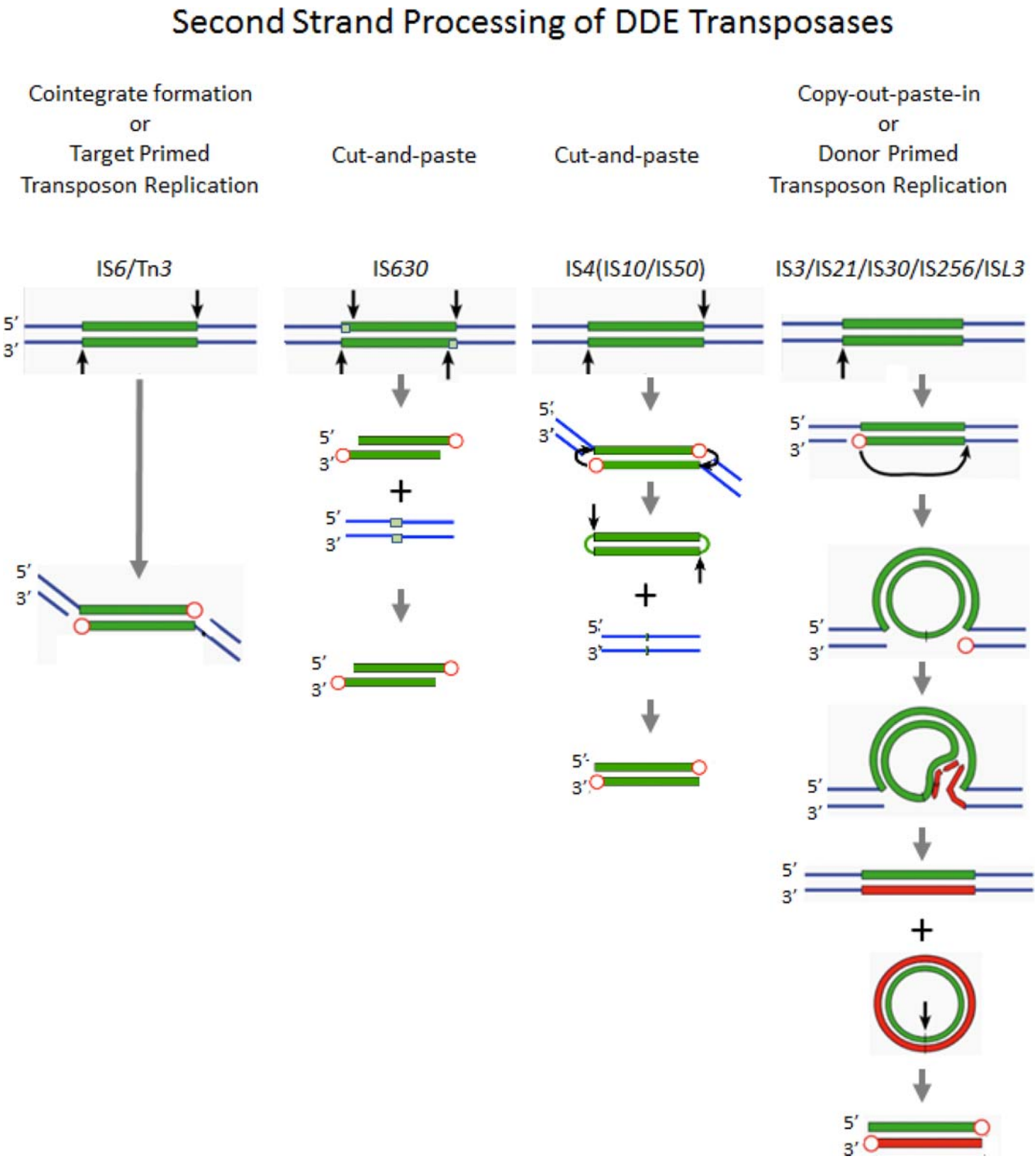
This nomenclature is clearly complex and stems directly from the history of the field. It has often led to confusion in genome annotations in the public databases. The reader is referred to Dyda and Hickman (this volume), He et al., (this volume) and to chapters in the section *Conservative Site-Specific Recombination* for more detailed descriptions of the overlapping issues of mechanism, function, and structure.

FUZZY BORDERS

With our increasing knowledge of mobilome diversity (the ensemble of mobile genetic elements including TE, ICE, Genomic Islands [GI], plasmids, phages, and integrons), the distinction between IS and other TE is becoming increasingly unclear. The major feature used to distinguish IS from transposons was that the former (Figure 2A) lack phenotypically detectable passenger genes (genes not involved in the transposition process) whereas the latter include one or more such genes (for antibiotic resistance, virulence and pathogenicity functions or genes permitting the use of unusual compounds). This is no longer the case (Figure 2). Many examples have now been identified in which passenger genes are located within IS or in which TE with typical transposon structures are devoid of transposition proteins.

Transporter IS: IS and relatives with passenger genes

Over the past few years, a number of TE have been identified that are very closely related to known IS but that carry passenger genes not directly involved in transposition. These are called transporter IS (tIS) (43) (Figure 2C). Passenger genes include transcription regulators (e.g., IS*Nha5*, members of the IS*1595* family), methyltransferases (e.g., IS*220*, IS*1380* family), and antibiotic resistance (e.g., IS*Cgl1*, IS*481* family) genes. They can include a significant amount of DNA with no clear coding capacity (e.g., IS*Bse1*, IS*Spo3*, and IS*Spo8*, IS*1595* family) and are longer than typical IS (e.g., IS*Causp2*, 7,915 bp, IS*1595* family). This has presumably delayed their identification. As the second IS end would occur at an unexpectedly distant position, they would resemble partial IS copies lacking a second end. They are never present in high numbers and often only in



single copy, suggesting that their transposition activity may be limited.

IS derivatives of Tn*3* family transposons

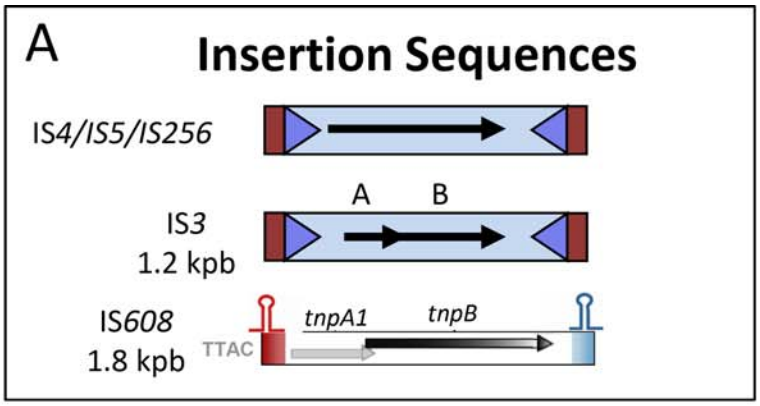
Another source of ambiguity for classification purposes occurs in the Tn*3* family (Nicolas et al., this volume) (Figure 2B). Tn*3* family members are quite variable. They include a number of diverse passenger genes that can represent entire operons, notably mercury resistance, or individual genes involved in antibiotic resistance, breakdown of halogenated aromatics or virulence (44). They often carry integron recombination platforms enabling them to incorporate additional resistance genes by recruiting integron cassettes (45). These are small DNA segments that carry promoterless passenger genes and integrate by site-specific recombination into the integron recombination platform. This platform provides an appropriate resident promoter to govern their expression (see Escudero et al., this volume). Members are characteristic: they have long relatively well-conserved IR and a particularly long Tpase (950 to 1,025 amino acids). They also encode a site-specific recombination ("resolution") system necessary for completion of their transposition (Nicolas et al., this volume). IS*1071*, composed of Tn*3*-like IR and Tpase gene but lacking both the site-specific recombination system and passenger genes, was identified many years ago (46). This clearly accords with the definition of an IS. Several other examples have now been identified (e.g., IS*Vsa19*, IS*Shfr9*, IS*Busp1*).

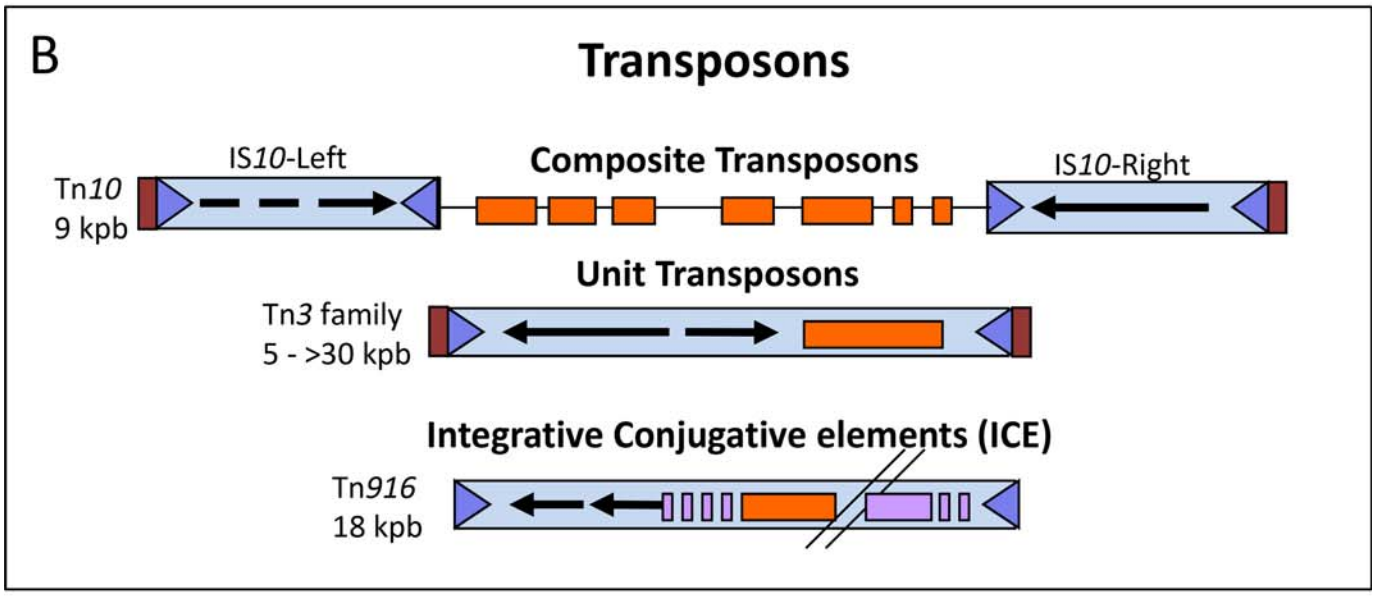
IS related to ICE

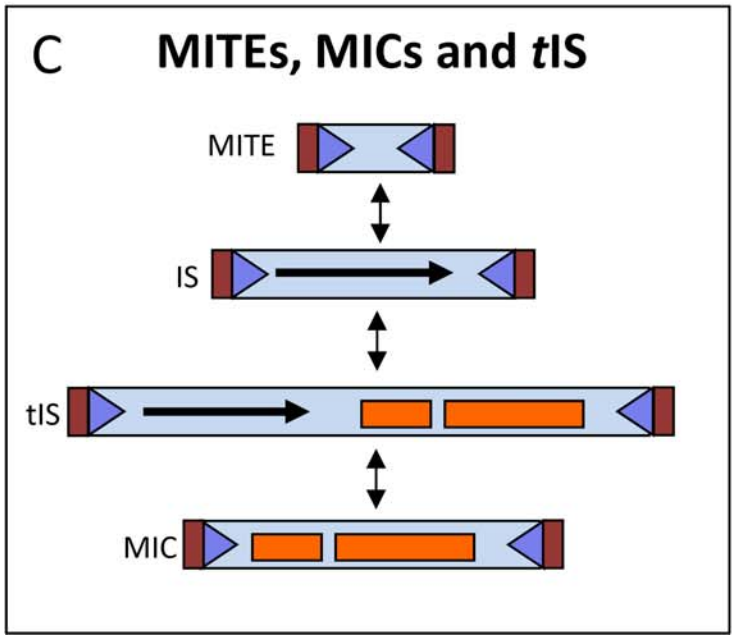
The ICE, which were initially identified integrate and excise from their host chromosomes using a tyrosine-based enzyme related to phage integrases (47) (Figure 2B) (see chapters by Wood and Gardner and by Carraro and Burrus, this volume). These also carry genes permitting intercellular transfer, although derivatives exist that are not capable of autonomous transfer and are known as IMEs or CIMEs (integrative mobilizable elements; or *cis*-mobilizable elements) (48, 49). Some, known as GI, also include other types of passenger gene. Depending on the type of passenger genes, GI have been called pathogenicity islands or symbiotic islands.

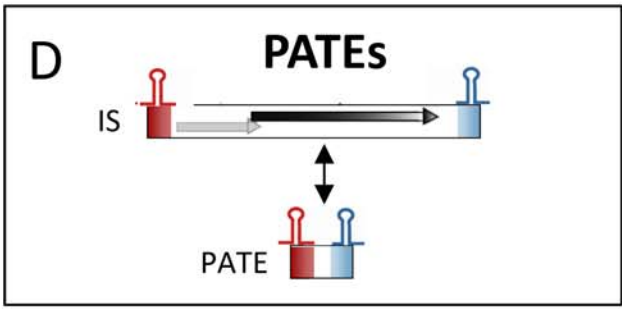
More recent studies (50–53) have identified ICE with typical DDE Tpases. One group, TnGBS, initially found in Group B Streptococcus, has led to the identification of an entire family of typical insertion sequences, the IS*Lre2* family (52), whereas another shows a close relationship to IS*30* family members (53). The ICE that use typical DDE Tpases also include IR with sequences resembling those of the related IS. Hence it is becoming difficult to draw a distinction between certain GI, ICE and IS.

FIGURE 1 Insertion sequence (IS) families with DDE transposases are distinguished by how the second ("nontransferred") strand is processed. IS are shown in green, flanking DNA in blue. Cleavage is shown as bold vertical arrows. 3′ OH residues are shown as red circles, replicated DNA is indicated in red. The first column shows initial cleavages which generate the 3′OH of the transferred strand and are subsequently used to attack target DNA (not shown) without prior liberation from the flanking donor DNA. Their transfer generates a forked molecule in which a donor and target strand are joined to the TE at each end and which provides a 3′ OH in the flanking target DNA that can prime replication of the transposable elements (TE). This might be called target primed transposon replication. TE of the Tn*3* and IS*6* families transpose in this way. The second column shows a pathway adopted by the IS*630* family. Here, the nontransferred strand is cleaved two bases within the TE (light green square) before cleavage of the transferred strand, which generates the 3′ OH. Repair of the donor molecule would lead to inclusion of a non-complementary 2-bp scar or footprint (light green square). This is a cut-and-paste mechanism without TE replication. The third column represents transposition using a hairpin intermediate in which the transferred strand is first cleaved and the resulting 3′ OH then attacks the opposite strand to form a hairpin at the TE ends liberating the TE from flanking donor DNA. This is then hydrolyzed to liberate the final transposition intermediate. This is a cut-and paste mechanism without TE replication. The fourth column shows a "copy out-paste" in mechanism adopted by a large number of IS families. It involves cleavage of one IS end and attack of the opposite end by the liberated 3′ OH, the TE then undergoes replication using the 3′ OH in the donor DNA, a process that might be called donor primed transposon replication. This generates a double-strand DNA transposon circle and regenerates the donor molecule. The circle then undergoes cleavage and insertion. Adapted from references 35 and 259. doi:10.1128/microbiolspec.MDNA3-0030-2014.f1









IS*91* and IS*CR*

A final example of the subtle line dividing IS and transposons is found in the IS*91*/IS*CR* group. IS*91* was identified some time ago (54) and carries a single Tpase *orf*. More recently, a group of related elements, IS*CR* (IS with a "common region") was described (reviewed in references 55 and 56). Although there has been no formal demonstration that these actually transpose, the CR is an *orf* that resembles the IS*91* family HUH Tpases (41). The major feature of IS*CR* elements is that they are associated with a diverse variety of antibiotic resistance genes and, particularly in the case of *Pseudomonas* IS*CR*, with aromatic degradation pathways, both upstream and downstream of the Tpase *orf*. This is therefore another example of a very particular group of IS derivatives that appear to include multiple passenger genes.

NON AUTONOMOUS IS DERIVATIVES

Many prokaryotic genomes are littered with IS fragments and small, non autonomous IS derivatives whose transposition can, in principal, be catalyzed in *trans* by the Tpase of a related complete IS (Figure 2C). These miniature inverted repeat transposable elements (MITE) were first identified in plants (57) and are related to IS with DDE Tpases. They are short (~300 bp), include terminal IR but no Tpase and generally generate flanking DR. Equivalent MITE-like structures called palindrome-associated transposable elements (PATE) (58) including the ends of IS*200*/IS*605* family members (see *IS families*, below; He et al., this volume) with their subterminal secondary structures have also been identified (Figure 2D). Both MITE and PATE probably derive from IS by internal deletion. The first MITE identified were related to eukaryotic Tc/mariner elements (distantly related to the bacterial IS*630* family) and IS*630*-related MITE were also the first described bacterial examples (59–61). MITE showing similarities to other IS families (e.g., IS*1*, IS*4*, IS*5*, IS*6*, and even Tn*3* family members) have since been identified in bacteria and archaea (62, 63).

Another group of IS derivatives related to MITE are called mobile insertion cassettes (MICs) (64) (Figure 2C). These, like MITE, are flanked by IR, do not include a Tpase gene, and generate flanking DR. They carry various coding sequences, and are present in relatively low copy number. The IS*231* subgroup of the large IS*4* family includes examples of many of these IS-derivatives (canonical IS, MITE, MIC and tIS) (65).

RELATIONSHIP BETWEEN IS AND EUKARYOTIC TE

In spite of their obvious similarities, there is often poor transfer of knowledge between studies of prokaryotic and eukaryotic TE. This artificial barrier is reflected in their nomenclature systems: Prokaryotic TE are named following the basic logic of bacterial genetics built on the initial Demerec rules (66); Eukaryotic TE, on the other hand, have more colorful names in keeping with the culture of nomenclature used in eukaryotic genetics. To a certain extent, this idiosyncratic nomenclature camouflages the diversity and relationships between members of the eukaryotic TE superfamilies and their prokaryotic cousins.

