REVIEW ARTICLE

Bacterial insertion sequences: their genomic impact and diversity

Patricia Siguier, Edith Gourbeyre & Mick Chandler

Laboratoire de Microbiologie et Génétique Moléculaires, Unité Mixte de Recherche 5100, Centre National de Recherche Scientifique, Toulouse Cedex, France

Abstract

Insertion sequences (ISs), arguably the smallest and most numerous autonomous transposable elements (TEs), are important players in shaping their host genomes. This review focuses on prokaryotic ISs. We discuss IS distribution and impact on genome evolution. We also examine their effects on gene expression, especially their role in activating neighbouring genes, a phenomenon of particular importance in the recent upsurge of bacterial antibiotic resistance. We explain how ISs are identified and classified into families by a combination of characteristics including their transposases (Tpases), their overall genetic organisation and the accessory genes which some ISs carry. We then describe the organisation of autonomous and nonautonomous IS-related elements. This is used to illustrate the growing recognition that the boundaries between different types of mobile element are becoming increasingly difficult to define as more are being identified. We review the known Tpase types, their different catalytic activities used in cleaving and rejoining DNA strands during transposition, their organisation into functional domains and the role of this in regulation. Finally, we consider examples of prokaryotic IS domestication. In a more speculative section, we discuss the necessity of constructing more quantitative dynamic models to fully appreciate the continuing impact of TEs on prokaryotic populations.

Introduction

The idea that many prokaryotic genomes are mosaic, composed of a 'central genome backbone' of essential and house-keeping genes (the core genome) interspersed with DNA segments constituting the 'mobilome' (a variety of accessory genes that form part of the pan genome), is now common currency (Medini *et al.*, 2005; Tettelin *et al.*, 2008). The mobilome embraces several types of genetic unit which, as their collective name indicates, can move from place to place in a particular genome or from cell to cell. These mobile genetic elements (MGEs) can be divided into two major groups: those, such as plasmids and bacteriophages, that are transmissible from cell to cell (the intercellular MGEs), and those that cannot themselves undergo transfer but which are transferred following integration into members of the first group (the intracellular MGEs). Intracellular MGE or transposable elements (TEs) include transposons (Tn) and insertion

sequences (ISs) but can embrace integrons (In) and introns (Craig *et al.*, 2002). Originally, Tn were distinguished from ISs because they carry passenger (also called cargo) genes not involved in catalysing or regulating TE movement. Most eukaryotic DNA transposons have relatives among the prokaryotic ISs (see (Hickman *et al.*, 2010a, b)) and it is not surprising that a variety of these elements carrying passenger genes are now also being identified (Bao *et al.*, 2009; Bao & Jurka, 2013). Prokaryotes harbour a host of such elements as well as several types of structure possessing characteristics of both groups (e.g. integrative conjugative elements, ICEs, originally called conjugative transposons, and other types of genomic island) (Burrus & Waldor, 2004) (Dobrindt *et al.*, 2004; Guerillot *et al.*, 2013).

TEs insert into many different sites within a genome using mechanisms, which do not involve large regions of DNA homology between the TE and its target. In contrast to MGEs in eukaryotes, which include a large proportion

whose movement uses RNA intermediates, the vast majority of known prokaryotic TEs transpose using DNA intermediates.

This review will be limited to the simplest TEs: the ISs. ISs are classified into families (Fig. 1, Table 1) using a variety of characteristics (Mahillon & Chandler, 1998) (see ISfinder: www.is.biotoul.fr, below). This includes (1) the length and sequence of the short imperfect terminal inverted repeat sequences (IRs) carried by many ISs at their ends (TIRs or ITRs in eukaryotes); (2) the length and sequence of the short flanking direct target DNA repeats (DRs) (TSD, target site duplication, in eukaryotes) often generated on insertion; (3) the organisation of their open reading frames; or (4) the target sequences into which they insert. However, the principal factor in IS classification is the similarity, at the primary sequence level, of the enzymes which catalyse their movement, their transposases (Tpases) (see 'Major IS groups are defined by transposase type' below).

As ISs have been reviewed a number of times over the past years [e.g. (Mahillon & Chandler, 1998; Chandler & Mahillon, 2002; Curcio & Derbyshire, 2003; Hickman *et al.*, 2010a, b; Montano & Rice, 2011; Dyda *et al.*, 2012)], we have not included a detailed in-depth description of each different IS family. Instead, we first discuss IS distribution and impact on genome evolution and expression and explain how they are identified and classified into families by their Tpases and accessory genes. We then describe various mobile elements related to ISs to illustrate the growing recognition that the boundaries between different types of mobile element are becoming increasingly difficult to define as more are being identified. We also describe the different types of Tpase and their activities as well as their organisation into domains

and its role in regulation. Finally, we consider the few known examples of prokaryotic IS domestication.