It is important to appreciate that the basic chemistry of transposition is identical for both prokaryotic and

FIGURE 2 Organization of different insertion sequence (IS) -related derivatives. IS with DDE transposases (Tpases) and their derivatives are shown as blue boxes, terminal inverted repeats as light blue triangles and flanking direct target repeats as red boxes. The Tpase *orf*s are shown as black horizontal arrows. Passenger genes are shown as orange boxes and transfer functions (in the case of ICE) are shown as purple boxes. The single-strand IS are indicated with their left (red) and right (blue) subterminal secondary structures indicated. (A) IS organization. From top to bottom: a typical IS with a single Tpase *orf*; an IS in which the Tpase reading frame is distributed over two reading phases and requires frameshifting for expression; and the organization of a typical member of the single-strand IS family IS*200*/IS*605*. (B) Different IS-related TE. From top to bottom: composite transposon Tn*10* with inverted flanking copies of IS*10* (note that the left IS*10* copy is not autonomously transposable); a unit transposon of the Tn*3* family; and an integrative conjugative element (ICE). (C) Relationship between IS, miniature inverted repeat transposable elements (MITE), transporter IS (tIS) and mobile insertion cassettes (MIC). (D) Generation of palindrome-associated transposable elements (PATE) from IS*200*/IS*605* family members. Adapted from references 20 and 43. doi:10.1128/microbiolspec.MDNA3-0030-2014.f2

eukaryotic elements (Hickman and Dyda, this volume). Moreover, many eukaryotic DNA transposons have similar sizes and organization to those of prokaryotic IS and, as most do not carry additional "passenger" genes, they are not transposons in the prokaryotic sense and should strictly be considered as eukaryotic IS. The major differences lie in how Tpase expression and activity are regulated (67). One important difference is that most eukaryotic transposons are "insulated" by constraints of the nucleus (which physically separate the transposition process from that of Tpase expression) whereas those of prokaryotes are not, because prokaryotic transcription and translation are coupled. In addition, eukaryotic transposons are subject to a hierarchy of regulation via small RNAs (68, 69). In prokaryotes, it is possible that CRISPRs (clustered regularly interspaced short palindromic repeats) may impose some control at this level but, although it has been demonstrated that CRISPRs are active against mobile genetic elements and may regulate some endogenous gene expression (70), these are limited to plasmids and phage and to our knowledge have not yet been demonstrated to act on intracellular MGE such as IS and transposons.

In spite of these differences, a significant number of eukaryotic DNA TE are related to prokaryotic IS, and moreover, eukaryotic TE including passenger genes are now being identified (71). This reinforces the view that the borders between different types of TE are "fuzzier" than previously recognized.

IS IMPACT

There is an increasing body of data describing the impact of IS on their host genomes. Although outside the scope of this review, it is important to provide a brief overview here because this is a crucial feature of IS.

IS expansion, elimination and genome streamlining

Perhaps one of the most spectacular characteristics of IS populations is their capacity to expand within genomes and also participate in genome streamlining or trimming by facilitating DNA deletion. There are many striking examples of IS expansion (20). These are often observed in bacteria with recently adopted host-restricted lifestyles. Current opinion is that the nutritionally rich environment of the host reduces the requirement for many genes that are essential for free-living bacteria, allowing fixation of slightly deleterious mutations in the population by random stochastic IS transposition and concomitant increase in IS copy number.

Additionally, IS activity may also eliminate genes responsible for surface antigens and interfere with regulatory circuitry, providing increased protection against host defenses (72, 73). Fixation would be facilitated by successive population bottlenecks (74–77) and the effect would be more marked the more genetically isolated the bacterial population. Intracellular endosymbionts provide many examples of this. Accumulation of IS-generated endosymbiont pseudogenes in functions that can be supplied by the host would increase the host dependence of the bacterial population (78–80). The capacity of IS to generate deletions would also be expected to eventually lead to complete or partial elimination of the IS themselves. This is thought to have occurred in the IS-free ancient endosymbionts. Genetic isolation of the population is key to the process of IS elimination because reinfection could occur by lateral gene transfer from other bacterial populations, a phenomenon that is thought to have occurred several times in certain *Wolbachia* sp. (81).

IS and gene expression

Another important aspect of IS impact on their bacterial hosts is their ability to modulate gene expression. In addition to acting as vectors for gene transmission from one replicon to another in the form composite transposons (two IS flanking any gene; Figure 2B) and tIS (Figure 2C) and their ability to interrupt genes, it has been known for some time that IS can also activate gene expression (5, 82). This capacity has recently received much attention because of the increase in resistance to various antibacterials (83, 84), a worrying public health threat (85–87).

They can accomplish this in two ways: either by providing internal promoters whose transcripts escape into neighboring DNA (82, 88) or by hybrid promoter formation. Many IS carry -35 promoter components oriented towards the flanking DNA. In a number of cases this plays an important part in their transposition because a significant number of IS transpose using an excised transposon circle (see *Major IS groups are defined by the type of transposase they use* section, above) with abutted left and right ends. For these IS, the other end carries a –10 element oriented inwards towards the Tpase gene. Together with the –35, this generates a strong promoter on formation of the circle junction to drive the Tpase expression required for catalysis of integration (Chandler et al., this volume) (89–91). Hence, if integration occurs next to a resident –10 sequence, the IS –35 sequence can contribute to a hybrid promoter to drive expression of neighboring genes, see reference 92.

At present this phenomenon had been reported to occur with over 30 different IS in at least 17 bacterial species (20, 93).

Target choice

The influence of different IS on genome architecture will depend not only on their levels of activity but also on the type of target into which they insert. It was initially believed that TE show no or only low sequence specificity in their target choice. For example IS*630* and the eukaryotic Tc/*mariner* families (Tellier et al., this volume) both require a TA dinucleotide in the target (36, 37) whereas others such as the IS*200*/IS*605* and IS*91* families require short tetra- or penta-nucleotide sequences (94, 95). Yet others, such as IS*1* and IS*186* (of the IS*4* family), show some regional specificity (for AT–rich and GC-rich sequences, respectively) (96–98).

Although, from a global genome perspective, insertion may appear to occur without significant sequence specificity, accumulation of more statistically robust data has uncovered rather subtler insertion patterns. For example, there is some indication from the public databases suggesting that IS density is generally significantly higher in conjugative bacterial plasmids than in their host chromosomes with the exception of special cases in which the host has undergone IS expansion as described above. Such plasmids are major vectors in lateral gene transfer providing good delivery systems for TE. Some TE, including IS, appear to be attracted to replication forks (99, 100) and show a strong orientation bias indicating strand preference at the fork (99–101). Moreover, in certain cases, insertion may target stalled replication forks (He et al., this volume). A link between replication (in this case, replication origins) and insertion has also now been observed for a eukaryotic TE: the P element of *Drosophila* (102).

Some transposons such as Tn*7* in *Escherichia coli* (103) and Tn*917* in *Bacillus subtilis* (104), *Enterococcus faecalis* (105) and *Streptococcus equi* (but not in *Listeria monocytogenes* or *Streptococcus suis*) (106) also show a preference for integration into the replication terminus region and sites of DNA breakage may also attract insertions (103). It remains to be seen whether any IS has adopted these types of target preference.

Potential topological characteristics or secondary structures are another feature that can attract certain TE. Changes in topology induced by the nucleoid protein, H-NS, for example, may explain the effects of H-NS mutants on the target choice of IS*903* and Tn*10* (IS*10*) (107) (Haniford and Ellis, this volume). Members of the IS*110*, IS*3*, and IS*4* families are examples of IS that insert into potential secondary structures such as repeated extragenic palindromes (He et al., this volume) (108–110), integrons (Escudero et al., this volume; 111, 112) or even the ends of other TE (113, 114).

In addition, IS*21*, IS*30*, and IS*911* have all been observed to insert close to sequences that resemble their own IR (115–118). Although these IS are members of different families, they have in common the formation of a dsDNA excised circular transposon intermediate with abutted left and right ends (Chandler et al., this volume). Insertion next to a resident "target" IR such that IR of the IS are abutted "head-to-head" presumably reflects the capacity of the Tpase to form a synaptic complex between one IR present in the transposon circle and the target IR. This type of structure is extremely active in transposition and will continue to generate genome rearrangements.

Other factors that may be influential in determining target choice are interactions (direct or indirect) with DNA-associated host proteins. For example, Tn*7* appears to be drawn to replication forks by interaction with the fork-associated β-clamp (119). Indeed, a recent large-scale analysis of IS insertions in bacterial chromosomes has pointed out the potential implication of sliding clamp in targeting and provided evidence for direct interaction of various transposases with the conserved replication processing factor (120). A more detailed exploration of these results would be useful. Another potential interaction, in this case with RNA polymerase, is suggested by the recent observation that the transposon TnGBS (an ICE from *Streptococcus agalactiae*) and members of the closely related IS*Lre2* family insert preferentially 15–17 bp upstream of σA promoters (50, 51). Targeting of upstream regions of transcription units has also been extensively documented for certain eukaryotic transposons (121).

Influence of transposition mechanisms

The way in which strand cleavages and transfers occur during transposition also affects the outcome of the transposition events and therefore impinges on genome structure. For example Tn*3* and IS*6* family members, both with DDE Tpases, generate fusions or cointegrates between the donor and target replicons by a process of replicative transposition, presumably by target primed transposon replication (Hickman and Dyda, this volume) (Figure 1 first column). However, in the event of intramolecular transposition, this type of mechanism is expected to give rise to inversions with a copy of the IS at each junction or inversions with a single IS copy remaining and a second copy segregating with a circular-

ized deletion (122). Note that similar effects are also known to occur by homologous recombination between two inverted or directly repeated IS copies in a replicon. Other known mechanisms such as cut-and-paste, or Copy out-Paste in (donor primed transposon replication) (34) would not generate this type of genomic rearrangement but could contribute to genomic modifications in other ways such as "nearly precise excision" (123, 124) or by using alternative sequences that resemble their IR (125, 126).

Summary

The above considerations serve to provide an overview of the diversity of IS and their close TE relatives together with an outline of their behavior and its impact on prokaryotic genomes. Below we provide a guide to the present classification of IS and present a more detailed description of the individual IS families, their characteristics, differences, similarities, transposition mechanism (where known), and their distribution. There are clearly large disparities in our level of knowledge from family to family. Although we have tried to treat each family individually, we have grouped those that have proved to be related. We also include information concerning the identification of the different types of derivative elements such as MITES or tIS, which are related to given individual IS families.

IS FAMILIES

Insertion sequences in ISfinder are classified into families using a variety of characteristics (Table 1) (127) such as: transposition chemistry; length and sequence of the short imperfect terminal IR sequences (TIR in eukaryotes) carried by many IS at their ends; length and sequence of short flanking direct target DNA repeats; *orf* organization; and the nature of their target sequences. In many cases, these distinguishable differences are associated with different properties and behavior of the IS. An additional criterion is based on the presence/absence and order of various Tpase domains. Transposases are multidomain proteins. They include domains or modules involved in catalysis (e.g., DDE and HUH—see *Major IS groups are defined by the type of transposase they use* section, above), in sequence-specific binding to the TE ends (e.g., helix-turn-helix, HTH; zinc finger, ZF) and in multimerization (leucine zipper). The sequence-specific binding module is generally located in the N-terminal region permitting folding and binding of the nascent peptide before completion of translation (see *IS3 and IS481 families* section, below). In addition, they may include specific signatures for interaction with host proteins (Peters, this volume).

At present, IS in ISfinder are grouped into 29 families and many of these are further divided into subgroups based on shared characteristics. Although there are very well-defined homogeneous families (such as IS*3*, IS*30*, and IS*256*) others have been redefined over time as more and more IS are identified (e.g., IS*4* and IS*5*).

It is important to note that the number of IS in a given ISfinder family does not necessarily reflect their relative abundance in nature. Some estimates of this have been made from time to time using different approaches, although the accuracy of these is difficult to assess (74, 128, 129). IS inclusion in ISfinder has not involved a systematic global search of the public databases.

We emphasize that this is a description of the ISfinder classifications at this time and that some groups will certainly emerge with further additions to the database and more detailed analysis.

IS WITH DDE TRANSPOSASES

IS*1* and IS*1595*: two related families

IS*1*.

(i) General

IS*1* was among the first bacterial insertion sequences to be identified (13). It is also one of the shortest (768 bp) and has been identified in over 40 different bacterial and archaeal species. IS*1* is a component of several compound transposons (12, 130) where it is present in direct or inverted orientation. It is also found in several conjugative plasmids flanking large regions carrying a number of antibiotic resistance genes (resistance determinant or r-det) (131) and can participate in homologous recombination to generate circular r-det forms or tandem multimers resulting in increased antibiotic resistance (132).

(ii) Organization

The family has been extensively reviewed previously (26, 127) (and references therein). Integration generates a 9 bp target DR but DR of 8, 10, and 14 bp also occasionally occur. The classic IS*1* includes 23 bp imperfect inverted repeats (IRL and IRR) and two partly overlapping *orf*s (*insA* and *insB*') located in the 0 and –1 relative translational phases and expressed from a promoter partially located in IRL. Their integrity is essential for transposition. The *insA* product, InsA, binds both IR and regulates expression and probably transposition activity. The

Tpase, InsAB', is a transframe fusion protein produced by programmed ribosomal frameshifting, between *insA* and *insB'* with typical frameshift signals. Programmed ribosomal frameshifting occurs at a frequency of about 1%. No InsB' protein species has been detected.

IS*1*-related derivatives carrying only a single *orf* have been identified. These tend to be longer (~1,000 bp) than the classic IS*1* with slightly longer Tpases due to an N-terminal extension (Table 1). They retain the characteristic IS*1* IR sequences (Table 1). They were first observed in the archaeal Sulfolobiales (IS*C1773a* and *b* and IS*Sto7*) where this arrangement appears to be the rule (63) but are not restricted to the Archaea. Several examples occur in Eubacteria (e.g., IS*Aba3*, *Acinetobacter baumannii*, and IS*Pa14*, *Pseudomonas aeruginosa*) (43). It is possible, as in the case of *dnaX* (133), that the upstream N-terminal protein is indeed produced but by frameshifting to generate the smaller derivative from the full-length protein.

Overall transposition activity appears to depend on the ratio of InsA/InsAB', serving to regulate activation of transposition by uncontrolled Tpase expression from external transcription. It had been suggested that a translational restart within the *insA* frame giving rise to an InsAB' protein with an N-terminal deletion generates the true Tpase (134). However, although the importance of this protein cannot be ruled out, the establishment of an *in vitro* IS*1* transposition system based on partially purified engineered InsAB' suggests that the translational restart product may not play a central role (135).

(iii) Mechanism

In vivo, direct visualization of 13 DNA species obtained following induction of IS*1* transposition and the kinetics of their appearance and disappearance clearly identified forms corresponding to reciprocal products of IS-mediated deletions as well as excised transposon circles. This suggests that IS*1* can transpose using both the cointegrate (target primed transposon replication) and Copy out-Paste in (donor primed transposon replication) pathways (136).

(iv) Transposase organization

Alignment of InsAB' from different members of the family confirmed the presence of a C-terminal DDE catalytic domain (137) and also revealed potential N-terminal ZF and HTH motifs (135, 138). Addition of 1,10-phenanthroline, which shows a high affinity for zinc, prevented binding to IS*1* IR, as did mutations in either of the two motifs whereas mutation of the DDE motif confirmed its importance in catalysis but not in binding (135, 138). All three motifs are observed in the longer Tpases with a single long reading frame.

However, members of the IS*Mhu11* subgroup lack the N-terminal ZF while retaining the HTH motif (43). They include a 30–120-residue C-terminal extension that is unrelated in different members of this group and the spacing between the second D and E residues is 40–60 amino acids longer. Three different organizations of IS*Mhu11* subgroup members were identified: examples with two *orf*s and a potential frameshift zone (IS*Mhu11*, IS*Mac25*, IS*Arch18*, and IS*Acma3*); a single example with additional noncoding DNA upstream of the Tpase *orf* (IS*Beg1*); and members that carry passenger genes (Table 1) generally with no known function, but often with other relatives in different bacteria. An exception is tIS*Sce1*, which includes *orf*s resembling a DNA methyltransferase, a possible sigma factor, and member of the HTH_XRE family of transcription regulators. However, only a single example of each type with passenger genes was identified, suggesting that these IS have low or no transposition activity.

More extensive comparisons have indicated that IS*1* is distantly related to another relatively newly recognized family, IS*1595*.

IS*1595*.

(i) General

IS*1595* was identified in *Xanthomonas campestris* (43) and closely related IS (e.g., IS*Xo2*, IS*Xo5*, IS*Xo16*, and IS*Xca4*) are present in high copy number in other *Xanthomonas* species. The IS*1595* family is less homogeneous than the IS*1* family. BLAST analysis of ISfinder using IS*1595* as a query confirmed a distant relationship with IS*1* family members and with IS*1016* (139), a multicopy *Neisseria* element previously binned with the "unclassified" IS (IS*NYC*). Seven subgroups have been identified: IS*1595*, IS*1016*, IS*Pna2*, the halobacterial IS*H4* group (140), IS*Sod1*, IS*Nwi1* and IS*Nha5*. Members of the IS*Pna2*, IS*Nwi1* and IS*Nha5* subgroups may contain passenger genes or additional noncoding DNA (43). As in the case of IS*Mhu11*, only a single example of each member was identified, suggesting that these IS have low or no transposition activity. It is also related to IS*Sag10* from *Streptococcus agalactiae*. This was originally called MTnSag1 and thought to be a member of the IS*1* family (141) but now called tIS*Sag10* due to the presence of an O-lincosamide nucleotidyltransferase passenger gene. tIS*Sag10* also carries an origin of transfer and can be mobilized by transfer functions of Tn*916*. It can therefore also be defined as a non autonomous

ICE, again underlining the increasing difficulty in TE classification.

(i) Organization

Most are flanked by 8 bp AT-rich DR and have a single Tpase *orf*. Like IS*1* family Tpases, they include an N-terminal ZF, an HTH motif, and a C-terminal catalytic motif with some exceptions. For example IS*1016* group members lack the N-terminal ZF but, as IS*1016* is present in multiple copies, it is probably active. The catalytic sites of all family members show group-specific variation particularly around the final E residue. More surprising is the apparent substitution of the final E residue for N or H in certain members. The exact nature of these possible noncanonical catalytic sites will require experimental determination.

It has been reported that the IS*1595* family (in particular IS*1016*) is related to the Merlin family of eukaryotic TE (142), especially at the level of the DDE motif and the position of an upstream HTH. They also have comparable lengths and similar IR. The Merlin Tpase is longer at the N-terminus than that of IS*1016* and more similar in size to the other members of the IS*1595* family although it does not exhibit the IS*1595* N-terminal ZF.

IS*3* and IS*481* families

IS*3*.