Distribution

ISs are widespread and can occur in very high numbers in prokaryotic genomes. A recent study concluded that proteins annotated as Tpases, or as proteins with related functions, are by far the most abundant functional class in both the prokaryotic and eukaryotic genomic and metagenomic public data bases (Aziz *et al.*, 2010).

Since the last surveys [e.g. (Mahillon & Chandler, 1998; Chandler & Mahillon, 2002)], many new ISs have been identified largely as a result of the massive increase in available sequenced prokaryotic genomes. Careful analysis of a number of these has also revealed that some genomes contain significant levels of truncated and partial ISs devoid of Tpase genes. These genomic 'scars' represent traces of numerous ancestral transposition events. However, genome annotations are often based simply on the presence of Tpase genes and do not include the entire DNA sequence with the IS ends. Indeed, a significant number of solo IS-related IRs have been identified in various genomes (www.is.biotoul.fr; genome section). Small IS fragments are rarely taken into account even though they can provide important insights into the evolutionary history of the host genome. Not only can this seriously impair studies attempting to provide an overview of the evolutionary influence of TEs on bacterial and archeal genomes, but such fragments may encode truncated proteins and these could influence gene regulation [e.g. (Cordaux *et al.*, 2006; Liu *et al.*, 2007; Shaheen *et al.*, 2010)]. In bacteria (Gueguen *et al.*, 2006) (Stalder *et al.*, 1990) (Salvatore *et al.*, 2001) and eukaryotes (Vos
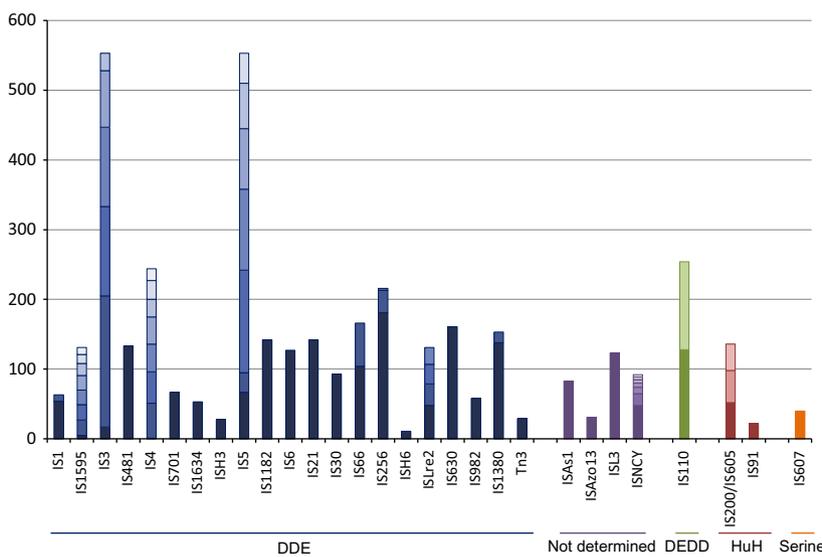
Fig. 1. Distribution of IS families in the ISfinder database. The histogram shows the number of IS of a given family, as defined in the text, in the ISfinder database (June 2013). The horizontal boxes indicate the number and relative size of different subgroups (see Table 1 for the subgroups names) within the family. They are grouped by colour to indicate the type of Tpase used: DDE, blue; undetermined, purple; DEDD, green; HUH, red; Serine, orange.