(i) General

The IS*3* family is one of the most coherent and largest IS families (143) (see Chandler et al., this volume). It is very widely distributed and at present ISfinder includes 554 different members from more than 270 bacterial species. The family is quite homogeneous in spite of its wide distribution in bacteria exhibiting a large range of G+C contents and of the presence of members in hosts such as *Mycoplasma* with a nonuniversal genetic code (e.g., IS*1138*) or in bacteria that use stop codon read-through by insertion of the unusual amino acid selenocysteine (e.g., IS*Dvu3* from *Desulfovibrio vulgaris*). In the case of both copies of IS*1138*, which participates in high-frequency rearrangements of the *Mycoplasma pulmonis* chromosome, the Tpase *orf* carries 11 UGA codons, which are decoded as tryptophan (144).

Although most members had been limited to the eubacteria (63), an example, IS*Mco1*, has now been identified in the Archaea *Methanosaeta concilii*. As this Archaea is widespread in nature (145), it is possible that this represents a case of recent horizontal transfer. The presence of eight copies implies that IS*Mco1* is active in its archaeal host.

(ii) Organization

Members carry relatively well conserved IR and invariably express their Tpases as fusion proteins using programmed -1 frameshifting. All have characteristic terminal IR generally terminating with the dinucleotide 5′-CA-3′ and carrying a short block of GC-rich sequence. They generate 3 or 4 bp DR on insertion. Their Tpase, OrfAB, like those of the majority of IS*1* family members, is expressed as a fusion protein by programmed -1 programmed ribosomal frameshifting from two consecutive, partially overlapping reading frames with the second located in the -1 reading phase compared to the upstream frame (Chandler et al., this volume).

A clade carrying noncanonical ends has recently been identified. This is currently composed of 15 members. These IS include seven supplementary base pairs on each end flanking canonical IS*3* ends: a conserved stretch of five C residues is located 5′ to the left IR and a less conserved motif (CGG) is located 3′ to the right end. When these additional bases are taken into account every member of this clade exhibits a 4 bp DR characteristic of the IS*3* family (Table 1) (E. Gourbeyre, unpublished). This conclusion is supported by the presence of multiple IS copies (e.g., IS*Psy31*) and also by identification of "empty sites" (ISfinder). This clearly requires further experimental investigation.

Recently, an additional subgroup has been proposed, which includes IS*Ppy1* (146). However, all members belong to the IS*150* subgroup and their Tpases are not separated by our standard multiple alignment and Markov Cluster Analysis (MCL) analysis. Although they do exhibit some variation in the sequence of their terminal dinucleotides, similar variations are found for IS*2* and members of other IS*3* subgroups.

(iii) Mechanism

IS*3* family members transpose using a Copy out-Paste in, donor primed transposon replication, mechanism that produces dsDNA circular intermediates (Figure 1). The recognition sequences and cleavage points have been well characterized experimentally in several IS*3* family members. *In vitro* transposition systems have been developed for several family members (Chandler et al., this volume). Bridged molecules are formed in an asymmetric manner by cleavage of one IS end to generate a 3′ OH, which then attacks the other end to generate a single-strand bridge between the IS ends. A copy of the IS appears to be replicated out as a circle leaving the other IS copy in the reconstituted donor DNA replicon (Chandler et al., this volume).

One member of this family, IS*911*, has been used to address the mechanism, common to many IS, in which the Tpase shows a preference for acting on the IS copy from which it was synthesized (*cis* activity). It was shown that this behavior depends on coupling between transcription and translation and results from co-translational binding of the nascent Tpase peptide to the IS ends (147).

IS*481*

Initially, IS*481* appeared to be an IS*3* family derivative that had been truncated for the N-terminal end of the Tpase and includes a C-terminal extension. The DDE active site domain and the IR (ending in 5′ TGT 3′) are similar to those of IS*3* family members. Their presence in high copy number in some species and the identification of at least 130 distinct but related IS from over 90 species strongly suggests that these represent a distinct transpositionally active family. Different members generate DR of between 4 and 15 bp. Moreover, certain members (e.g., IS*Sav7*) insert specifically into the tetranucleotide CTAG, which becomes the flanking DR and provides the UAG termination codon for the Tpase. In contrast to the vast majority of IS*3* family members, the IS*481* Tpase is not produced by frameshifting. There is no evidence for a leucine zipper as in IS*3*.

Some members include passenger genes including antibiotic resistance (Cm^R for IS*5564* and IS*Cgl1*), or potential transcriptional regulators (IS*Krh1*, IS*Pfr21*, IS*Sav7*). IS*481* itself has played a fundamental role in the evolution of the genomes of the Bordetellae where, in *Bordetella pertussis* it has undergone extensive amplification to several hundred copies with accompanying genome decay (72, 73).

These IS are distantly related to the eukaryotic Banshee transposon, which at present is restricted to the anaerobic flagellated protozoan *Trichomonas vaginalis* (Pritham per. comm.) (148). They share the highly conserved Pfam integrase core domain identified initially in the IS*3* family and retroviruses (28, 29). They also show a conserved 5′ TG 3′ tip to the IR, which is typical of this and other types of mobile element. It would be interesting to determine whether Banshee transposes using a dsDNA circular intermediate as do IS*3* family members.

IS*1202* group (IS*NCY*)

A small group including IS*1202* (149), composed of 10 IS which had been included in the IS*NCY* (not classified yet) group appears distantly related to IS*481*. Members are between 1,400 and 1,700 bp (except for IS*Kpn21*, which includes a passenger gene annotated as "hypothetical protein") with a Tpase *orf* of between 400 and 500 amino acids in a single reading frame. Their IR begin with TGT as do those of the IS*3* and IS*481* families. They generate DR of between 5 and, unusually, 27 bp.

They appear to have similarities at the level of their Tpases particularly in their DDE domains (e.g., IS*1202* has 39% amino acid similarity to IS*Pfr5* of the IS*481* family). They include a glutamine (Q) seven residues C-terminal to the conserved E instead of the characteristic K/R. Identification of additional IS will be necessary to clearly define this group.

The redefined IS*4* and related families

The accumulation of additional related IS has permitted a more detailed analysis of the IS*4* family, which had already become heterogeneous displaying extremely elevated levels of internal divergence (150). Based on more than 200 IS*4*-related sequences from bacterial and archaeal genome sequences, seven subgroups and three emerging families (IS*701*, IS*H3*, and IS*1634*) were defined. Separation into three emerging families was principally due to variations in an important conserved YREK motif, a division that is supported by the IR sequences and the associated DR.

Members of all these encode a Tpase with an insertion domain rich in β-strand and located between the second D and the E of the DDE motif (31). That of the IS*50* Tpase (151) is the only example that has structural support (152), although bioinformatic analysis (31) indicated that IS*H3* (e.g., IS*C1359* and IS*C1439*), IS*701* (e.g., IS*701* and IS*Rso17;* (153)) and IS*1634* (e.g., IS*1634*, IS*Mac5*, IS*Plu4*; (154)) family members also exhibit a similar insertion domain.

IS*4*

The IS*4* family originally included a diverse collection of IS characterized by three conserved domains in the Tpase [N2, N3, and C1 containing D, D, and E respectively; (151)]. In addition to the conserved DDE triad, this family is defined by the presence of an additional tetrad YREK (150) which, in the case of IS*50*, is thought to be involved in coordination of a terminal phosphate group at the 5′ end of the cleaved IS (155) (Haniford and Ellis, this volume).

IS*701*

The IS*701* family was distinguished from the IS*4* family by a highly conserved 4 bp target duplication, 5′-YTAR-3′. MCL analysis also indicated that the Tpases form a

defined and separate group and alignments indicated the absence of Y in the Tpase YREK motif. There are several clades within this family. A new clade, IS*Aba11*, was proposed as a new family based on five IS (156). Members of this group generate 5 bp target duplications (instead of 4 bp), exhibit conserved IR and include HHEK instead of YREK. However, additional examples (ISfinder) exhibiting the conserved IR did not universally contain HHEK and MCL cluster analysis did not strongly support the notion that IS*Aba11* constitutes a new family. At present, we have retained IS*Aba11* as a subgroup in the IS*701* family.

IS*H3*

The IS*H3* family is restricted to the Archaea. In roughly half of the 30 members identified, the Tpase lacked the K/R residue of the DDE motif whereas all except IS*Fac10* displayed a Y(2)R(3)E(3)R motif. A characteristic of this family is the presence of 5 bp DR flanked at one end by A and at the other by T.

IS*1634*

The IS*1634* family (previously IS*1549*) is characterized by large Tpases due to a β-strand insertion domain located between the conserved second D and E residues, which is 35 to 79 amino acids longer than that of IS*4* and members of the other related families. They generate 5 bp to 6 bp AT-rich DR and are present in both Archaea and bacteria (150). Certain members generate very long variable DR (e.g., IS*1634* from 17 to 478 bp, (154); IS*Csa8* from 16 to 131 bp; IS*Mhp1*, 80 bp).

The redefined IS*5* and related IS*1182* families

Like the IS*4* family, growth in the number of identified members of the IS*5* (564 members) and IS*1182* families (142 members) has revealed that the two are related and has allowed a more detailed analysis and a separation into various subgroups and families The IS*5* family is now partitioned into six subgroups: IS*5*, IS*903*, IS*L2*, IS*H1*, IS*1031*, and IS*427*. Some of these may prove to be emerging families.

There is a distant relationship, about 30% similarity, with the Pif/Harbinger group of eukaryotic TE (148, 157).

IS*5*

Although the majority of members have a single Tpase *orf*, about 20% of family members may express Tpase by frameshifting because it is distributed between two translation phases similar to most of the IS*427* subgroup (82/116) (127). In these cases the frameshifting signals appear more appropriate for a programmed transcriptional realignment frameshift mechanism rather than for classical translation frameshifting (programmed ribosomal frameshifting) as there are no obvious downstream enhancement signals (21). Similar split reading frames have now been identified in several other subgroups: IS*1031* (13/65 members); IS*L2* (7/43); and few in the IS*5* subgroup (7/149). There is no experimental evidence that these frameshift signals are functional but many of the IS are in multiple copies, suggesting that the derivatives are active. In view of their diversity compared with families such as IS*3*, the subgroups will certainly be partitioned into additional groups as more IS are identified. At present, the IS*903* and the archaeal IS*H1* subgroups do not contain members with potential frameshifting.

In addition to their Tpases and the presence or absence of potential frameshifting, a further distinction between these elements resides in their target specificities. Certain IS*427* subgroup members and IS*1182* family members do not carry a termination codon for their Tpases but generate this on insertion into a specific target sequence, CTAG, which is duplicated on insertion. Other IS such as IS*1031*, duplicate a sequence TNA while others such IS*L2* appear to duplicate ANT.

IS*1182*

IS*1182* family members exhibit a diverse set of target specificities. Some duplicate 4 bp. These are of two types: those specific for CTAG and those that show no apparent target sequence specificity. Yet others target palindromic sequences. These are also of different types: some insert at the 3′ foot of a stem-loop and duplicate the entire structure whereas others insert 3′ of the loop and simply duplicate the loop (P. Siguier, E. Gourbeyre and M. Chandler, unpublished).

IS*Dol1* group (IS*NCY*)

Another small group, IS*Dol1*, with 17 members from 15 bacterial species has recently emerged from the IS*NCY* "orphan" group. Members have a length of between 1,600 and 1,900 bp and generate DR of 6–7 bp.

IS*6*

(i) General

There are at present 130 family members in ISfinder from nearly 80 bacterial and archaeal species but this represents only a fraction of those present in the public databases. The family was named after the directly repeated insertion sequences in transposon Tn*6* (158) to standardize the various names that had been attributed

to identical elements (e.g., IS*15*, IS*26*, IS*46*, IS*140*, IS*160*, IS*176*). Many are found as part of compound transposons (159–161) invariably as flanking direct repeats, a consequence of their transposition mechanism.

Recent activity resulting in horizontal dissemination is suggested by the observation that copies identical to *Mycobacterium fortuitum* IS*6100* occur in other bacteria: as part of a plasmid-associated catabolic transposon carrying genes for nylon degradation in *Arthrobacter* sp. (162), from the *Pseudomonas aeruginosa* plasmid R1003 (164), and within the *Xanthomonas campestris* transposon Tn*5393b* (164).

One member, IS*257*, has played an important role in sequestering a variety of antibiotic resistance genes in clinical methicillin-resistant *Staphylococcus aureus* isolates. It provides an outward oriented promoter that drives expression of genes located proximal to the left end. Moreover, both left and right ends appear to carry a –35 promoter component, which would permit formation of hybrid promoters on insertion next to a resident –10 element (165).

A single member, IS*Dsp3*, present in single copy in *Dehalococcoides* sp. BAV1 carries a passenger gene annotated as hypothetical protein.

IS*6* family elements are abundant in Archaea and cover almost all of the traditionally recognized archaeal lineages (methanogens, halophiles, thermoacidophiles, and hyperthermophiles) (63).

They form a monophyletic group related to bacterial IS from Firmicutes but can be further divided into three phylogenetic groups present in the halophiles, the sulfolobales, and the pyrococcales/methanosarcinales. The IR of the archaeal IS*6* members are variable compared with the bacterial members and generally terminate with 5′-GT or 5′-GA, as opposed to the 5′-GG found in bacteria. The large phylogenetic distribution of IS*6* family members in the Archaea and the monophyly of the IS*6* archaeal group suggest that these elements were ancestrally present in the Archaea rather than being recently acquired by lateral gene transfer from bacteria.

IS*26* is encountered with increasing frequency in plasmids of clinical importance where it is involved in expression of antibiotic resistance genes and plasmid rearrangements (166–171).

(ii) Organization

The putative Tpases are very closely related and show identity levels ranging from 40 to 94%. They generally range in length from 789 bp (IS*257*) to 880 bp (IS*6100*). However, a separate group represented by seven members are somewhat larger (approximately 1,200 bp) as a consequence of an N-terminal extension with a predicted ZF. Several members (e.g., IS*Rle39a*, IS*Rle39b*, and IS*Enfa1*) apparently require a frameshift for Tpase expression. It is at present unclear whether this is biologically relevant. However, alignment with similar sequences in the public databases suggests that IS*Enfa1* itself has an insertion of 10 nucleotides and is therefore unlikely to be active.

All carry short related (15–20 bp) terminal IR and generally create 8 bp DR. A single *orf* is transcribed from a promoter at the left end and stretches across almost the entire IS. In the case of IS*26* this is located within the first 82 bp of the left end and the intact *orf* is required for transposition activity.

(iii) Mechanism and insertion specificity

The predicted amino acid sequence of the Tpase exhibits a strong DDE motif. Translation products of this frame have been demonstrated for several members (e.g., IS*240*). Little is known about Tpase expression although transposition activity of IS*6100* in *Streptomyces lividans* is significantly increased when the element is placed downstream of a strong promoter.

Where analyzed, members of the IS*6* family give rise exclusively to replicon fusions (cointegrates) in which the donor and target replicons are separated by two directly repeated IS copies (e.g., IS*15D*, IS*26*, IS*257*, IS*1936*) (172). Transposition of these elements therefore presumably occurs in a replicative manner by target primed transposon replication. No known specific resolvase system such as that found in Tn*3*-related elements has been identified in this family and it is assumed that cointegrate resolution occurs via homologous recombination. It is for this reason that compound IS*6*-based transposons carry directly repeated flanking IS copies. Recent results suggest that the IS*6* family member IS*26* may transpose in an unusual manner (173), an observation which merits further investigation.

No marked target selectivity has been observed.

IS*21*

(i) General

The IS*21* family is fairly homogeneous. It has more than 140 members from about 80 bacterial species. IS*21* was discovered in plasmid R68 where it was subsequently observed to undergo a tandem duplication, which greatly facilitated the insertion of the plasmid into the *Pseudomonas* host genome. This led to the formation of Hfr strains (174, 175) and the subsequent demonstration that the *Pseudomonas* chromosome was genetically circular (176).

(ii) Organization

IS*21* family members encode two genes, the Tpase *istA* and a "helper" gene, *istB*, which exhibits some similarity to the DnaA replication initiator protein due to the presence of an ATP-binding motif, and often appears in BLAST searches of complete genomes. In many members, the termination codon of *istA* and the initiation codon of *istB* overlap, suggesting that IstB is produced by translational coupling. The IS*21* family terminal IR are complex and carry several tandem repeated sequences thought to be Tpase binding sites. Like a number of IS families, the ends of the element terminate with the dinucleotide 5′-CA-3′.

(iii) Mechanism

IS*21* family members transpose using a two-step mechanism by formation of a reactive junction, similar to those formed in the copy-paste mechanism of IS*3* and other families, in which two abutted IS*21* ends are separated by several nucleotides. This is consistent with the marked tendency of IS*21* to insert in, or close to, an IS*21* end. The reactive junction is subsequently integrated into the target DNA to give a DR of 4 to 8 bp. Integration is optimal when the distance separating the two ends in the junction is 4 bp. It is efficient with a 2 or 3 bp separation but inefficient with smaller or larger intervening sequences. IstA carries the DDE motif and is the Tpase. The molecular details of IstB activity are not known. Both IstA and IstB are required for efficient transposition whereas a product of an alternative translation initiation within the Tpase gene may facilitate integration. Using IstA-enriched *Escherichia coli* cell extracts, it was shown that this protein is responsible for 3′ end cleavage of IS*21* and of both ends in the IR junction (177).

IS*30*

(i) General

The IS*30* family currently comprises 94 members from over 70 bacterial species and an example, ISC*1041*, has also been found in the Archaea (178). IS*30*-like Tpases have also been found as integral parts of certain ICE from methicillin-resistant *Staphylococcus aureus* (53).

Members of this family are capable of activating neighboring genes by creation of a hybrid promoter on insertion next to a −10 promoter element (83, 179–181).

(ii) Organization

IS*30* family members have a single Tpase *orf* spanning almost their entire length. The Tpases show several well-conserved regions. One, in the N-terminus includes a potential helix–turn–helix (HTH) motif, which, in the case of IS*30*, is responsible for IR binding (182, 183). Another in the C-terminal region contains the DD(33)E motif. The terminal IR are between 20 and 30 bp and contain some conserved sequence signatures (183), and their tips show significant sequence variation although in most of the elements the IR have not been experimentally confirmed. Insertion generally generates DR of 2–3 bp but there are several exceptions in which the DR is between 12 and 32 bp (e.g., IS*1630*, IS*1470*, IS*658*, IS*L7*, IS*Lpl1*) (183).

(iii) Mechanism and insertion specificity

The founding member of the family, IS*30*, is the best characterized at the mechanistic level (117, 182–189) and an *in vitro* transposition system has been developed (190). This 1,221 bp long *Escherichia coli* element belongs to a growing class of IS known to transpose through an intermediate formed by abutting the IR, donor primed transposon replication. Here the IR are separated by 2 bp (179, 184, 186). Such an IR junction can be created by formation of a dimer of two directly repeated IS copies or by the formation of transposon circles. Both IS minicircles and dimers have been observed. IR–IR junctions have also been detected in some other IS*30* family members such as IS*18* (181), IS*4351* (191), and IS*1470* (193). A structure in which two IS*30* ends are linked by a single-strand bridge (forming a figure of eight structure on a circular plasmid, has been identified (190).

IS*30* also shows a preference for two distinct types of target sequence: "natural" hot spots, characterized by a 24 bp symmetric consensus, and, like many IS which transpose via a circular dsDNA intermediate, the IR of the element itself. A similar type of insertion specificity was observed for IS*1655* from *Neisseria meningitidis* (189).

IS*66*

(i) General

IS*66* was first identified in the Ti plasmid pTi66 of *Agrobacterium tumefaciens* (193). The vast majority of IS*66* members originate from the Proteobacteria with several from the Bacteriodetes/Chlorobi and the Firmicutes. A second group of closely related IS, widely spread among both bacteria and Archaea are thought to represent a subgroup within the IS*66* family (194). The founder member, IS*Bst12*, originally isolated from *Bacillus stearothermophilus*, was described as a novel family (195), but identification of many additional examples suggests that the IS*Bst12* and IS*66* groups should

be considered a single family. IS*Bst12* are found in Actinobacteria, Cyanobacteria, Deinococcus/Thermus, Firmicutes, and Planctomycetes as well as in Proteobacteria. They are also found in the Euryarchaeota phylum of Archaea (but have not yet been identified in the Crenarchaeota).

(ii) Organization

The IS*66* reference copy from a plasmid of the enteropathogenic *Escherichia coli* B171, IS*679*, (196) is defined by three *orf*s: *tnp*A, *tnp*B, and *tnp*C and relatively well conserved terminal IR of about 20–30 bp flanked by an 8 bp DR at their insertion sites. Orf *tnp*C is 1,572 bp and its predicted product includes a typical DDE motif. It also carries an insertion domain between the second D and the E of the DDE motif (e.g., IS*679*, IS*Psy5*, and IS*Mac8*) (31)

The role of the products of *tnpA* (651 bp) and *tnpB* (345 bp) is less clear. Mutation of each *orf* separately (by introduction of an in-frame deletion) reduced transposition by at least two orders of magnitude (196). The three frames are disposed in a pattern suggesting translational coupling: *tnpB* is in general in translational reading frame –1 compared with *tnpA* and in most cases the termination codon of *tnpA* and the initiation codon of *tnpB* overlap (ATGA). An initiation codon for *tnpC* occurs slightly downstream separated from *tnpB* by about 20 bp.

However, rather surprisingly, in the light of a requirement for all three *orf*s for transposition of the canonical IS*66* family member IS*679*, members of the IS*Bst12* group are devoid of *tnpA* and *tnpB* and carry only the *tnpC* reading frame. Although both IS*Bst12* and IS*66* members contain IR, which start with 5′-GTAA-3′, they are clearly distinguishable due to a single conserved A at bp 11 In IS*Bst12*, which is not conserved in IS*66*.

IS*66* members can be grouped into three classes based on their organization: those including all three *orf*s, A, B, and C transcribed in the same direction; those with additional passenger genes invariably present downstream of *orf*C and transcribed in the same direction; and those which lack *orf*A but retain both *orf*s B and C. Each of these organizations includes members with multiple copies, implying that they are active in transposition. In addition to the DDE catalytic domain (194), TnpC also exhibits a highly conserved CwAH-rR motif downstream of the second D residue, a relatively conserved CX2(C)X33CX2C motif characteristic of a ZF further upstream and a leucine-rich region which might form a leucine zipper necessary in multimerization of other Tpases (197), at the N-terminus.

(iii) Mechanism and insertion specificity

Nothing is known about the transposition mechanism of this group of IS and they exhibit no substantial target sequence specificity.

The IS*256* cluster

Recently, a study of ICE elements identified examples from type B Streptococcus (TnGBS) (50) and Mycoplasma (199) that include a DDE type Tpase rather than the more common phage integrase-like gene. Using a cascade PSI-Blast approach not only revealed two new IS families (IS*Lre2* and IS*Kra4*) but established a distant relationship with the IS*256* and IS*H6* families (52).

Analysis of the N-terminal Tpase region (52) also identified two shared domains (N1 and N2). N2 corresponds to a potential HTH domain in the region of the IS*256* Tpase, which recognizes the terminal IR (199).

The cluster can be divided into five clades containing nine groups based on branching of the Tpases phylogenetic tree: two types of closely related TnGBS, TnGBS1, and TnGBS2, and IS*Lre2* (MULT3); the mycoplasma ICE; IS*256* (MULT1); IS*H6* (MULT2); IS*Azba1*, IS*Mich2*, IS*Kra4* (MULT4) (52).

There is a distant relationship with the Tpase of the eukaryotic Mutator TE and, like MuDR from *Zea mays*, many generate 8 bp to 9 bp target repeat on insertion. They have therefore been called MULE (for Mutator-like Elements). Like MuDr/Foldback, members of these groups carry a largely α-helical insertion domain (31) between the second D and E catalytic residues. This includes a conserved C/D(2)H signature present in the eukaryotic and prokaryotic IS (52, 200).

IS*256*

The IS*256* family can be subdivided into three groups: IS*256*, IS*1249*, and IS*C1250*. The classical IS*256* group has more than 180 members in both bacteria and Archaea. They are between 1,200 and 1,500 bp long with IR of 20–30 bp and generate DR of between 8 and 9 bp.

IS*256* itself was originally isolated as a component of the compound transposon Tn*4001* (201, 202). This family is quite homogeneous. Members carry related IR of between 24 and 41 bp, and most generate 8–9 bp duplications. A single long *orf* carrying a potential DDE motif with a spacing of 112 residues between the second D and E residues, together with a correctly placed K/R residue. The catalytic residues have been validated by mutagenesis (203). It was shown several years ago that the Tpase of IS*256* family elements shares some similarities with the eukaryotic Mutator element (204), a

relationship that has been explored recently in more detail (205).

Members of this family transpose using an excised circular dsDNA transposon intermediate (203, 206).

IS*1249* group

There are more than 30 members confined at present to the actinobacteria and the firmicutes. They are about 1,300 bp in length with IR of about 26 bp and generally generate DR of 8 bp (with variations of between 0 and 10).

IS*C1250* group

At present there are only three members of this family in ISfinder. All are found in the Archaea *Sulfolobus solfataricus*.

IS*H6*

This group (MULT2) was originally observed uniquely in Archaea (63). There are 11 members of about 1,450 bp with highly conserved IR of 24–27 bp, DR of 8 bp and a single Tpase *orf* encoding a protein of 450 bp.

IS*Lre2*

There are 48 entries for IS*Lre2* family members in ISfinder. They are restricted at present to the bacteria. They are between 1,500 and 2,000 bp long, with IR from 15 to 29 bp and generate 9 bp DR. Together with the related TnGBS ICE, show strong target specificity and insert 13-17 bp upstream of σA promoters (50, 52) in oriented fashion with the right IS end (RE) proximal. PCR analysis has detected a transposon circle junction, as with the related ICE, suggesting that transposition may occur via a donor primed transposon replication process.

IS*Kra4*

This new family includes 83 members and is divided into three related groups: IS*Azba1*, IS*Mich2*, and IS*Kra4*.

(i) IS*Azba1*

There are presently 28 members of this group. They encode a Tpase of between 450 and 480 amino acids, are 1,400 to 2,900 bp long with IR of about 20 bp and no DR. Six (IS*Afe13*, IS*Cot1*, IS*Ec51*, IS*Kpn19*, IS*Sysp7*) carry an *orf* in addition to the Tpase and this specifies a protein related to serine-recombinases or resolvases. Four of these also include a third *orf* annotated as hypothetical protein. The fifth, IS*Afe13*, carries the Tpase, a resolvase and an alternative *orf* annotated as ORF-3-like from plasmid pRiA4b. Other proteins found in this family are annotated as being hypothetical or putative TnpR resolvases although no direct evidence for resolvase function is available. Eight other members simply encode the Tpase and the ORF-3 like protein. While IS*Cep1* includes the ORF-3-like protein and a third annotated as phage integrase or *xer*C/D.

(ii) IS*Mich2*

This includes 24 members which are presently limited to the cyanobacteria. Twenty-two have a Tpase *orf* distributed between two reading phases whereas in the remaining two the Tpase forms a unique continuous *orf*. However, all show a potential but atypical frameshift motif, TTTTTT which could be involved in either programmed ribosomal frameshifting or programmed transcriptional realignment recoding. Further experimental analysis would be necessary to confirm or refute this. Members are between 1,250 and 1,400 bp long with a Tpase of 360 amino acids, IR of between 18 and 39 bp with 8 bp DR. Three members (IS*Cysp26*, IS*Mic1*, and IS*Mich2*) carry a passenger gene annotated as hypothetical protein.

(iii) IS*Kra4*

The 31 elements in this group range in size from 1,400 to 3,700 bp due to the presence in some of various passenger genes. They have IR of 18 to 31 bp and generate DR of 9 bp. Three carry passenger genes: IS*Ldr1*, a hypothetical protein and a reverse transcriptase; IS*Sri1*, a transcriptional regulator; and IS*Tni1*, a hypothetical protein. Six members may express their Tpases by frameshifting (five include a 7A motif and one with a motif, 5TC).

IS*630*

(i) General

There are over 160 members from over 80 bacterial and archaeal genomes.

(ii) Organization

Members are between 1,100 and 1,200 bp long with terminal IR and generally include a single *orf*. However, in about 90 members, the Tpase *orf* is distributed over two reading frames, suggesting that it may be produced as a fusion protein by frameshifting.

(iii) Mechanism

IS*630* transposition has been addressed *in vitro* (207) using IS*Y100* (IS*TcSa*) first identified in *Synechocystis* sp. PCC6803 (208). The Tpase was shown to specifically bind IS*Y100* IR using an N-terminal domain containing

two potential HTH motifs. It is the only protein required for IS*Y100* excision and integration and introduces double-strand breaks on mini-IS*Y100* on a supercoiled DNA substrate. Tc*1*/*mariner* element transposition has also been extensively studied *in vitro* (37) and a Tpase structural model is available (209, 210). IS*630* Tpase cleaves exactly at the 3′ (transferred strand) IS ends and two nucleotides inside the 5′ (nontransferred strand) ends. Cleavage is less precise on linear substrates. Both single-end and, less frequently, double-end insertion occur *in vitro* in a TA-target-specific manner (208). Transposition does not involve a hairpin intermediate.

(iv) Insertion specificity

Family members show high target specificity inserting into and duplicating a TA dinucleotide with a preference for the sequence 5′-NTAN-3′ (211). As the cleavages of the nontransferred strand occur 2 nucleotides within the 5′ end of the IS, repair of the donor molecule after excision of the IS can result in a 2 bp scar at the excision site.

The IS*630* family is related to the Tc1/mariner family of eukaryotic TE particularly at the level of the DDE signature. There is also an N-terminal HTH motif. Moreover, IS*630* and the Tc1/mariner families target similar sequences, have similar DR and transposition of both involves cleavage of two nucleotides inside the 5′ ends (37).

IS*982*

(i) General

The IS*982* family has nearly 100 entries in ISfinder from over 40 bacterial and archaeal species. In the case of IS*Lpl4* from *Lactobacillus plantarum*, identical copies have been detected in *Leuconostoc mesenteroides*, *Oenococcus oeni*, and *Lactobacillus sakei* indicating horizontal gene transfer. At least two members, IS*982*B (212) and IS*1187* (83, 213, 214) can provide a −35 hexamer in their right IR capable of forming a hybrid promoter with a resident −10 and activating neighboring genes.

(ii) *Organization*

IS*982* family members are between 962 and 1,155 bp long and carry similar terminal IR of between 18 and 35 bp with conserved ends: 5′-ACCC-3′. They encode a single *orf* of between of 271 and 313 amino acids with a possible DDE motif but without a convincing conserved downstream K/R residue. Little is known about the transposition of these elements. They generate DR of 6 to 8 bp.

Although the Tpase of a majority of members occupies a single reading phase, there are several examples in which the gene is distributed over two phases. It has been reported that a +1 nucleotide insertion in the Tpase *orf* of an IS*Lpl4* from *Oenococcus oeni* may undergo programmed translational frameshifting at a low rate (215). Although this must be confirmed, it would represent the first functional case of +1 frameshifting in IS. The Tpase of archaeal element IS*Pfu3* is also distributed over two phases. IS*Pfu3* carries a potential transcriptional frameshift signal A7 (programmed transcriptional realignment) present in all five copies suggesting that IS*Pfu3* is active.

IS*1380*

This family is represented by 153 members from nearly 100 bacterial species in ISfinder. They show conserved ends terminating with CCt/c. Although the majority of their Tpases often include the canonical DDE(6)K/R, several members exhibit other residues in place of the K/R. These include DDE(6)Q or DDE(6)I. A subgroup, IS*942*, composed of 13 members all restricted to Bacteriodetes, include DDE(6)N. None of these differences appear to affect the predicted secondary structure. In addition, to the host-restricted IS*942* group, two other branches of the Tpase tree are restricted to the *Actinobacteria*. A single, poorly characterized NCBI database entry (WP_018034290) probably corresponds to an archaeal IS*1380* member and intriguingly, an *orf* (XP_002337507) with a 100% match to IS*Lsp5* (*Leptospirillum* sp.) has been identified in the genomic sequence of the black cottonwood tree, *Populus trichocarpa*. Tpases of this family include an insertion domain with a predominantly β-strand secondary structure (31).

IS*1380* itself (216) is present in high copy number in the *Acetobacter pasteurianus* NCI1380 genome and in several strains of acetic acid bacteria. At present the family contains two tIS present in a single copy: IS*Msm12* (*Mycobacterium smegmatis*; tetR + methyltransferase) and IS*Rop1* (*Rhodococcus opacus*; reverse transcriptase).

This IS family is distantly related to the eukaryotic PiggyBac TE family (see Yusa, this volume) which also includes an insertion domain largely in the form of β-strand (148).

IS*As1*

There are over 80 entries for this family in ISfinder from over 50 bacterial species. There are currently no archaeal members. IS*As1* family members are between 1,200 and

1,326 bp long and generally carry terminal IR of between 14 and 22 bp. A single *orf* of between 294 and 376 amino acids occupies almost the entire length. There are several conserved D and E residues. The putative Tpase of IS*1358* has been visualized using a phage T7 promoter-driven gene and that of ISAs*1* has been detected in *Escherichia coli* minicells. The family also includes "H-repeats", which form part of several so-called rearrangement hot spots (RHS) elements containing another repeated sequence, the H-rpt element (Hinc repeat). H-rpt display features of typical insertion sequences although no transposition activity has yet been detected. For the sake of clarity, the H-rpt DNA sequences B (RhsB), C1 to C3 (RhsC), E (RhsE), and min.5 as well as H-rptF were renamed ISEc*1* to ISEc*7*, respectively (127). The Tpases of this family include a β-strand insertion domain (31).

Little is known about the transposition properties of this family of elements. However, recent experiments with the *Vibrio cholerae* element IS*1358* have demonstrated that insertion generates 10 bp DR and that, in *Escherichia coli*, it undergoes simple insertion into a target plasmid, pOX38 (217).

IS*L3*

(i) General

There are more than 120 members from nearly 80 bacterial species. The family also includes archaeal members, particularly in the Methanomicrobia. A potential tIS derivative that includes a mercury-resistance operon has also been identified in a conjugative plasmid in *Enterococcus faecium* (218).

(ii) Organization

Members range in size from 1,186 bp to 1,553 bp, carry closely related IR of between 15 and 39 bp and generate DR of 8 bp. They generally have a single *orf* of between 400 and 440 amino acids, which shows good alignment and includes an α-helical insertion domain (31).

However, IS*1096* harbors two *orf*s: the upstream *orf* exhibits similarities to the IS*L3* family Tpases; the second, *tnp*R, a MerR-like transcription factor, is related to *orf*s from *Agrobacterium rhizogenes* and *Rhizobium* sp. plasmids. TnpR appears to regulate transposition activity of IS*Ppu12* (219).

In IS*1167*, the reading frame appears to be distributed between two consecutive *orf*s with a potential for translational coupling suggested by overlapping initiation and termination codons (ATGA). Small sequences (130 to 340 bp) related to the IR of IS*1167* have been detected in *Streptococcus sanguis*, *Streptococcus pneumoniae*, and *Streptococcus agalactiae* (3, 417).

(iii) Mechanism

Transposition of most of these elements has been demonstrated, but no detailed analysis of their transposition mechanism has yet appeared. IS*1411* from *Pseudomonas putida* forms a circular species with abutted IR separated by 5 bp (220). Transposon circles are also formed by IS*Pst9* (221) and an isoform of IS*Ppu12* (219) and IS*31831* forms DNA species with a size expected for an excised transposon (222). There is some evidence indicating that transposition of these IS can be induced by a form of zygotic induction following conjugative transfer (219).

No obvious target sequence specificity has yet been observed although there is some suggestion that there may be a preference for AT-rich regions.

Tn*3*

(i) General

This represents a large and highly homogeneous group in terms of their transposition enzymes and terminal IR (Nicolas et al., this volume). Many are complicated in structure and include multiple antibiotic resistances, virulence and other "accessory" genes. These are often carried by another type of transposable element, integron cassettes (45). The Tn*3* family is included here because certain family members resemble IS (e.g., IS*1071*, IS*101*, IS*Xc4*/IS*Xc5*) encoding only the Tpase flanked by IR.

(ii) Organization

The IR are generally 38 bp long, start with the sequence GGGG (occasionally GAGG) and terminate internally with TAAG. The Tn*3* family encodes large (>900 amino acids) DDE Tpases with an α-helical insertion domain (31). Classical Tn*3* family members also encode a site-specific recombinase. The major differences between members of this family are in the number and location of the many passenger genes and in the type of site-specific recombinase present.

(iii) Mechanism

The replicative transposition mechanism of this family involves formation of a cointegrate in which donor and recipient replicons are fused and separated at each junction with a directly repeated transposon copy (223). These structures must be "resolved", by recombination between the two transposon copies, to generate the donor and target replicons each retaining a single transposon

copy. This is accomplished by a site-specific recombinase (resolvase), which acts at a unique DNA sequence, the Res site. Apart from passenger genes, the major difference between various Tn*3*-like elements is the nature of their resolvases (Nicolas et al., this volume).