Table 1. General characteristics of IS families

Families	Subgroups	Typical size-range	DR (bp)	Ends	IRs	Nb ORF	Frameshift	Catalytic residus	Comments	Mechanism
IS1	–	740–1180	8–9	GGnnnTG	Y	2	ORFAB	DDE		Copy-and-paste
	Single ORF	800–1200	0–9			1				co-integrate
	ISMhu11	900–4600	0–10		Y	2	ORFAB			
IS1595	ISPna2	1000–1150	8	**GG**Cnn**TG**	Y	1		DDNK		Copy-and-paste ?
	ISPna2 + pass	1500–2600	8			1				
	ISH4	1000	8	**C**G**C**TCTT		1		DDNK		
	IS1016	700–745	7–9	**GG**Ggctg		1		DDEK		
	IS1595	900–1100	8	CcTGATT		1		DDNK + ER4R7		
	ISSod11	1000–1100	8	nnn**G**cnTATC		1		DDHK + ER4R7		
	ISNwi1	1080–1200	8	**gg**nnatTAT		1		DDEK + ER4		
	ISNwi1 + pass	1750–4750	8			1				
	ISNha5	3450–7900	8	**C**G**G**nn**TT**		1		DDER/K		
IS3	IS150	1200–1600	3–4	TG	Y	2	ORFAB	DDE	ORFB coordinates defined between two STOP codons since there are no inherent initiation signals and the orf is not translated alone.	Copy-paste
	IS407	1100–1400	4	TG						
	IS51	1000–1400	3–4	TG						
	IS3	1150–1750	3–4	TGa/g						
	IS2	1300–1400	5	TG						
IS481	–	950–1300	4–15	TGT	Y	1				Copy-paste ?
IS4	IS10	1200–1350	9	CT	Y	1		DDE	Hairpin intermediate	Cut-and paste
	IS50	1350–1550	8–9	C					Hairpin intermediate	
	ISPepr1	1500–1600	7–8	–T–AA						
	IS4	1400–1600	10–13	–AAT						
	IS4Sa	1150–1750	8–10	CA						
	ISH8	1400–1800	10	CAT						
	IS231	1450–5400	10–12	CAT		1 or +*			*Passenger genes	
IS701	–	1400–1550	4		Y	1		DDE		
ISH3	–	1225–1500	4–5	C–GT	Y	1		DDE		
IS1634	–	1500–2000	5–6	C	Y	1		DDE		
IS5	IS903	950–1150	9	GG	Y	1		DDE		
	ISL2	850–1200	2–3			1				
	ISH1	900–1150	8	–GC		1				
	IS5	1000–1500	4	Ga/g		1				
	IS1031	850–1050	3	GAa/g		1				
	IS427	800–1000	2–4	Ga/g		2	ORFAB			
IS1182	–	1330–1950	0–60		Y	1		DDE		
IS6	–	700–900	8	GG	Y	1		DDE		Co-integrate
IS21	–	1750–2600	4–8	TG	Y	2*		DDE	*istB : helper of transposition	
IS30	–	1000–1700	2–3		Y	1		DDE		Copy and paste

Table 1. Continued

Families	Subgroups	Typical size-range	DR (bp)	Ends	IRs	Nb ORF	Frameshift	Catalytic residus	Comments	Mechanism
IS66	–	2000–3000	8–9	GTAA	Y	3*		DDE*	*TnpC has the DDE domain	
	ISBst12	1350–1900	8–9	GTAA	Y	1		DDE		
IS91	–	1500–2000	0		N	1		HUH/Y2	Target site GAAC () CAAG	Rolling circle
IS110	–	1200–1550	0		N	1		DEDD		
	IS1111				Y*				*IRs not at the termini of the IS	
IS200/IS605	IS200	600–750	0		N	1*		HUH/Y1	*TnpA	Peel and paste
		1300–2000				2*		HUH/Y1**	*TnpA + TnpB; **Y1 on TnpA	
	IS1341	1200–1500				1*			*TnpB	
IS607	–	1700–2500	0		N	2*		Serine**	*TnpA + TnpB; **Y1 on TnpA	
IS256	–	1200–1500	8–9	Ga/g	Y	1		DDE		Copy-paste
IS630	–	1000–1400	2*		Y	1 or 2	ORFAB	DDE	*Target site : often NTAN with duplication of the TA	Cut and paste
IS982	–	1000	3–9	AC	Y	1		DDE		
IS1380	–	1550–2000	4–5	CC	Y	1		DDE		
ISAs1	–	1200–1500	8–10	CAGGG	Y	1				
ISL3	–	1300–2300	8	GG	Y	1				
Tn3	–	> 3000	0	GGGG	Y	> 1		DDE		Co-integrate
ISAzo13	–	1250–2200	0–4	Ga/g	Y	1				

IS families as defined in ISfinder are listed in related groups. Column 1: IS families; column 2: IS subgroups; column 3: typical length in base pairs; column 4: length of direct target repeats generated on insertion in base pairs; column 5: typical sequence of the IS ends; column 6: presence of imperfect terminal inverted repeats (Y = yes); column 7: presence of a frameshift; column 8: catalytic residus (DDE = OH as nucleophile; Y1 = single nucleophilic tyrosine; Y2 = two nucleophilic tyrosines; Ser = nucleophilic serine); column 9: mechanism if known; column 10: general comments; ? clearly means we do not know.

et al., 1993) (Rio, 1991), truncated transposases have been shown to inhibit transposition. One example where annotation of IS fragments has provided important information is in the obligatory intracellular insect endosymbiont, *Wolbachia*, which also carries high numbers of full-length ISs. The sequence divergence observed suggests that several waves of IS invasion and elimination have occurred over evolutionary time (Cerveau *et al.*, 2011).