IS*Azo13*

This family, represented by 37 members in ISfinder, emerged from the IS*NCY* orphan group. It is based on both Tpase and IR sequence similarities. Insertion generates a 3 bp AT-rich DR and the ends have a consensus GGa/g. Their Tpases are highly conserved with a probable DDE motif and an HTH motif at the N-terminus, which could function as a DNA binding domain. Two members encode two *orf*s with a possible programmed ribosomal frameshifting motif of 8 or 9 Å while the other members encode a unique *orf* which includes a triple lysine at the equivalent position (E. Gourbeyre unpublished).

IS WITH DEDD TRANSPOSASES

IS*110*

(i) General

At present, only a single IS family, IS*110*, is known to encode this type of DEDD enzyme, related to the RuvC Holliday junction resolvase (40). There are over 250 examples from nearly 130 bacterial and archaeal species.

The Tpase is closely related to the *Piv* and *MooV* invertases from *Moraxella lacunata/Moraxella bovis* (224, 225) and *Neisseria gonorrhoeae* (39, 226). *Piv* catalyzes inversion of a DNA segment permitting expression of a type IV pilin. However, the organization of IS*110* family members and the inversion systems are different. In the inversion systems, the recombinase is located outside the invertible segment, whereas it is located within the IS element (40).

The family includes two subgroups. It has been suggested that these may even represent two distinct families (114, 227): IS*110* and IS*1111*. Although their Tpases are very similar to those of the classical IS*110*, members of the second subgroup, IS*1111*, are distinguished from those of the IS*110* group principally by the presence of small (7 to 17 bp) subterminal IR. This would be the only family that is not defined by differences in the Tpase but by the nature of the IS ends.

However, the entire group of IS exhibit significant differences: some have subterminal IR whereas others do not; some appear to generate small DR while others do not; and the entire family show significant variations in their target preferences.

(ii) Organization

The organization of IS*110* family members is quite different from that of the DDE IS: they do not contain the typical terminal IR of the DDE IS and do not generally generate flanking target DR on insertion. This implies that their transposition occurs using a different mechanism to the DDE IS. They encode a single Tpase gene that spans the entire length of the IS. One characteristic which distinguishes IS*110* family members from all other elements whose Tpases exhibit a predicted RNase fold is that the predicted catalytic domain of their DEDD Tpases is located N-terminal to the DNA-binding domain (39, 228). In the DDE Tpases it is generally located upstream.

(iii) Mechanism

It has proved difficult to determine the activity of these Tpases in detail *in vitro*. Transposition of IS with DEDD Tpases may be unusual and involve HJ intermediates, which must be resolved using a RuvC-like mechanism. This type of recombination would be consistent with the close relationship between DEDD Tpases and the Piv/MooV invertases, which presumably resolve HJ structures during inversion (229). The difference in domain organization between the DEDD and DDE Tpases reinforces the idea that the two IS types possess a different transposition mechanism.

Members of this family produce double-strand circular transposon intermediates (e.g., IS*492*:(90), (230); IS*Pa11* (114); IS*Ec11*, (231); IS*117* (232, 233); IS*1383* (234)). However, it remains to be determined whether they use a copy-paste mechanism or whether the circular intermediate is formed by excision. For example, IS*492* was identified within the extracellular polysaccharide production (*eps*) gene of *Pseudoalteromonas atlantica* (90, 235). IS*492* clearly undergoes Tpase-dependent precise excision to regenerate a functional *eps* gene. Like many other IS that use double-strand circular intermediate, circle formation results in the assembly of a junction promoter from a –35 promoter element in the right end oriented outwards and a –10 promoter element in the left end oriented inwards (90). In the case of IS*1383*, the –10 component of the promoter appears to lie within the DNA sequence located between the abutted IS ends. This promoter presumably serves to drive Tpase expression for the final integration step (90, 234). For IS*1383*, amplification of a putative circle junction suggested that the abutted IS ends were separated by 10 bp composed of the 5 bp flanking each end in the original target site. For IS*492*, this is distance is 5 bp. The IS*492* copy within the *eps* gene is flanked by a 5-bp

DR that is required for Tpase-dependent precise excision and for transposition of the IS (90).

(iv) Insertion specificity

Another characteristic of the IS*110* family is their particular insertion specificities. IS*492*, IS*Ptu2*, and IS*Spi5* recognize a 7 bp sequence, 5′-CTTGTTA-3′, and duplicate the first 5 bp (CTTGT). Excision of the IS regenerates the original target sequence (236). In the case of IS*492*, the flanking 5 bp were also essential for formation of the dsDNA circular form (90).

At least six different members of the IS*1111* subgroup (IS*Kpn4*, IS*Pa21*, IS*Pst6*, IS*UnCu1*, IS*Azvi12*, and IS*Pa25*) show a preference for *attC* sequences of integrons (111) whereas others appear to prefer repeated extragenic palindrome (REP) sequences (109). Both targets are capable of assuming hairpin-like secondary structures. The integron *attC* is central to integration of circular integron cassettes (45) whereas REP sequences are small repeated extragenic palindromic sequences often present in many hundreds of copies in bacterial genomes and which play a variety of structural and regulatory roles (He et al., this volume). There are at least seven examples (IS*621*, IS*Pa11*, IS*Ppu9*, IS*Ppu10*, IS*Rm19*, IS*Psy7*, and IS*1594*) from both the classical IS*110* and IS*1111* groups, which have been identified as insertions into REP sequences. These are two types of insertion. In type 1, the IS inserts at the same position in the REP whereas type 2 insertions occur adjacent to a REP. Only one IS*1111* member, IS*Psy7*, has been identified with type 2 insertion specificity (109). Certain IS that insert into REP (e.g., IS*621*, IS*Ppu9*, IS*Ppu10* and IS*1594*) generate a duplication of 2 bp.

Other IS of this family also appear to insert into conserved target sequences: IS*1533* occurs in 84 copies in *Leptospira borgpetersenii* and inserts into a partially conserved sequence (ttAGACAAAA[IS*1533*]TATCAG agcc-gtct–aaa); IS*Rfsp2* from *Roseiflexus* sp. RS-1, present in 40 copies in the host genome, is flanked by the sequence, CTCtGCGaaCGCtGCGc[IS*Rfsp2*]CTCtGCG Gtg. Yet other IS of this family appear to target the ends of other IS. Hence IS*1383* present in six copies in *Pseudomonas putida* plasmids, inserts into the sequence TTCAGATGGT[IS1383]ATAAG contained within the end of another IS of the IS*5* family, IS*1384* (227, 234); IS*4321* and IS*5075* (both members of the IS*1111* group) target the ends of the transposon Tn*21*; and IS*Sba8* (IS*1111*) inserts into a REP-like sequence located in the end of the IS*3* family member, IS*Sba5*.

Therefore, as a general rule, IS*110* family elements appear to insert in a sequence-specific and oriented way (236).

IS WITH HUH TRANSPOSASES

There are two major HUH Tpase families: Y1 and Y2 enzymes (41) according to whether they carry one or two Y residues involved in catalysis. One (Y1) is associated with the IS*200*/IS*605* family. Although these enzymes use the same Y-mediated cleavage mechanism, they appear to carry out the transposition process in quite different ways. The TE that use this type of Tpase do not terminate in IR, which often makes it difficult to determine their ends by bioinformatics analysis unless there are a number of identical copies in a genome.

IS*91*

The IS*91* family is fairly homogeneous. The canonical IS*91* identified three decades ago (54) carries only a single *orf*, encoding an HUH Y2 Tpase with an N-terminal zinc finger motif (237) (see also reference 41). Both Y residues are necessary for transposition (238). The IS does not terminate with extensive IR but includes some potential secondary structure. Several other family members (e.g., IS*Azo26*, IS*CARN110*, IS*Mno23*, IS*Sde12*, IS*Shvi3*, IS*Sod25*, and IS*Wz1*) include a second upstream *orf* related to the phage integrase/Y-recombinase family. Its role, if any, in transposition remains unclear.

More recently, a second related group, the IS*CR* (IS with a Common Region) was identified (55) associated with multiple antibiotic-resistance genes both upstream and downstream of the Tpase gene. They form a distinct class. Although their Tpases are closely related to the IS*91* Tpase, they carry a single catalytic Y residue. However, it has yet to be demonstrated that IS*CR* indeed transpose.

IS*91* is thought to transpose using a rolling circle-type mechanism (239) involving an initiation event at one IS end (ori, 3′ to the Tpase), polarized transfer of the IS strand into a target molecule and termination at the second end (ter, 5′ to the Tpase) (238). Sequestration of flanking genes is proposed to occur when the termination mechanism fails and rolling circle transposition extends into neighboring DNA, where it may encounter a second surrogate end (238). This has been estimated to occur at a frequency of about 1%. Indeed, *ori* is essential for activity whereas removal of *ter* reduces but does not eliminate transposition.

Insertion of IS*91* is oriented with *ori* adjacent to the 3′ of a specific tetranucleotide target (5′-CTTG or 5′-GTTC), which is essential for further transposition (238, 239). In the transposition model, displacement of an IS*91* active transposon strand would be driven by leading strand replication of the donor replicon from a 3′

OH generated by cleavage at *ori*. The original model proposed that the cleaved IS end is transferred to the target DNA and the IS is replicated "into" the target replicon. However, it is difficult to explain the occurrence of single- and double-strand IS*91* circles on this model and their role remains to be determined. Clearly, since IS*CR* may play an important role in the assembly and transmission of multiple antibiotic resistance (55, 240) it is important to address the transposition mechanism of these elements.

IS*91* and IS*CR* Tpases and organization are distantly related to the eukaryotic Helitrons (241) (Thomas and Pritham, this volume).

IS*200*/IS*605*

(i) General

The IS*200*/IS*605* family (242–244) (He et al., this volume) is divided into three major groups based on the presence or absence of two reading frames, *tnpA* (encoding a Y1 Tpase) and *tnpB* (whose exact role is unknown) (245). The IS*200* group includes *tnpA* only, the IS*1341* (246) group carries only *tnpB* (and as such its status as an autonomous TE remains questionable), while the third group, IS*605*, carries both *tnpA* and *tnpB*. These can be expressed divergently or sequentially and, in some cases overlap slightly, suggesting translational coupling. TnpB is not required for transposition (245, 247) but may have a role in regulation because expression of the protein under control of inducible promoter from IS*Dra2* reduces transposition in its original host, *Deinococcus radiodurans*, and in *Escherichia coli* (248). TnpB has also been identified in the IS*607* family (below) not only in prokaryotes but also in eukaryotes (71, 249) where it is sometimes associated with other TE.

(ii) Mechanism

The IS*200*/IS*605* transposition mechanism is well understood from a combination of genetic, biochemical and structural studies (94, 100, 245, 250–255). Briefly, it occurs using a single-strand "peel-and-paste" mechanism (He et al., this volume) in which a specific single transposon strand (the "top" strand) is excised to form a single-strand transposon circle. This then inserts into a single-strand target. Family members include subterminal secondary structures, which are recognized by TnpA. The cleavage site occurs a short distance 5′ to the left and 3′ to the right of the structure. These are not directly recognized by TnpA but form a complex set of interactions with the internal sequence permitting their cleavage. Transposition occurs by insertion of the left end 3′ to a specific tetra- or penta-nucleotide, which is also essential for excision and further transposition. Insertion does not generate DR and occurs preferentially into the lagging strand template of replication forks. This results in a clear orientation bias at the genome level reflecting the direction of replication of the target replicon. This can be detected in numerous bacterial genomes. It is also possible that this family targets stalled replication forks (He et al., this volume).

IS WITH SERINE TRANSPOSASES

The serine recombinase family includes three groups: the resolvase/integrase group (whose activity has been well characterized); the large serine recombinases; and the serine Tpases (256). For the two former groups, the catalytic domain is invariably N-terminal followed by the sequence-specific DNA-binding domain—a simple HTH for the resolvase/invertase group, or a much larger domain of unknown structure in the large serine recombinases (257); (see chapters in the section *Conservative Site-Specific Recombination*).

IS*607*

(i) General

IS*607* was first identified in *Helicobacter pylori* (258). Family members encode a Tpase related to serine site-specific recombinase which uses serine as a nucleophile for cleavage of the DNA strand (223). This family is linked to the IS*200*/IS*605* family (above) by the presence of a TnpB analogue. IS*607*-like elements have been observed in the Mimi virus and other nucleocytoplasmic large DNA viruses (NCLDV) (140). Interestingly it appears to be one of the rare prokaryotic IS identified in eukaryotic genomes (71, 249).

(ii) Mechanism

Little is known about IS*607* family transposition although it is thought that circular intermediates are involved (N.D.F. Grindley pers. comm. cited in reference 63). The enzyme presumably catalyzes similar cleavages and strand transfers as its site-specific serine recombinase cousins using a transitory 5′ phosphor-serine covalent intermediate to excise a double-strand circular IS DNA copy. A mechanistic transposition model has been proposed based on Tpase structures from structural genomics studies and detailed knowledge of the general serine recombinase mechanism (256). This imagines a synaptic Tpase tetramer (as for classical serine recombinases) and explains the lack of IS*607* target specificity (258), unusual for this type of recombinase.

(iii) Organization

In contrast to the serine recombinases, the DNA-binding domain of the serine Tpases is located N-terminal to the catalytic domain (249, 256) in a similar way to that of DDE Tpases and may reflect a similar function: folding of the nascent peptide and co-translational binding to the IS ends (see *IS3 and IS481 families* section above)

ORPHAN IS

In addition to the major IS groupings, there are also several families for which the Tpase signature is not yet clear, either because there are a number of potential catalytic residues or because there are not a sufficient number of examples to define the highly conserved residues. In the former cases, definition of the important conserved residues will require experimental analysis. In the latter case, the IS are grouped as IS*NCY* (not classified yet), which includes small numbers of unclassified IS or orphans. Members of this group often emerge as families, or new distant groups of a known family, as more examples are added to the database. For example, both IS*1202 and ISDol1* have emerged as groups related to the IS*3*/IS*481* and IS*5* families, respectively.

Others include:

IS*892*: These are 1,600 bp long and represented by two sequences (with additional examples in the public databases). Their Tpases may be produced by frameshifting and include a Pfam MULE-like motif.

IS*Lbi1*: These are 1,400–1,500 bp long and represented by two sequences (with additional examples in the public databases) with a single *orf*, 30 bp IR and 5 bp DR.

IS*Mae2*: These are 1,400–2,400 bp long and represented by three sequences (with additional examples in the public databases) with a single *orf* carrying a potential DDE motif, IR and 9 bp DR. One of these, IS*Acif1*, includes a passenger gene of unknown function located downstream of the Tpase.

IS*Plu15*: These are 800–1,000 bp long and represented by two sequences (with additional examples in the public databases). They include a single *orf* and there are no apparent IR or DR, which makes definition of the IS ends difficult unless they are present in multiple copies.

IS*A1214*: These are archaeal-specific IS with a length of 1,000–1,200 bp. They are represented by five sequences with IR, with DR of 8 to 12 bp and a small *orf* upstream of the Tpase expressed in the opposite direction.

IS*C1217*: These archaeal-specific IS are 1,200 bp long and are represented by four sequences from the Sulfolobales with a single Tpase *orf*, IR, and DR of 6–8 bp.

IS*M1*: These archaeal-specific IS are 1,300–1,600 bp long and are represented by six sequences with low conservation and no clear DDE but with IR of about 24 bp, DR of 8 or 9 bp, and a single *orf*.

EUKARYOTIC TE AND PROKARYOTIC RELATIVES

Eukaryotic DNA transposons have been classified into "superfamilies" and, of those that have been analyzed in some detail, many have prokaryotic cousins. These relationships have been highlighted throughout this text. They are Tc1/mariner (IS*630*), Mutator(MuDR)/foldback (IS*256* and IS*Lre2*), PiggyBac (IS*1380*), PIF/Harbinger (IS*5*), Merlin (IS*1595*), Banshee (IS*3*/IS*481*), and Helitron (IS*91*). However, several elements with DDE transposases such as hAT, P, CACTA (En/Spm), Transib, Chapaev, Sola, Zator, and Ginger (148, 200) have yet to find prokaryotic cousins. It seems highly probable that, as the diversity of TE is explored more extensively aided by the massive accumulation of sequence data and the development of software designed to detect TE, more phylogenetic relationships between prokaryotic and eukaryotic TE will become evident.

CONCLUDING REMARKS

We have divided this review into two major sections. In one, we have attempted to present an overview of our current understanding of prokaryotic IS, their diversity in sequence, in organization and in mechanism, their distribution and impact on their host genome, and their relation to their eukaryotic cousins. We have discussed several IS-related TE that have been identified since the previous edition of *Mobile DNA*. This includes IS that use single-strand DNA intermediates and their related "domesticated" relations, and IS*CR* and ICE that use IS-related transposases for excision and integration. Several more specialized chapters in this volume include additional detailed information concerning a number of these topics. One of the major conclusions from this section is that the frontiers between the different types of TE are becoming less clear as more are identified.

In the second part, we have provided a detailed description of the expanding variety of IS, which we have divided into families for convenience. We emphasize that there is no "quantitative" measure of the weight of each of the criteria we use to define a family. Our perception of these families continues to evolve and families emerge regularly from the IS*NCY* "class" as more IS are added. This section is designed to be an aid and a source of information for consultation by interested specialist readers.

It is clear from this survey that there are many important and unanswered questions concerning both mechanistic and regulatory aspects of IS and the way in which they have and continue to spread and colonize their host genomes. An area of particular importance is in understanding the dynamics of TE-driven genome remodeling. A growing view is that the prokaryote genome is an ecological niche populated by a diverse collection of TE, in particular IS, including both endogenous elements and those acquired serendipitously by horizontal transfer. These form an integral part of a genomic landscape which is continuously modified by their activity. Understanding the evolutionary relationship between genomes and TE in this ecological niche and of the dynamics of TE-mediated genome remodeling is one of the fundamental challenges in genomics.

REFERENCES

1. Fiandt M, Szybalski W, Malamy MH. 1972. Polar mutations in lac, gal and phage lambda consist of a few IS- DNA sequences inserted with either orientation. *Mol Gen Genet* **119:**223–231.

2. Hirsch HJ, Saedler H, Starlinger P. 1972. Insertion mutations in the control region of the galactose operon of *E. coli*. II. Physical characterization of the mutations. *Mol Gen Genet* **115:**266–276.