Impact of ISs on genome evolution

ISs have had an important impact on genome structure and function. Several of these effects are considered in the following sections. In this context, it is useful to understand the time scales involved in these processes because they are often confounded. Evolutionary time is used to compare species (10^6 years), historical time in comparisons within or between populations (10^2–10^4 years), variety time in selection experiments (1–10^2 years) and laboratory time for ongoing events such as experimental measurement of transposition frequencies or in biochemical analyses (10^{-3}–1 years) (A. Schulman, pers. commun.). Thus, a 'burst' of transposition in evolutionary time is many orders of magnitude longer than a 'burst' of transposition in experimental biology.

IS expansion, elimination and genome streamlining

Perhaps, one of the most striking concerns IS expansion. ISs can undergo massive expansion and loss accompanied by gene inactivation and decay, genome rearrangement and genome reduction. Clearly, host lifestyle strongly influences these IS-mediated effects on genome structure, presumably by determining the level of genetic isolation of the microbial population. Factors affecting this include: whether the bacteria are ectosymbionts, primary endosymbionts having long evolutionary histories with their hosts, or secondary endosymbionts with more recent associations; whether they are transmitted in a strictly vertical manner or pass through a step of horizontal transfer via reinfection or passage through a second host vector (Bordenstein & Reznikoff, 2005; Moya *et al.*, 2008).

IS expansion has been commonly observed in bacteria with recently adopted fastidious, host-restricted lifestyles. Those which may have more ancient host-restricted lifestyles (e.g. *Wigglesworthia* in the Tsetse fly; *Buchnera aphidicola* in the aphid; *Blochmannia floridanus* in the ant) tend to possess small streamlined genomes with few pseudogenes or MGEs (see (Bordenstein & Reznikoff, 2005; Moya *et al.*, 2008); Supporting Information, Table S1).

One view is that IS expansion is an early step in this genome reduction process (Moran & Plague, 2004; Touchon & Rocha, 2007; Gil *et al.*, 2008; Plague *et al.*, 2008) (Fig. 2). This results from a decrease in strength and efficacy of purifying selection due to the shift from free to intracellular lifestyles (Moran & Plague, 2004). It is reinforced by a phenomenon known as Muller's ratchet, which leads to the irreversible accumulation of mutations in a confined intracellular environment (Moran, 1996; Andersson & Kurland, 1998; Silva *et al.*, 2003). In the nutritionally rich environment of the host, many genes of free-living bacteria are inessential. Enhanced genetic drift
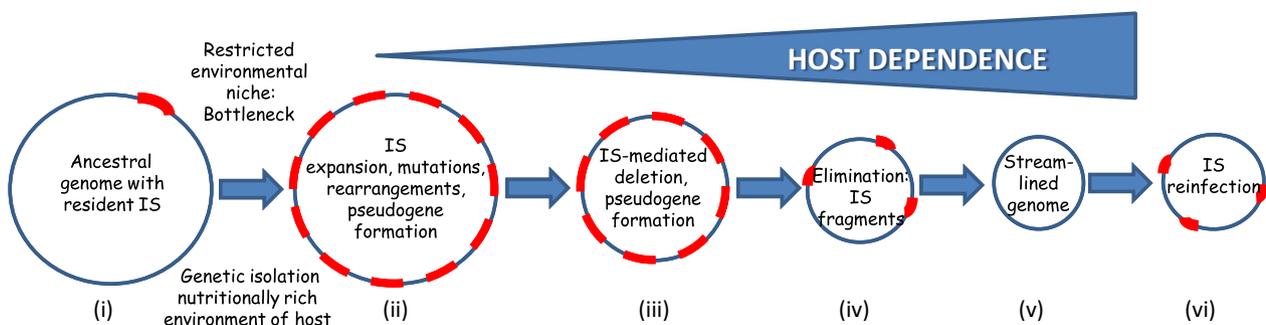
Fig. 2. IS expansion, elimination and genome 'streamlining'. The figure shows schematically from left to right events leading to the evolution of host-dependence in bacteria. (i) The parental (ancestral) chromosome including a low number of resident IS (red arcs). Note that the entire genome might also include transmissible plasmids carrying their own IS load, which can in principle undergo transposition into the chromosome. (ii) IS expansion occurs as a result of isolation and the formation of population bottlenecks within a host organism. This is accompanied by mutation promoted by insertion of new IS copies and by their related transposition activities of deletion and rearrangement. These genome rearrangements can also occur by homologous recombination between identical IS copies. (iii) With time IS will have a tendency to undergo deletion with adjacent DNA sequences in the absence of direct selection. This leads to a reduction in genome size. (iv) Eventually, extensive deletion will lead to the generation of nonautonomous IS fragments and their elimination. (vi) This gives rise to streamlined, IS 'free' genomes which may become 'reinfected' by IS on rare contact with other, IS-carrying strains or infection by IS-carrying bacteriophage (v).

would allow fixation of slightly deleterious mutations in the population, facilitated by the occurrence of successive population bottlenecks. The more genetically isolated the bacterial population, the more acute would be the effect. Indeed, many examples of this can be found among intracellular endosymbionts. This initial stage of transition from free-living to host-dependence would therefore result in the accumulation of pseudogenes, which will eventually be eliminated by so-called deletional bias (Mira et al., 2001). Clearly, the activities of MGEs, and of ISs in particular, make them important instruments in these processes. IS expansion would contribute to pseudogenisation by IS-mediated intrachromosomal recombination and genome reduction (Andersson & Andersson, 1999; Lawrence et al., 2001; Mira et al., 2001) by their capacities to generate deletions (Mahillon & Chandler, 1998). Such deletions would also eventually lead to complete or partial elimination of the ISs themselves. These processes are shown schematically in Fig. 2.

There are many striking examples of IS expansions in bacterial genomes (Table S1). The first to be identified was Shigella from the pregenomics era (Nyman et al., 1981; Ohtsubo et al., 1981). But IS expansion identified from sequenced genomes has been implicated in generating the present day Bordetella pertussis and B. parapertussis, Yersinia pestis, Enterococcus faecium, Mycobacterium ulcerans and many others. In at least some of these cases, it has been argued that large-scale genome rearrangements and deletions associated with IS expansion have improved the ability of the bacterium to combat host defences for example by changing surface antigens and regulatory circuitry (Parkhill & Thomson, 2003). This has been particularly well documented in the Bordetellae (Parkhill et al., 2003; Preston et al., 2004).

The phenomenon is also common among endosymbionts such as Wolbachia sp. These are considered ancient endosymbionts, which might be expected to possess more streamlined genomes. However, evidence has been presented that they have been subjected to several waves of invasion and elimination of ISs (Cerveau et al., 2011). This may be related to the fact that they are not strictly transmitted vertically but may also undergo relatively low levels of horizontal transmission and coinfection. Other symbionts or host-restricted bacteria also contain high IS loads. These include organisms such as Orientia tsutsugamushi, various Rickettsia, Sodalis glossinidius, Amoebophilus asiaticus 5a2, the γ1 symbiont of the marine oligochaete Olavius algarvensis, the Bacteroidete Cardinium hertigii, a symbiont of the parasitic wasp Encarsia pergandiella, and the primary symbionts of grain weevils. These obligate intracellular bacteria may carry intercellular MGEs such as phage (Hsia et al., 2000; Read et al., 2000) and conjugative elements (Blanc et al., 2007)

capable of acting as IS vectors and motors of horizontal gene transfer. Similar arguments might be used for other niche-restricted prokaryotes to explain increased IS loads found in some extremophiles (e.g. Sulfolobus solfataricus and certain cyanobacteria) (Brugger et al., 2004; Filee et al., 2007) (Papke et al., 2003; Allewalt et al., 2006).

Although IS expansion is generally assumed to occur stochastically over periods of evolutionary time, it has recently been observed that the Olavius algarvensis symbionts express significant levels of transposase (Kleiner et al., 2013). This raises the possibility that transposase expression is deregulated in this symbiont system. However, another symbiont, Amoebophilus asiaticus, with a high IS load, shows no evidence of recent transposition activity in spite of extensive IS transcription (Schmitz-Esser et al., 2011). In view of the time scales involved, only a very small but sustained increase in transposition activity might be needed to give rise to the high loads observed. Further exploration of the relationship between IS gene expression and transposition activity is clearly essential to understanding the dynamics of ISs in these and other systems.

Of course, different ISs are involved in different expansions and it is therefore important to understand IS diversity and properties. This is clearly evident in studies concerning the behaviour of IS on storage of bacterial strains where certain ISs appear more active than others (Naas et al., 1994, 1995).Their detailed effects on the host genome will depend on their particular transposition mechanisms. For example, IS target specificity will have profound effects on the way the host genome is shaped.