3. Hirsch HJ, Starlinger P, Brachet P. 1972. Two kinds of insertions in bacterial genes. *Mol Gen Genet* **119:**191–206.

4. Saedler H, Heiss B. 1973. Multiple copies of the insertion-DNA sequences IS1 and IS2 in the chromosome of *E. coli* K-12. *Mol Gen Genet* **122:**267–277.

5. Reif HJ, Saedler H. 1974. IS1 is Involved in Deletion Formation in the gal Region of *E.coli* K12. *Mol Gen Genet* **137:**17–28.

6. Saedler H, Reif HJ, Hu S, Davidson N. 1974. IS2, a genetic element for turn-off and turn-on of gene activity in *E. coli*. *Mol Gen Genet* **132:** 265–289.

7. Hu S, Ohtsubo E, Davidson N. 1975. Electron microscopic heteroduplex studies of sequence relations among plasmids of *Escherichia coli*: structure of F13 and related F-primes. *J Bacteriol* **122:**749–763.

8. Barth PT, Datta N, Hedges RW, Grinter NJ. 1976. Transposition of a deoxyribonucleic acid sequence encoding trimethoprim and streptomycin resistances from R483 to other replicons. *J Bacteriol* **125:**800–810.

9. Hedges RW, Jacob AE. 1974. Transposition of ampicillin resistance from RP4 to other replicons. *Mol Gen Genet* **132:**31–40.

10. Heffron F, Sublett R, Hedges RW, Jacob A, Falkow S. 1975. Origin of the TEM-beta-lactamase gene found on plasmids. *J Bacteriol* **122:** 250–256.

11. Arber W, Iida S, Jutte H, Caspers P, Meyer J, Hanni C. 1979. Rearrangements of genetic material in *Escherichia coli* as observed on the bacteriophage P1 plasmid. *Cold Spring Harb Symp Quant Biol* **43**(Pt 2)**:**1197–1208.

12. So M, Heffron F, McCarthy BJ. 1979. The *E. coli* gene encoding heat stable toxin is a bacterial transposon flanked by inverted repeats of IS1. *Nature* **277:**453–456.

13. Nevers P, Saedler H. 1977. Transposable genetic elements as agents of gene instability and chromosomal rearrangements. *Nature* **268:**109–115.

14. McClintock B. 1956. Controlling elements and the gene. *Cold Spring Harb Symp Quant Biol* **21:**197–216.

15. Nyman K, Nakamura K, Ohtsubo H, Ohtsubo E. 1981. Distribution of the insertion sequence IS1 in gram-negative bacteria. *Nature* **289:**609–612.

16. Ohtsubo H, Nyman K, Doroszkiewicz W, Ohtsubo E. 1981. Multiple copies of iso-insertion sequences of IS1 in *Shigella dysenteriae* chromosome. *Nature* **292:**640–643.

17. Aziz RK, Breitbart M, Edwards RA. 2010. Transposases are the most abundant, most ubiquitous genes in nature. *Nucleic Acids Res* **38:**4207–4217.

18. Berg DE, Howe MM. 1989. *Mobile DNA*. American Society for Microbiology, Washington DC.

19. Craig NL, Craigie R, Gellert M, Lambowitz A. 2002. *Mobile DNA II*. American Society of Microbiology, Washington DC.

20. Siguier P, Gourbeyre E, Chandler M. 2014. Bacterial insertion sequences: their genomic impact and diversity. *FEMS Microbiol Rev* **38:**865–891.

21. Sharma V, Firth AE, Antonov I, Fayet O, Atkins JF, Borodovsky M, Baranov PV. 2011. A pilot study of bacterial genes with disrupted ORFs reveals a surprising profusion of protein sequence recoding mediated by ribosomal frameshifting and transcriptional realignment. *Mol Biol Evol* **28:**3195–3211.

22. Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M. 2006. ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res* **34:**D32–D36.

23. Campbell A, Berg DE, Botstein D, Lederberg EM, Novick RP, Starlinger P, Szybalski W. 1979. Nomenclature of transposable elements in prokaryotes. *Gene* **5:**197–206.

24. Mahillon J, Chandler M. 2000. Insertion sequence nomenclature. *ASM News* **66:**324.

25. Galas DJ, Chandler M. 1989. Bacterial insertion sequences, p 109–162. *In* Berg D, Howe M (ed), *Mobile DNA*. American Society for Microbiology, Washington DC.

26. Chandler M, Mahillon J. 2002. Insertion sequences revisited, p 305–366. *In* Craig NL, Craigie R, Gellert M, Lambowitz A (ed), *Mobile DNA*, vol II. ASM Press, Washington DC.

27. Mizuuchi K, Baker TA. 2002. Chemical mechanisms for mobilizing DNA, p 12–23. *In* Craig NL, Craigie R, Gellert M, Lambowitz A (ed), *Mobile DNA*, vol II. ASM press, Washington DC.

28. Fayet O, Ramond P, Polard P, Prere MF, Chandler M. 1990. Functional similarities between retroviruses and the IS3 family of bacterial insertion sequences? *Mol Microbiol* **4:**1771–1777.

29. Kulkosky J, Jones KS, Katz RA, Mack JP, Skalka AM. 1992. Residues critical for retroviral integrative recombination in a region that is highly conserved among retroviral/retrotransposon integrases and bacterial insertion sequence transposases. *Mol Cell Biol* **12:**2331–2338.

30. Khan E, Mack JP, Katz RA, Kulkosky J, Skalka AM. 1991. Retroviral integrase domains: DNA binding and the recognition of LTR sequences [published erratum appears in Nucleic Acids Res 1991 Mar 25;19(6):1358]. *Nucleic Acids Res* **19:**851–860.

31. Hickman AB, Chandler M, Dyda F. 2010. Integrating prokaryotes and eukaryotes: DNA transposases in light of structure. *Crit Rev Biochem Mol Biol* **45:**50–69.

32. Davies DR, Braam LM, Reznikoff WS, Rayment I. 1999. The three-dimensional structure of a Tn5 transposase-related protein determined to 2.9-A resolution. *J Biol Chem* **274:**11904–11913.

33. Hickman AB, Perez ZN, Zhou L, Musingarimi P, Ghirlando R, Hinshaw JE, Craig NL, Dyda F. 2005. Molecular architecture of a eukaryotic DNA transposase. *Nat Struct Mol Biol* **12:**715–721.

34. Curcio MJ, Derbyshire KM. 2003. The outs and ins of transposition: from Mu to Kangaroo. *Nat Rev Mol Cell Biol* **4:**865–877.

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

36. Feng X, Colloms SD. 2007. *In vitro* transposition of ISY100, a bacterial insertion sequence belonging to the Tc1/mariner family. *Mol Microbiol* **65:**1432–1443.

37. Plasterk RH. 1996. The Tc1/mariner transposon family. *Curr Top Microbiol Immunol* **204:**125–143.

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

39. Choi S, Ohta S, Ohtsubo E. 2003. A novel IS element, IS621, of the IS110/IS492 family transposes to a specific site in repetitive extragenic palindromic sequences in *Escherichia coli*. *J Bacteriol* **185:**4891–4900.

40. Buchner JM, Robertson AE, Poynter DJ, Denniston SS, Karls AC. 2005. Piv site-specific invertase requires a DEDD motif analogous to the catalytic center of the RuvC Holliday junction resolvases. *J Bacteriol* **187:**3431–3437.

41. Chandler M, de la Cruz F, Dyda F, Hickman AB, Moncalian G, Ton-Hoang B. 2013. Breaking and joining single-stranded DNA: the HUH endonuclease superfamily. *Nat Rev Microbiol* **11:**525–538.

42. Taylor KL, Churchward G. 1997. Specific DNA cleavage mediated by the integrase of conjugative transposon Tn916. *J Bacteriol* **179:**1117–1125.

43. Siguier P, Gagnevin L, Chandler M. 2009. The new IS1595 family, its relation to IS1 and the frontier between insertion sequences and transposons. *Res Microbiol* **160:**232–241.

44. Liebert CA, Hall RM, Summers AO. 1999. Transposon Tn21, Flagship of the Floating Genome. *Microbiol Mol Biol Rev* **63:**507–522.

45. Mazel D. 2006. Integrons: agents of bacterial evolution. *Nat Rev Microbiol* **4:**608–620.

46. Nakatsu C, Ng J, Singh R, Straus N, Wyndham C. 1991. Chlorobenzoate catabolic transposon Tn5271 is a composite class I element with flanking class II insertion sequences. *Proc Natl Acad Sci USA* **88:**8312–8316.

47. Burrus V, Waldor MK. 2004. Shaping bacterial genomes with integrative and conjugative elements. *Res Microbiol* **155:**376–386.

48. Adams V, Lyras D, Farrow KA, Rood JI. 2002. The clostridial mobilisable transposons. *Cell Mol Life Sci* **59:**2033–2043.

49. Pavlovic G, Burrus V, Gintz B, Decaris B, Guedon G. 2004. Evolution of genomic islands by deletion and tandem accretion by site-specific recombination: ICESt1-related elements from *Streptococcus thermophilus*. *Microbiology* **150:**759–774.

50. Brochet M, Da Cunha V, Couve E, Rusniok C, Trieu-Cuot P, Glaser P. 2009. Atypical association of DDE transposition with conjugation specifies a new family of mobile elements. *Mol Microbiol* **71:**948–959.

51. Guerillot R, Da Cunha V, Sauvage E, Bouchier C, Glaser P. 2013. Modular evolution of TnGBSs, a new family of integrative and conjugative elements associating insertion sequence transposition, plasmid replication, and conjugation for their spreading. *J Bacteriol* **195:**1979–1990.

52. Guerillot R, Siguier P, Gourbeyre E, Chandler M, Glaser P. 2014. The diversity of prokaryotic DDE transposases of the mutator superfamily, insertion specificity, and association with conjugation machineries. *Genome Biol Evol* **6:**260–272.

53. Smyth DS, Robinson DA. 2009. Integrative and sequence characteristics of a novel genetic element, ICE6013, in *Staphylococcus aureus*. *J Bacteriol* **191:**5964–5975.

54. Diaz-Aroca E, de la Cruz F, Zabala JC, Ortiz JM. 1984. Characterization of the new insertion sequence IS91 from an alpha-hemolysin plasmid of *Escherichia coli*. *Mol Gen Genet* **193:**493–499.

55. Toleman MA, Bennett PM, Walsh TR. 2006. ISCR Elements: Novel Gene-Capturing Systems of the 21st Century? *Microbiol Mol Biol Rev* **70:**296–316.

56. Toleman MA, Walsh TR. 2010. ISCR elements are key players in IncA/C plasmid evolution. *Antimicrob Agents Chemother* **54:**3534; author reply 3534.

57. Feschotte C, Zhang X, Wessler S. 2002. Miniature inverted repeat transposable elements and their relationship to established DNA transposons, p 1147–1158. *In* Craig NL, Craigie R, Gellert M, Lambowitz A (ed), *Mobile DNA*, vol II. ASM Press, Washington DC.

58. Dyall-Smith ML, Pfeiffer F, Klee K, Palm P, Gross K, Schuster SC, Rampp M, Oesterhelt D. 2011. *Haloquadratum walsbyi*: limited diversity in a global pond. *PLoS ONE* **6:**e20968.

59. Correia FF, Inouye S, Inouye M. 1988. A family of small repeated elements with some transposon-like properties in the genome of *Neisseria gonorrhoeae*. *J Biol Chem* **263:**12194–12198.

60. Buisine N, Tang CM, Chalmers R. 2002. Transposon-like Correia elements: structure, distribution and genetic exchange between pathogenic *Neisseria* sp. *FEBS Lett* **522:**52–58.

61. Oggioni MR, Claverys JP. 1999. Repeated extragenic sequences in prokaryotic genomes: a proposal for the origin and dynamics of the RUP element in *Streptococcus pneumoniae*. *Microbiology* **145**(Pt 10)**:**2647–2653.

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

63. Filee J, Siguier P, Chandler M. 2007. Insertion sequence diversity in archaea. *Microbiol Mol Biol Rev* **71:**121–157.

64. Chen Y, Braathen P, Léonard C, Mahillon J. 1999. MIC231, a naturally occurring mobile insertion cassette from *Bacillus cereus*. *Mol Microbiol* **32:**657–668.

65. De Palmenaer D, Vermeiren C, Mahillon J. 2004. IS231-MIC231 elements from *Bacillus cereus sensu lato* are modular. *Mol Microbiol* **53:**457–467.

66. Demerec M, Adelberg EA, Clark AJ, Hartman PE. 1966. A proposal for a uniform nomenclature in bacterial genetics. *Genetics* **54:**61–76.

67. Nagy Z, Chandler M. 2004. Regulation of transposition in bacteria. *Res Microbiol* **155:**387–398.

68. Dumesic PA, Madhani HD. 2014. Recognizing the enemy within: licensing RNA-guided genome defense. *Trends Biochem Sci* **39:**25–34.

69. Fedoroff NV. 2012. Transposable elements, epigenetics, and genome evolution. *Science* **338:**758–767.

70. Bikard D, Marraffini LA. 2013. Control of Gene Expression by CRISPR-Cas systems. *F1000Prime Rep* **5:**47.

71. Bao W, Jurka J. 2013. Homologues of bacterial TnpB_IS605 are widespread in diverse eukaryotic transposable elements. *Mob DNA* **4:**12.

72. Parkhill J, Sebaihia M, Preston A, Murphy LD, Thomson N, Harris DE, Holden MT, Churcher CM, Bentley SD, Mungall KL, Cerdeno-Tarraga AM, Temple L, James K, Harris B, Quail MA, Achtman M, Atkin R, Baker S, Basham D, Bason N, Cherevach I, Chillingworth T, Collins M, Cronin A, Davis P, Doggett J, Feltwell T, Goble A, Hamlin N, Hauser H, Holroyd S, Jagels K, Leather S, Moule S, Norberczak H, O'Neil S, Ormond D, Price C, Rabbinowitsch E, Rutter S, Sanders M, Saunders D, Seeger K, Sharp S, Simmonds M, Skelton J, Squares R, Squares S, Stevens K, Unwin L, et al. 2003. Comparative analysis of the genome sequences of *Bordetella pertussis*, *Bordetella parapertussis* and *Bordetella bronchiseptica*. *Nat Genet* **35:**32–40.

73. Preston A, Parkhill J, Maskell DJ. 2004. The Bordetellae: lessons from genomics. *Nat Rev Microbiol* **2:**379–390.

74. Touchon M, Rocha EP. 2007. Causes of insertion sequences abundance in prokaryotic genomes. *Mol Biol Evol* **24:**969–981.

75. Gil R, Belda E, Gosalbes MJ, Delaye L, Vallier A, Vincent-Monegat C, Heddi A, Silva FJ, Moya A, Latorre A. 2008. Massive presence of insertion sequences in the genome of SOPE, the primary endosymbiont of the rice weevil *Sitophilus oryzae*. *Int Microbiol* **11:**41–48.

76. Plague GR, Dunbar HE, Tran PL, Moran NA. 2008. Extensive proliferation of transposable elements in heritable bacterial symbionts. *J Bacteriol* **190:**777–779.

77. Moran NA, Plague GR. 2004. Genomic changes following host restriction in bacteria. *Curr Opin Genet Dev* **14:**627–633.

78. Andersson JO, Andersson SG. 1999. Insights into the evolutionary process of genome degradation. *Curr Opin Genet Dev* **9:**664–671.

79. Lawrence JG, Hendrix RW, Casjens S. 2001. Where are the pseudogenes in bacterial genomes? *Trends Microbiol* **9:**535–540.

80. Mira A, Ochman H, Moran NA. 2001. Deletional bias and the evolution of bacterial genomes. *Trends Genet* **17:**589–596.

81. Cerveau N, Leclercq S, Leroy E, Bouchon D, Cordaux R. 2011. Short- and long-term evolutionary dynamics of bacterial insertion sequences: insights from *Wolbachia* endosymbionts. *Genome Biol Evol* **3:**1175–1186.

82. Glansdorff N, Charlier D, Zafarullah M. 1981. Activation of gene expression by IS2 and IS3. *Cold Spring Harb Symp Quant Biol* **45**(Pt 1)**:**153–156.

83. Soki J, Eitel Z, Urban E, Nagy E. 2013. Molecular analysis of the carbapenem and metronidazole resistance mechanisms of *Bacteroides* strains reported in a Europe-wide antibiotic resistance survey. *Int J Antimicrob Agents* **41:**122–125.

84. Aubert D, Naas T, Heritier C, Poirel L, Nordmann P. 2006. Functional characterization of IS1999, an IS4 family element involved in mobilization and expression of beta-lactam resistance genes. *J Bacteriol* **188:**6506–6514.

85. Kieny M-P. 2012. *The evolving threat of antimicrobial resistance: options for action*. World Health Organization, Geneva.

86. McKenna M. 2013. The Last Resort. *Nature* **499:**394–396.

87. Mole B. 2013. Farming up trouble. *Nature* **499:**398–400.

88. Simons RW, Hoopes BC, McClure WR, Kleckner N. 1983. Three promoters near the termini of IS10: pIN, pOUT, and pIII. *Cell* **34:**673–682.

89. Ton-Hoang B, Betermier M, Polard P, Chandler M. 1997. Assembly of a strong promoter following IS911 circularization and the role of circles in transposition. *EMBO J* **16:**3357–3371.

90. Perkins-Balding D, Duval-Valentin G, Glasgow AC. 1999. Excision of IS492 requires flanking target sequences and results in circle formation in *Pseudoalteromonas atlantica*. *J Bacteriol* **181:**4937–4948.

91. Duval-Valentin G, Normand C, Khemici V, Marty B, Chandler M. 2001. Transient promoter formation: a new feedback mechanism for regulation of IS911 transposition. *EMBO J* **20:**5802–5811.

92. Prentki P, Teter B, Chandler M, Galas DJ. 1986. Functional promoters created by the insertion of transposable element IS1. *J Mol Biol* **191:**383–393.

93. Depardieu F, Podglajen I, Leclercq R, Collatz E, Courvalin P. 2007. Modes and modulations of antibiotic resistance gene expression. *Clin Microbiol Rev* **20:**79–114.

94. Guynet C, Achard A, Hoang BT, Barabas O, Hickman AB, Dyda F, Chandler M. 2009. Resetting the site: redirecting integration of an insertion sequence in a predictable way. *Mol Cell* **34:**612–619.

95. Mendiola MV, de la Cruz F. 1989. Specificity of insertion of IS91, an insertion sequence present in alpha-haemolysin plasmids of *Escherichia coli*. *Mol Microbiol* **3:**979–984.