Target choice

The choice of a target DNA can have an important influence on the impact of TE on their host genomes. Initially, it was often assumed that TE show no or perhaps only low sequence specificity in their choice of a target. While this remains approximately true, the enormous increase in available sequences has provided more statistically robust data which have revealed that several TE use rather subtle mechanisms in choosing a target. Inspection of the public databases suggests that IS density is significantly higher in bacterial plasmids than in their host chromosomes and it seems likely that plasmids are major vectors in IS transmission (as well as in transmission of accessory traits such as resistance to antibacterials). Indeed, plasmids, in particular those that use a rolling circle mechanism either for replication or in conjugation, appear to be preferential targets for certain TE. For example, transposon Tn7 has two modes of transposition: in one, a specific sequence within the highly conserved glmS is recognised and insertion occurs next to this essential

gene (Craig, 2002); in the second, insertion occurs into replication forks directed by interactions with the β-clamp (Parks *et al.*, 2009). This results in a strong orientation bias of Tn7 insertions, consistent with insertion into the lagging strand of the replication fork formed during conjugative transfer. Although studies with IS are less advanced, a similar orientation bias was observed with IS903 (Hu & Derbyshire, 1998), suggesting that it too may use the β-clamp in directing insertions. It seems probable that many other IS use this type of protein–protein interaction.

A second example of specialised target choice was observed in members of the IS200/IS605 family (see Transposases and mechanism: *IS200/IS605-family transposases* below). These transpose using a strand-specific single-strand intermediate and insertion occurs 3′ to a tetra- or penta-nucleotide on the lagging strand (Ton-Hoang *et al.*, 2010). Clear vestiges of this specificity can still be detected in a large number of bacterial genomes where the orientation of insertion is strongly correlated with the direction of replication. There are clearly incidences of insertion in the 'wrong' orientation, but many of these may be explained by postinsertion genome rearrangements involving inversions. This would place the 'active' strand of the IS on the lagging rather than on the leading strand. Interestingly, those IS that are not oriented in the 'correct' orientation with respect to replication are almost certainly inactive and unable to transpose further (Ton-Hoang *et al.*, 2010). Replication fork targeting was subsequently observed in eukaryotes (Spradling *et al.*, 2011).

Other examples of sequence-specific target choice have been described. IS1, for example, shows a preference for regions rich in AT, whereas the transposon TnGBS and members of the related ISLre2 family show a preference for insertion 15–17 bp upstream of σA promoters (Brochet *et al.*, 2009; Guerillot *et al.*, 2013). Targeting of upstream regions of transcription units has also been extensively documented for certain eukaryotic transposons [e.g.(Qi *et al.*, 2012)].

There are also examples of IS (e.g. some members of the IS110, IS3 and IS4 families) which insert into potential secondary structures such as repeated extragenic palindromes (REP) (see The REP system below) (Clement *et al.*, 1999); (Wilde *et al.*, 2003), (Tobes & Pareja, 2006), integrons (Tetu & Holmes, 2008; Post & Hall, 2009) or even the ends of other TE) (Partridge & Hall, 2003) (Hallet *et al.*, 1991).

These examples represent only a small part of the literature concerning factors influencing target choice but serve to illustrate the impact this can have on genomes. Further notable aspects of the effects of target choice are considered in the next section.

Impact of ISs on genome expression

While massive IS-mediated genomic changes leading to streamlined genomes with increased pathogenicity and virulence are important and spectacular, they do not reflect the full impact of ISs. ISs can incorporate additional genes and subsequently act as vectors for these genes (see section 'ISs and relatives with passenger genes' below). Areas of topical relevance in this respect are the transmission of antibacterial resistance and virulence.

ISs also play more subtle roles. They can insert upstream of a gene and activate its expression, a phenomenon known for some time (Glansdorff *et al.*, 1981) and suspected for even longer (Reif & Saedler, 1974). This has recently received much attention due to the increase in resistance to various antibacterials (see (Aubert *et al.*, 2006; Soki *et al.*, 2013), which has become a worrying public health threat (Kieny, 2012; McKenna, 2013; Mole, 2013).

Activation of neighbouring gene expression can occur in two principal ways: either via promoters contained entirely within the IS driving transcripts that escape into neighbouring DNA (Glansdorff *et al.*, 1981), or by the formation of hybrid promoters following insertion. Many ISs contain outward oriented -35 promoter components in their ends and insertion at the correct distance from a suitable -10 box can generate a strong promoter [e.g. (Prentki *et al.*, 1986)]. This property has been noted for a very large number of ISs in a variety of bacterial strains. A preliminary survey of the literature, presented in Table S2, shows that the phenomenon is associated with over 30 different ISs and has occurred in at least 17 bacterial species (Depardieu *et al.*, 2007). Indeed, specific vector plasmids have been designed to identify activating insertions [e.g. (Szeverenyi *et al.*, 1996)].

IS activity can affect efflux mechanisms resulting in increased resistance: IS1 or IS10 insertion can up-regulate the AcrAB-TolC pump in *Salmonella enterica* (Olliver *et al.*, 2005); IS1 or IS2 insertion upstream of AcrEF (Jellen-Ritter & Kern, 2001; Kobayashi *et al.*, 2001) and IS186 insertional inactivation of the AcrAB repressor, AcrR, in *Escherichia coli* (Jellen-Ritter & Kern, 2001), all lead to increased resistance to fluoroquinolones. Insertional inactivation of specific porins can also play a significant role (Wolter *et al.*, 2004).

The IS families

Diversity and classification

IS classification is needed to cope with the high numbers and diversity of ISs. It also permits identification of the many IS fragments present in numerous genomes, contributes to understanding their effects on their host

genomes and can provide insights into their regulation and transposition mechanism. This role has been assumed by the ISfinder database (www-is.biotoul.fr) (Siguier *et al.*, 2006) following the closure of the Stanford repository (Lederberg, 1981).

IS identification

The families in ISfinder are defined using an initial manual BLAST analysis often followed by reiterative BLAST analyses with the primary transposase sequence of representative elements used as a query in a BLASTP (Altschul *et al.*, 1990) search of microbial genomes. Potential full-length Tpases are retained and that with the lowest score then used as a query in a second BLASTP search. This is continued until no new potential candidates are detected. The CLUSTALW multiple alignment algorithm (Thompson *et al.*, 1994) is then used and the results displayed using the Jalview alignment editor (Clamp *et al.*, 2004) for assessment. The corresponding DNA together with 1000 base pairs upstream and downstream is then extracted and examined manually for the IRs or other typical features such as secondary structures and flanking DRs. This, together with comparison of the DNA extremities of various elements, allows identification of both ends of the collected elements. In cases where more than a single IS copy is identified, BLASTN can be used to define the IS ends. Where only a single copy is found, the ends can often be defined by identifying and comparing with empty sites.

In a second step, we use the Markov cluster algorithm (MCL) (http://micans.org/mcl/) (Van Dongen, 2000; Enright *et al.*, 2002) to weigh the relationships between clusters of ISs and to validate prior ISfinder classification of ISs into families and subgroups (Siguier *et al.*, 2009). This is explained in detail in Siguier *et al.* (2009) and is based on the parameters used in the MCL in addition to characteristics, such as the specificity of target site duplications, the detailed sequence of the ends and genetic organisation. It should be understood that the distinction between families and subgroups can evolve as the number of ISs in the database increases.

The ISfinder repository contains over 4000 entries grouped into *c.* 26 families some of which can be conveniently divided into subgroups (Fig. 1, Table 1). This classification evolves continuously with the accumulation of additional ISs.

Several semi-automatic IS annotation pipelines are now available. The interested reader is directed to three of these: ISsaga (which is integrated into the ISfinder platform (Siguier *et al.*, 2006; Varani *et al.*, 2011), ISScan (Wagner *et al.*, 2007) and Oasis (Robinson *et al.*, 2012). At present, *de novo* prediction of ISs is not efficient and these pipelines all employ the ISfinder database to function. While all three pipelines permit identification of IS fragments as well as full-length ISs, a certain level of manual assessment is essential.

Major IS groups defined by transposase type

The primary difference between ISs is the nature of their transposases based on the type of chemistry they catalyse. These include DDE, DEDD, HUH and Ser transposases and are described in more detail in the following sections. Figure 1 shows the different families associated with each transposase type as of June 2013. For each family, the histogram is colour-coded to indicate the different subgroups. The majority of these are classical ISs and encode Tpases of the DDE superfamily (see 'Transposases' below).

It is important to note that this is certainly not an unbiased sample. ISs are generally identified by their similarity to those already in the database and rarely, as was the case in the pregenomic era, by their transposition activity (i.e. by analysis of mutations they produce by transposition). Moreover, IS inclusion in ISfinder has not involved a systematic global search of the public databases. We therefore emphasise that this distribution should not be taken as a true representation of the relative abundance of different IS families in prokaryotes.

ISs with DDE transposases

Classical ISs with DDE transposases (named for a conserved amino acid triad, Asp, Asp, Glu, the active site) are small (*c.* 0.7–2.5 kb long) genetically compact DNA segments with one or two open reading frames (a Tpase and possibly a protein involved in regulation). They end with imperfect IRs and generate short flanking DRs on insertion (Table 1). The DR length is specific for each IS type. There are clearly several dominant families among the DDE IS group. These include IS3, whose Tpases are perhaps the most closely related to the retroviral integrase catalytic core (IN) by the spacing of the DDE triad and by the appearance of additional conserved residues [e.g. (Haren *et al.*, 1999)]. The IS3 family contains several well-defined clades delimiting subgroups. With subsequent accumulation of additional members, it has been necessary to redefine several large families, such as IS4, into a number of individual families (De Palmenaer *et al.*, 2004). A similar situation has also occurred for the IS5 family (Table 1; Fig 1) whose original members are at present distributed over several families and subgroups. Some of these will certainly develop into separate families (P. Siguier, E. Gourbeyre and M. Chandler, unpublished).