96. Galas DJ, Calos MP, Miller JH. 1980. Sequence analysis of Tn9 insertions in the lacZ gene. *J Mol Biol* **144:**19–41.

97. Meyer J, Iida S, Arber W. 1980. Does the insertion element IS1 transpose preferentially into A+T- rich DNA segments? *Mol Gen Genet* **178:**471–473.

98. Sengstag C, Iida S, Hiestand-Nauer R, Arber W. 1986. Terminal inverted repeats of prokaryotic transposable element IS186 which can generate duplications of variable length at an identical target sequence. *Gene* **49:**153–156.

99. Peters JE, Craig NL. 2001. Tn7: smarter than we thought. *Nat Rev Mol Cell Biol* **2:**806–814.

100. Ton-Hoang B, Pasternak C, Siguier P, Guynet C, Hickman AB, Dyda F, Sommer S, Chandler M. 2010. Single-stranded DNA transposition is coupled to host replication. *Cell* **142:**398–408.

101. Hu WY, Derbyshire KM. 1998. Target choice and orientation preference of the insertion sequence IS903. *J Bacteriol* **180:**3039–3048.

102. Spradling AC, Bellen HJ, Hoskins RA. 2011. *Drosophila* P elements preferentially transpose to replication origins. *Proc Natl Acad Sci U S A* **108:**15948–15953.

103. Peters JE, Craig NL. 2000. Tn7 transposes proximal to DNA double-strand breaks and into regions where chromosomal DNA replication terminates. *Mol Cell* **6:**573–582.

104. Shi Q, Huguet-Tapia JC, Peters JE. 2009. Tn917 Targets the Region Where DNA Replication Terminates in *Bacillus subtilis*, Highlighting a Difference in Chromosome Processing in the Firmicutes. *J Bacteriol* **191:**7623–7627.

105. Garsin DA, Urbach J, Huguet-Tapia JC, Peters JE, Ausubel FM. 2004. Construction of an *Enterococcus faecalis* Tn917-mediated-gene-disruption library offers insight into Tn917 insertion patterns. *J Bacteriol* **186:**7280–7289.

106. Slater JD, Allen AG, May JP, Bolitho S, Lindsay H, Maskell DJ. 2003. Mutagenesis of *Streptococcus equi* and *Streptococcus suis* by transposon Tn917. *Vet Microbiol* **93:**197–206.

107. Swingle B, O'Carroll M, Haniford D, Derbyshire KM. 2004. The effect of host-encoded nucleoid proteins on transposition: H-NS influences targeting of both IS903 and Tn10. *Mol Microbiol* **52:**1055–1067.

108. Clement J-M, Wilde C, Bachellier S, Lambert P, Hofnung M. 1999. IS1397 Is Active for Transposition into the Chromosome of *Escherichia coli* K-12 and Inserts Specifically into Palindromic Units of Bacterial Interspersed Mosaic Elements. *J Bacteriol* **181:**6929–6936.

109. Tobes R, Pareja E. 2006. Bacterial repetitive extragenic palindromic sequences are DNA targets for Insertion Sequence elements. *BMC Genomics* **7:**62.

110. Wilde C, Escartin F, Kokeguchi S, Latour-Lambert P, Lectard A, Clement JM. 2003. Transposases are responsible for the target specificity of IS1397 and ISKpn1 for two different types of palindromic units (PUs). *Nucleic Acids Res* **31:**4345–4353.

111. Tetu SG, Holmes AJ. 2008. A Family of Insertion Sequences That Impacts Integrons by Specific Targeting of Gene Cassette Recombination Sites, the IS1111-attC Group. *J Bacteriol* **190:**4959–4970.

112. Post V, Hall RM. 2009. Insertion sequences in the IS1111 family that target the attC recombination sites of integron-associated gene cassettes. *FEMS Microbiol Lett* **290:**182–187.

113. Hallet B, Rezsohazy R, Delcour J. 1991. IS231A from *Bacillus thuringiensis* is functional in *Escherichia coli*: transposition and insertion specificity. *J Bacteriol* **173:**4526–4529.

114. Partridge SR, Hall RM. 2003. The IS1111 family members IS4321 and IS5075 have subterminal inverted repeats and target the terminal inverted repeats of Tn21 family transposons. *J Bacteriol* **185:**6371–6384.

115. Loot C, Turlan C, Chandler M. 2004. Host processing of branched DNA intermediates is involved in targeted transposition of IS911. *Mol Microbiol* **51:**385–393.

116. Reimmann C, Haas D. 1987. Mode of replicon fusion mediated by the duplicated insertion sequence IS21 in *Escherichia coli*. *Genetics* **115:**619–625.

117. Olasz F, Farkas T, Kiss J, Arini A, Arber W. 1997. Terminal inverted repeats of insertion sequence IS30 serve as targets for transposition. *J Bacteriol* **179:**7551–7558.

118. Prere MF, Chandler M, Fayet O. 1990. Transposition in *Shigella dysenteriae*: isolation and analysis of IS911, a new member of the IS3 group of insertion sequences. *J Bacteriol* **172:**4090–4099.

119. Parks AR, Li Z, Shi Q, Owens RM, Jin MM, Peters JE. 2009. Transposition into replicating DNA occurs through interaction with the processivity factor. *Cell* **138:**685–695.

120. Gomez MJ, Diaz-Maldonado H, Gonzalez-Tortuero E, Lopez de Saro FJ. 2014. Chromosomal replication dynamics and interaction with the beta sliding clamp determine orientation of bacterial transposable elements. *Genome Biol Evol* **6:**727–740.

121. Qi X, Daily K, Nguyen K, Wang H, Mayhew D, Rigor P, Forouzan S, Johnston M, Mitra RD, Baldi P, Sandmeyer S. 2012. Retrotransposon profiling of RNA polymerase III initiation sites. *Genome Res* **22:**681–692.

122. Shapiro JA. 1979. Molecular model for the transposition and replication of bacteriophage Mu and other transposable elements. *Proc Natl Acad Sci USA* **76:**1933–1937.

123. Ross DG, Swan J, Kleckner N. 1979. Nearly precise excision: a new type of DNA alteration associated with the translocatable element Tn10. *Cell* **16:**733–738.

124. Ross DG, Swan J, Kleckner N. 1979. Physical structures of Tn10-promoted deletions and inversions: role of 1400 bp inverted repetitions. *Cell* **16:**721–731.

125. Ohtsubo E, Zenilman M, Ohtsubo H. 1980. Plasmids containing insertion elements are potential transposons. *Proc Natl Acad Sci U S A* **77:**750–754.

126. Polard P, Seroude L, Fayet O, Prere MF, Chandler M. 1994. One-ended insertion of IS911. *J Bacteriol* **176:**1192–1196.

127. Mahillon J, Chandler M. 1998. Insertion sequences. *Microbiol Mol Biol Rev* **62:**725–774.

128. Robinson DG, Lee MC, Marx CJ. 2012. OASIS: an automated program for global investigation of bacterial and archaeal insertion sequences. *Nucleic Acids Res* **40:**e174.

129. Wagner A, Lewis C, Bichsel M. 2007. A survey of bacterial insertion sequences using IScan. *Nucleic Acids Res* **35:**5284–5293.

130. MacHattie LA, Jackowski JB. 1977. Physical Structure and Deletion Effects of the Chloramphenicol Resistance Element Tn9 in Phage Lambda, p 219–228. *In* Bukhari AI, Shapiro JA, Adhya SL (ed), *DNA Insertion Elements, Plasmids, and Episomes*. Cold Spring Harbour Laboratory, New York.

131. Chandler M, Silver L, Lane D, Caro L. 1979. Properties of an autonomous r-determinant from R100.1. *Cold Spring Harb Symp Quant Biol* **43**(Pt 2)**:**1223–1231.

132. Peterson BC, Rownd RH. 1985. Recombination sites in plasmid drug resistance gene amplification. *J Bacteriol* **164:**1359–1361.

133. Blinkowa AL, Walker JR. 1990. Programmed ribosomal frameshifting generates the *Escherichia coli* DNA polymerase III gamma subunit from within the tau subunit reading frame. *Nucleic Acids Res* **18:**1725–1729.

134. Matsutani S. 1994. Genetic evidence for IS1 transposition regulated by InsA and the delta InsA-B'-InsB species, which is generated by translation from two alternative internal initiation sites and frameshifting. *J Mol Biol* **240:**52–65.

135. Ton-Hoang B, Turlan C, Chandler M. 2004. Functional domains of the IS1 transposase: analysis *in vivo* and *in vitro*. *Mol Microbiol* **53:**1529–1543.

136. Turlan C, Chandler M. 1995. IS1-mediated intramolecular rearrangements: formation of excised transposon circles and replicative deletions. *EMBO J* **14:**5410–5421.

137. Ohta S, Tsuchida K, Choi S, Sekine Y, Shiga Y, Ohtsubo E. 2002. Presence of a characteristic D-D-E motif in IS1 transposase. *J Bacteriol* **184:**6146–6154.

138. Ohta S, Yoshimura E, Ohtsubo E. 2004. Involvement of two domains with helix-turn-helix and zinc finger motifs in the binding of IS1 transposase to terminal inverted repeats. *Mol Microbiol* **53:**193–202.

139. Parkhill J, Achtman M, James KD, Bentley SD, Churcher C, Klee SR, Morelli G, Basham D, Brown D, Chillingworth T, Davies RM, Davis P, Devlin K, Feltwell T, Hamlin N, Holroyd S, Jagels K, Leather S, Moule S, Mungall K, Quail MA, Rajandream MA, Rutherford KM, Simmonds M, Skelton J, Whitehead S, Spratt BG, Barrell BG. 2000. Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491. *Nature* **404:**502–506.

140. Filée J, Siguier P, Chandler M. 2007. I am what I eat and I eat what I am: acquisition of bacterial genes by giant viruses. *Trends Genet* **23:**10–15.

141. Achard A, Leclercq R. 2007. Characterization of a small mobilizable transposon, MTnSag1, in *Streptococcus agalactiae*. *J Bacteriol* **189:**4328–4331.

142. Feschotte C. 2004. Merlin, a new superfamily of DNA transposons identified in diverse animal genomes and related to bacterial IS1016 insertion sequences. *Mol Biol Evol* **21:**1769–1780.

143. Rousseau P, Normand C, Loot C, Turlan C, Alazard R, Duval-Valentin G, Chandler M. 2002. Transposition of IS911, p 366–383. *In* Craig NL, Craigie R, Gellert M, Lambowitz A (ed), *Mobile DNA II*. American Society of Microbiology, Washington DC.

144. Bhugra B, Dybvig K. 1993. Identification and characterization of IS1138, a transposable element from *Mycoplasma pulmonis* that belongs to the IS3 family. *Mol Microbiol* **7:**577–584.

145. Smith KS, Ingram-Smith C. 2007. *Methanosaeta*, the forgotten methanogen? *Trends Microbiol* **15:**150–155.

146. Petrova M, Shcherbatova N, Gorlenko Z, Mindlin S. 2013. A new subgroup of the IS3 family and properties of its representative member ISPpy1. *Microbiology* **159:**1900–1910.

147. Duval-Valentin G, Chandler M. 2011. Cotranslational control of DNA transposition: a window of opportunity. *Mol Cell* **44:**989–996.

148. Feschotte C, Pritham EJ. 2007. DNA Transposons and the Evolution of Eukaryotic Genomes. *Annu Rev Genet* **41:**331–368.

149. Morona JK, Guidolin A, Morona R, Hansman D, Paton JC. 1994. Isolation, characterization, and nucleotide sequence of IS1202, an insertion sequence of *Streptococcus pneumoniae*. *J Bacteriol* **176:**4437–4443.

150. De Palmenaer D, Siguier P, Mahillon J. 2008. IS4 family goes genomic. *BMC Evol Biol* **8:**18.

151. Rezsohazy R, Hallet B, Delcour J, Mahillon J. 1993. The IS4 family of insertion sequences: evidence for a conserved transposase motif. *Mol Microbiol* **9:**1283–1295.

152. Davies DR, Goryshin IY, Reznikoff WS, Rayment I. 2000. Three-dimensional structure of the Tn5 synaptic complex transposition intermediate. *Science* **289:**77–85.

153. Mazel D, Bernard C, Schwarz R, Castets AM, Houmard J, Tandeau de Marsac N. 1991. Characterization of two insertion sequences, IS701 and IS702, from the cyanobacterium *Calothrix* species PCC 7601. *Mol Microbiol* **5:**2165–2170.

154. Vilei EM, Nicolet J, Frey J. 1999. IS1634, a Novel Insertion Element Creating Long, Variable-Length Direct Repeats Which Is Specific for *Mycoplasma mycoides* subsp. *mycoides* Small-Colony Type. *J Bacteriol* **181:**1319–1323.

155. Klenchin VA, Czyz A, Goryshin IY, Gradman R, Lovell S, Rayment I, Reznikoff WS. 2008. Phosphate coordination and movement of DNA in the Tn5 synaptic complex: role of the (R)YREK motif 10.1093/nar/gkn577. *Nucl Acids Res* **36:**5855–5862.

156. Rieck B, Tourigny DS, Crosatti M, Schmid R, Kochar M, Harrison EM, Ou HY, Turton JF, Rajakumar K. 2012. *Acinetobacter* insertion sequence ISAba11 belongs to a novel family that encodes transposases with a signature HHEK motif. *Appl Environ Microbiol* **78:**471–480.

157. Zhang X, Jiang N, Feschotte C, Wessler SR. 2004. PIF- and Pong-like transposable elements: distribution, evolution and relationship with Tourist-like miniature inverted-repeat transposable elements. *Genetics* **166:**971–986.

158. Berg DE, Davies J, Allet B, Rochaix JD. 1975. Transposition of R factor genes to bacteriophage lambda. *Proc Natl Acad Sci USA* **72:**3628–3632.

159. Miriagou V, Carattoli A, Tzelepi E, Villa L, Tzouvelekis LS. 2005. IS26-Associated In4-Type Integrons Forming Multiresistance Loci in Enterobacterial Plasmids. *Antimicrob Agents Chemother* **49:**3541–3543.

160. Partridge SR, Zong Z, Iredell JR. 2011. Recombination in IS26 and Tn2 in the Evolution of Multiresistance Regions Carrying blaCTX-M-15 on Conjugative IncF Plasmids from *Escherichia coli*. *Antimicrob Agents Chemother* **55:**4971–4978.

161. Zhu Y-G, Johnson TA, Su J-Q, Qiao M, Guo G-X, Stedtfeld RD, Hashsham SA, Tiedje JM. 2013. Diverse and abundant antibiotic resistance genes in Chinese swine farms. *Proc Natl Acad Sci USA* **110:**3435–3440.

162. Kato K, Ohtsuki K, Mitsuda H, Yomo T, Negoro S, Urabe I. 1994. Insertion sequence IS6100 on plasmid pOAD2, which degrades nylon oligomers. *J Bacteriol* **176:**1197–1200.

163. Hall RM, Brown HJ, Brookes DE, Stokes HW. 1994. Integrons found in different locations have identical 5′ ends but variable 3′ ends. *J Bacteriol* **176:**6286–6294.

164. Sundin GW, Bender CL. 1995. Expression of the strA-strB streptomycin resistance genes in *Pseudomonas syringae* and *Xanthomonas campestris* and characterization of IS6100 in *X. campestris*. *Appl Environ Microbiol* **61:**2891–2897.

165. Simpson AE, Skurray RA, Firth N. 2000. An IS257-derived hybrid promoter directs transcription of a tetA(K) tetracycline resistance gene in the *Staphylococcus aureus* chromosomal mec region. *J Bacteriol* **182:** 3345–3352.

166. Bertini A, Poirel L, Bernabeu S, Fortini D, Villa L, Nordmann P, Carattoli A. 2007. Multicopy blaOXA-58 gene as a source of high-level resistance to carbapenems in *Acinetobacter baumannii*. *Antimicrob Agents Chemother* **51:**2324–2328.

167. Zienkiewicz M, Kern-Zdanowicz I, Carattoli A, Gniadkowski M, Ceglowski P. 2013. Tandem multiplication of the IS26-flanked amplicon with the bla(SHV-5) gene within plasmid p1658/97. *FEMS Microbiol Lett* **341:**27–36.

168. Loli A, Tzouvelekis LS, Tzelepi E, Carattoli A, Vatopoulos AC, Tassios PT, Miriagou V. 2006. Sources of diversity of carbapenem resistance levels in Klebsiella pneumoniae carrying blaVIM-1. *J Antimicrob Chemother* **58:**669–672.

169. Doublet B, Praud K, Weill FX, Cloeckaert A. 2009. Association of IS26-composite transposons and complex In4-type integrons generates novel multidrug resistance loci in *Salmonella* genomic island 1. *J Antimicrob Chemother* **63:**282–289.

170. Nigro SJ, Farrugia DN, Paulsen IT, Hall RM. 2013. A novel family of genomic resistance islands, AbGRI2, contributing to aminoglycoside resistance in *Acinetobacter baumannii* isolates belonging to global clone 2. *J Antimicrob Chemother* **68:**554–557.

171. Cullik A, Pfeifer Y, Prager R, von Baum H, Witte W. 2010. A novel IS26 structure surrounds blaCTX-M genes in different plasmids from German clinical *Escherichia coli* isolates. *J Med Microbiol* **59:**580–587.

172. Trieu-Cuot P, Courvalin P. 1985. Transposition behavior of IS15 and its progenitor IS15-delta: are cointegrates exclusive end products? *Plasmid* **14:**80–89.

173. Harmer CJ, Moran RA, Hall RM. 2014. Movement of IS26-associated antibiotic resistance genes occurs via a translocatable unit that includes a single IS26 and preferentially inserts adjacent to another IS26. *MBio* **5:**e01801–01814.

174. Riess G, Holloway BW, Puhler A. 1980. R68.45, a plasmid with chromosome mobilizing ability (Cma) carries a tandem duplication. *Genet Res* **36:**99–109.

175. Willetts NS, Crowther C, Holloway BW. 1981. The insertion sequence IS21 of R68.45 and the molecular basis for mobilization of the bacterial chromosome. *Plasmid* **6:**30–52.

176. Watson JM, Holloway BW. 1978. Chromosome mapping in *Pseudomonas aeruginosa* PAT. *J Bacteriol* **133:**1113–1125.

177. Berger B, Haas D. 2001. Transposase and cointegrase: specialized transposition proteins of the bacterial insertion sequence IS21 and related elements. *Cell Mol Life Sci* **58:**403–419.

178. Ammendola S, Politi L, Scandurra R. 1998. Cloning and sequencing of ISC1041 from the archaeon Sulfolobus solfataricus MT-4, a new member of the IS30 family of insertion elements [In Process Citation]. *FEBS Lett* **428:**217–223.

179. Dalrymple B. 1987. Novel rearrangements of IS30 carrying plasmids leading to the reactivation of gene expression. *Mol Gen Genet* **207:**413–420.

180. Rasmussen JL, Odelson DA, Macrina FL. 1987. Complete nucleotide sequence of insertion element IS4351 from *Bacteroides fragilis*. *J Bacteriol* **169:**3573–3580.

181. Rudant E, Courvalin P, Lambert T. 1998. Characterization of IS18, an Element Capable of Activating the Silent aac(6′)-Ij Gene of *Acinetobacter* sp. 13 Strain BM2716 by Transposition. *Antimicrob Agents Chemother* **42:**2759–2761.

182. Stalder R, Caspers P, Olasz F, Arber W. 1990. The N-terminal domain of the insertion sequence 30 transposase interacts specifically with the terminal inverted repeats of the element. *J Biol Chem* **265:**3757–3762.

183. Nagy Z, Szabó M, Chandler M, Olasz F. 2004. Analysis of the N-terminal DNA binding domain of the IS30 transposase. *Mol Microbiol* **54:**478–488.

184. Olasz F, Stalder R, Arber W. 1993. Formation of the tandem repeat (IS30)2 and its role in IS30- mediated transpositional DNA rearrangements. *Mol Gen Genet* **239:**177–187.

185. Olasz F, Farkas T, Stalder R, Arber W. 1997. Mutations in the carboxy-terminal part of IS30 transposase affect the formation and dissolution of (IS30)2 dimer. *FEBS Lett* **413:**453–461.

186. Kiss J, Olasz F. 1999. Formation and transposition of the covalently closed IS30 circle: the relation between tandem dimers and monomeric circles. *Mol Microbiol* **34:**37–52.

187. Kiss J, Szabo M, Olasz F. 2003. Site-specific recombination by the DDE family member mobile element IS30 transposase. *Proc Natl Acad Sci USA* **100:**15000–15005.

188. Szeverenyi I, Nagy Z, Farkas T, Olasz F, Kiss J. 2003. Detection and analysis of transpositionally active head-to-tail dimers in three additional Escherichia coli IS elements. *Microbiology* **149:**1297–1310.

189. Kiss J, Nagy Z, Toth G, Kiss GB, Jakab J, Chandler M, Olasz F. 2007. Transposition and target specificity of the typical IS30 family element IS1655 from *Neisseria meningitidis*. *Mol Microbiol* **63:**1731–1747.

190. Szabó M, Kiss J, Nagy Z, Chandler M, Olasz F. 2008. Sub-terminal Sequences Modulating IS30 Transposition *in Vivo* and *in Vitro*. *J Mol Biol* **375:**337–352.

191. Hwa V, Shoemaker NB, Salyers AA. 1988. Direct repeats flanking the Bacteroides transposon Tn4351 are insertion sequence elements. *J Bacteriol* **170:**449–451.

192. Brynestad S, Granum PE. 1999. Evidence that Tn5565, which includes the enterotoxin gene in Clostridium perfringens, can have a circular form which may be a transposition intermediate. *FEMS Microbiol Lett* **170:**281–286.

193. Machida Y, Sakurai M, Kiyokawa S, Ubasawa A, Suzuki Y, Ikeda JE. 1984. Nucleotide sequence of the insertion sequence found in the T-DNA

region of mutant Ti plasmid pTiA66 and distribution of its homologues in octopine Ti plasmid. *Proc Natl Acad Sci USA* **81:**7495–7499.

194. Gourbeyre E, Siguier P, Chandler M. 2010. Route 66: investigations into the organisation and distribution of the IS66 family of prokaryotic insertion sequences. *Res Microbiol* **161:**136–143.

195. Egelseer EM, Idris R, Jarosch M, Danhorn T, Sleytr UB, Sara M. 2000. ISBst12, a novel type of insertion-sequence element causing loss of S- layer-gene expression in *Bacillus stearothermophilus* ATCC 12980. *Microbiology* **146**(Pt 9)**:**2175–2183.

196. Han CG, Shiga Y, Tobe T, Sasakawa C, Ohtsubo E. 2001. Structural and functional characterization of IS679 and IS66-family elements. *J Bacteriol* **183:**4296–4304.

197. Haren L, Polard P, Ton-Hoang B, Chandler M. 1998. Multiple oligomerisation domains in the IS911 transposase: a leucine zipper motif is essential for activity. *J Mol Biol* **283:**29–41.

198. Dordet Frisoni E, Marenda MS, Sagné E, Nouvel LX, Guérillot R, Glaser P, Blanchard A, Tardy F, Sirand-Pugnet P, Baranowski E, Citti C. 2013. ICEA of *Mycoplasma agalactiae*: a new family of self-transmissible integrative elements that confers conjugative properties to the recipient strain. *Mol Microbiol* **89:**1226–1239.

199. Hennig S, Ziebuhr W. 2010. Characterization of the Transposase Encoded by IS256, the Prototype of a Major Family of Bacterial Insertion Sequence Elements. *J Bacteriol* **192:**4153–4163.

200. Yuan Y-W, Wessler SR. 2011. The catalytic domain of all eukaryotic cut-and-paste transposase superfamilies. *Proc Natl Acad Sci USA* **108:** 7884–7889.

201. Lyon BR, Gillespie MT, Skurray RA. 1987. Detection and characterization of IS256, an insertion sequence in *Staphylococcus aureus*. *J Gen Microbiol* **133:**3031–3038.

202. Lyon BR, May JW, Skurray RA. 1984. Tn4001: a gentamicin and kanamycin resistance transposon in *Staphylococcus aureus*. *Mol Gen Genet* **193:***554–556.*

203. Loessner I, Dietrich K, Dittrich D, Hacker J, Ziebuhr W. 2002. Transposase-dependent formation of circular IS256 derivatives in *Staphylococcus epidermidis* and *Staphylococcus aureus*. *J Bacteriol* **184:**4709–4714.

204. Eisen JA, Benito MI, Walbot V. 1994. Sequence similarity of putative transposases links the maize Mutator autonomous element and a group of bacterial insertion sequences. *Nucleic Acids Res* **22:**2634–2636.

205. Hua-Van Al, Capy P. 2008. Analysis of the DDE Motif in the Mutator Superfamily. *J Mol Evol* **67:**670–681.

206. Prudhomme M, Turlan C, Claverys JP, Chandler M. 2002. Diversity of Tn4001 transposition products: the flanking IS256 elements can form tandem dimers and IS circles. *J Bacteriol* **184:**433–443.

207. Feng X, Bednarz AL, Colloms SD. 2010. Precise targeted integration by a chimaeric transposase zinc-finger fusion protein. *Nucl Acids Res* **38:**1204–1216.

208. Cassier-Chauvat C, Poncelet M, Chauvat F. 1997. Three insertion sequences from the *Cyanobacterium synechocystis* PCC6803 support the occurrence of horizontal DNA transfer among bacteria. *Gene* **195:**257–266.

209. Dawson A, Finnegan DJ. 2003. Excision of the Drosophila mariner transposon mos1. Comparison with bacterial transposition and v(d)j recombination. *Mol Cell* **11:**225–235.

210. Richardson JM, Dawson A, O'Hagan N, Taylor P, Finnegan DJ, Walkinshaw MD. 2006. Mechanism of Mos1 transposition: insights from structural analysis. *EMBO J* **25:**1324–1334.

211. Tenzen T, Ohtsubo E. 1991. Preferential transposition of an IS630-associated composite transposon to TA in the 5′-CTAG-3′ sequence. *J Bacteriol* **173:**6207–6212.

212. Lopez de Felipe F, Magni C, de Mendoza D, Lopez P. 1996. Transcriptional activation of the citrate permease P gene of Lactococcus lactis biovar diacetylactis by an insertion sequence- like element present in plasmid pCIT264. *Mol Gen Genet* **250:**428–436.

213. Kato N, Yamazoe K, Han CG, Ohtsubo E. 2003. New insertion sequence elements in the upstream region of cfiA in imipenem-resistant *Bacteroides fragilis* strains. *Antimicrob Agents Chemother* **47:**979–985.

214. Podglajen I, Breuil J, Rohaut A, Monsempes C, Collatz E. 2001. Multiple mobile promoter regions for the rare carbapenem resistance gene of *Bacteroides fragilis*. *J Bacteriol* **183:**3531–3535.

215. de Las Rivas B, Marcobal A, Gomez A, Munoz R. 2005. Characterization of ISLpl4, a functional insertion sequence in *Lactobacillus plantarum*. *Gene* **363:**202–210.

216. Takemura H, Horinouchi S, Beppu T. 1991. Novel insertion sequence IS1380 from *Acetobacter pasteurianus* is involved in loss of ethanol-oxidizing ability. *J Bacteriol* **173:**7070–7076.

217. Dumontier S, Trieu-Cuot P, Berche P. 1998. Structural and Functional Characterization of IS1358 from *Vibrio cholerae*. *J Bacteriol* **180:**6101–6106.

218. Davis IJ, Roberts AP, Ready D, Richards H, Wilson M, Mullany P. 2005. Linkage of a novel mercury resistance operon with streptomycin resistance on a conjugative plasmid in *Enterococcus faecium*. *Plasmid* **54:**26–38.

219. Christie-Oleza JA, Nogales B, Lalucat J, Bosch R. 2010. TnpR encoded by an ISPpu12 isoform regulates transposition of two different ISL3-like insertion sequences in *Pseudomonas stutzeri* after conjugative interaction. *J Bacteriol* **192:**1423–1432.

220. Kallastu A, Hõrak R, Kivisaar M. 1998. Identification and Characterization of IS1411, a New Insertion Sequence Which Causes Transcriptional Activation of the Phenol Degradation Genes in *Pseudomonas putida*. *J Bacteriol* **180:**5306–5312.

221. Christie-Oleza JA, Nogales B, Martin-Cardona C, Lanfranconi MP, Alberti S, Lalucat J, Bosch R. 2008. ISPst9, an ISL3-like insertion sequence from *Pseudomonas stutzeri* AN10 involved in catabolic gene inactivation. *Int Microbiol* **11:**101–110.

222. Vertes A, Asai Y, Inui M, Kobayashi M, Yukawa H. 1995. The Corynebacterial insertion sequence IS31831 promotes formation of an excised transposon fragment. *Biotechnol Lett* **17:**1143–1148.

223. Grindley NDF. 2002. The movement of Tn3-like elements: transposition and cointegrate resolution, p 230–271. *In* Craig NL, Craigie R, Gellert M, Lambowitz A (ed), *Mobile DNA II*. ASM Press, Washington DC.

224. Fulks KA, Marrs CF, Stevens SP, Green MR. 1990. Sequence analysis of the inversion region containing the pilin genes of *Moraxella bovis*. *J Bacteriol* **172:**310–316.

225. Rozsa FW, Meyer TF, Fussenegger M. 1997. Inversion of *Moraxella lacunata* type 4 pilin gene sequences by a *Neisseria gonorrhoeae* site-specific recombinase. *J Bacteriol* **179:**2382–2388.

226. Skaar EP, Lecuyer B, Lenich AG, Lazio MP, Perkins-Balding D, Seifert HS, Karls AC. 2005. Analysis of the Piv recombinase-related gene family of *Neisseria gonorrhoeae*. *J Bacteriol* **187:**1276–1286.

227. Lauf U, Muller C, Herrmann H. 1999. Identification and characterisation of IS1383, a new insertion sequence isolated from *Pseudomonas putida* strain H [In Process Citation]. *FEMS Microbiol Lett* **170:**407–412.

228. Tobiason DM, Buchner JM, Thiel WH, Gernert KM, Karls AC. 2001. Conserved amino acid motifs from the novel Piv/MooV family of transposases and site-specific recombinases are required for catalysis of DNA inversion by Piv. *Mol Microbiol* **39:**641–651.

229. Tobiason DM, Lenich AG, Glasgow AC. 1999. Multiple DNA binding activities of the novel site-specific recombinase, Piv, from *Moraxella lacunata*. *J Biol Chem* **274:**9698–9706.

230. Higgins BP, Carpenter CD, Karls AC. 2007. Chromosomal context directs high-frequency precise excision of IS492 in *Pseudoalteromonas atlantica*. *Proc Natl Acad Sci USA* **104:**1901–1906.

231. Prosseda G, Latella MC, Casalino M, Nicoletti M, Michienzi S, Colonna B. 2006. Plasticity of the P junc promoter of ISEc11, a new insertion sequence of the IS1111 family. *J Bacteriol* **188:**4681–4689.

232. Henderson DJ, Lydiate DJ, Hopwood DA. 1989. Structural and functional analysis of the mini-circle, a transposable element of *Streptomyces coelicolor* A3(2). *Mol Microbiol* **3:**1307–1318.

233. Smokvina T, Henderson DJ, Melton RE, Brolle DF, Kieser T, Hopwood DA. 1994. Transposition of IS117, the 2.5 kb *Streptomyces coelicolor* A3(2) 'minicircle': roles of open reading frames and origin of tandem insertions. *Mol Microbiol* **12:**459–468.

234. Muller C, Lauf U, Hermann H. 2001. The inverted repeats of IS1384, a newly described insertion sequence from *Pseudomonas putida* strain H, represent the specific target for integration of IS1383. *Mol Genet Genom* **265:**1004–1010.

235. Bartlett DH, Silverman M. 1989. Nucleotide sequence of IS492, a novel insertion sequence causing variation in extracellular polysaccharide production in the marine bacterium *Pseudomonas atlantica*. *J Bacteriol* **171:**1763–1766.

236. Higgins BP, Popkowski AC, Caruana PR, Karls AC. 2009. Site-specific insertion of IS492 in *Pseudoalteromonas atlantica*. *J Bacteriol* **191:**6408–6414.

237. 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:**3521–3521.

238. Garcillan-Barcia MP, Bernales I, Mendiola MV, De la Cruz F. 2002. IS91 rolling circle transposition, p 891–904. *In* Craig NL, Craigie R, Gellert M, Lambowitz A (ed), *Mobile DNA*, vol II. ASM Press, Washington DC.

239. del Pilar Garcillan-Barcia M, Bernales I, Mendiola MV, de la Cruz F. 2001. Single-stranded DNA intermediates in IS91 rolling-circle transposition. *Mol Microbiol* **39:**494–501.

240. Toleman MA, Walsh TR. 2011. Combinatorial events of insertion sequences and ICE in Gram-negative bacteria. *FEMS Microbiol Rev* **35:**912–935.

241. Kapitonov VV, Jurka J. 2007. Helitrons on a roll: eukaryotic rolling-circle transposons. *Trends Genet* **23:**521–529.

242. Lam S, Roth JR. 1983. IS200: a Salmonella-specific insertion sequence. *Cell* **34:**951–960.

243. Kersulyte D, Krishnan BR, Berg DE. 1992. Nonrandom orientation of transposon Tn5supF insertions in phage lambda. *Gene* **114:**91–96.

244. Kersulyte D, Velapatino B, Dailide G, Mukhopadhyay AK, Ito Y, Cahuayme L, Parkinson AJ, Gilman RH, Berg DE. 2002. Transposable element ISHp608 of *Helicobacter pylori*: nonrandom geographic distribution, functional organization, and insertion specificity. *J Bacteriol* **184:**992–1002.

245. Ton-Hoang B, Guynet C, Ronning DR, Cointin-Marty B, Dyda F, Chandler M. 2005. Transposition of ISHp608, member of an unusual family of bacterial insertion sequences. *EMBO J* **24:**3325–3338.

246. Murai N, Kamata H, Nagashima Y, Yagisawa H, Hirata H. 1995. A novel insertion sequence (IS)-like element of the thermophilic bacterium PS3 promotes expression of the alanine carrier protein-encoding gene. *Gene* **163:**103–107.

247. Kersulyte D, Akopyants NS, Clifton SW, Roe BA, Berg DE. 1998. Novel sequence organization and insertion specificity of IS605 and IS606: chimaeric transposable elements of *Helicobacter pylori*. *Gene* **223:**175–186.

248. Pasternak C, Dulermo R, Ton-Hoang B, Debuchy R, Siguier P, Coste G, Chandler M, Sommer S. 2013. ISDra2 transposition in *Deinococcus radiodurans* is downregulated by TnpB. *Mol Microbiol* **88:**443–455.

249. Gilbert C, Cordaux R. 2013. Horizontal Transfer and Evolution of Prokaryote Transposable Elements in Eukaryotes. *Genome Biol Evol* **5:**822–832.

250. Ronning DR, Guynet C, Ton-Hoang B, Perez ZN, Ghirlando R, Chandler M, Dyda F. 2005. Active site sharing and subterminal hairpin recognition in a new class of DNA transposases. *Mol Cell* **20:**143–154.

251. Guynet C, Hickman AB, Barabas O, Dyda F, Chandler M, Ton-Hoang B. 2008. *In vitro* reconstitution of a single-stranded transposition mechanism of IS608. *Mol Cell* **29:**302–312.

252. Barabas O, Ronning DR, Guynet C, Hickman AB, Ton-Hoang B, Chandler M, Dyda F. 2008. Mechanism of IS200/IS605 family DNA transposases: activation and transposon-directed target site selection. *Cell* **132:**208–220.

253. Hickman AB, James JA, Barabas O, Pasternak C, Ton-Hoang B, Chandler M, Sommer S, Dyda F. 2010. DNA recognition and the precleavage state during single-stranded DNA transposition in *D. radiodurans*. *EMBO J* **29:**3840–3852.

254. He S, Hickman AB, Dyda F, Johnson NP, Chandler M, Ton-Hoang B. 2011. Reconstitution of a functional IS608 single-strand transpososome: role of non-canonical base pairing. *Nucleic Acids Res* **39:**8503–8512.

255. He S, Guynet C, Siguier P, Hickman AB, Dyda F, Chandler M, Ton-Hoang B. 2013. IS200/IS605 family single-strand transposition: mechanism of IS608 strand transfer. *Nucleic Acids Res* **41:**3302–3313.

256. Boocock MR, Rice PA. 2013. A proposed mechanism for IS607-family serine transposases. *Mob DNA* **4:**24.

257. Van Duyne GD, Rutherford K. 2013. Large serine recombinase domain structure and attachment site binding. *Crit Rev Biochem Mol Biol* **48:**476–491.

258. Kersulyte D, Mukhopadhyay AK, Shirai M, Nakazawa T, Berg DE. 2000. Functional organization and insertion specificity of IS607, a chimeric element of *Helicobacter pylori*. *J Bacteriol* **182:**5300–5308.

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