In addition to the DDE encoding IS, three fundamentally different IS types have also been identified.

ISs with DEDD transposases

DEDD transposases (for Asp, Glu, Asp, Asp) are related to the Holiday junction resolvase, RuvC, itself related to DDE transposases. At present, only a single IS family, IS*110*, is known to encode this type of enzyme. The organisation of family members is quite different from that of the DDE ISs: they do not contain the typical terminal IRs of the DDE ISs and do not generate flanking target DRs on insertion. This implies that their transposition occurs using a different mechanism to the DDE ISs.

ISs with HUH Y1 or Y2 transposases

Two of these families encode a tyrosine (Y) Tpase. Note that these Tpases are not related to the well-characterised tyrosine site-specific recombinases such as phage integrases. Neither carries terminal IRs nor do they generate DRs on insertion. One family includes IS*91* (see below) (Zabala *et al.*, 1982; Garcillan-Barcia *et al.*, 2002), and the other includes IS*200*/IS*605* (Lam & Roth, 1983; Kersulyte *et al.*, 2002). Members of these families transpose using an entirely different mechanism to ISs with DDE transposases (del Pilar Garcillan-Barcia *et al.*, 2001; Ton-Hoang *et al.*, 2005) (see below). These ISs carry subterminal sequences, which are able to forms hairpin secondary structures. This is particularly marked in the IS*200*/IS*605* family elements.

These ISs are defined by the presence of a Tpase belonging to the HUH endonuclease superfamily (named for the conserved active site amino acid residues H = Histidine and U = large hydrophobic residue). There are two major HUH Tpase families: Y1 and Y2 enzymes (Chandler *et al.*, 2013) (see 'Transposases and Mechanism' below) according to whether they carry one or two Y residues involved in catalysis. One (Y1) is associated with the IS*200*/IS*605* family (Ton-Hoang *et al.*, 2005). The second (Y2) is associated with the IS*91* insertion sequence family (Mendiola & de la Cruz, 1992), with a related and newly defined group, the ISCR (Toleman *et al.*, 2006) (see 'IS*91*-related ISCRs' below) and with eukaryotic helitrons (Kapitonov & Jurka, 2007). Although these enzymes use the same Y-mediated cleavage mechanism, they appear to carry out the transposition process in quite different ways (see 'Transposases and Mechanism' below).

While the IS*91* family is fairly homogenous, the IS*200*/IS*605* family is divided into three major subgroups, IS*200*, IS*1341*, and those that resemble IS*605*. This is based on the presence or absence of two reading frames *tnpA* and *tnpB*, which can occur individually or together in several configurations (Ton-Hoang *et al.*, 2005) (Fig. 3): *tnpA* encodes the Tpase and *tnpB* encodes a
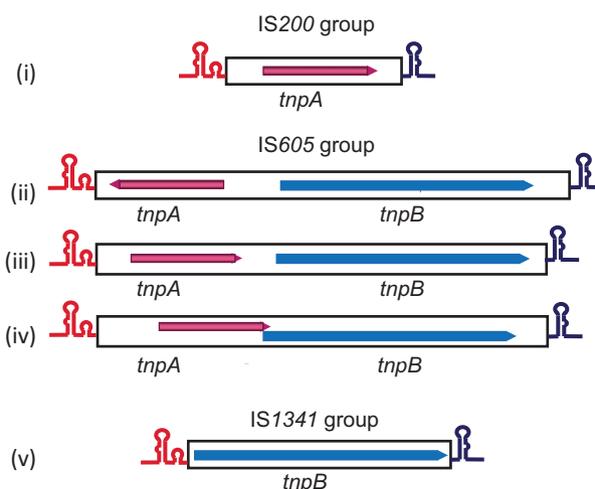
Fig. 3. Organisation of the IS*200*/IS*605* family. (i) IS*200* group with *tnpA* alone: examples include IS*200* (*Salmonella typhimurium*, *Escherichia coli*), IS*1541* (*Yersinia pseudotuberculosis*, *Yersinia pestis*), IS*1469* (*Clostridium perifringens*) and IS*W1* (*Wolbachia* sp.). (ii) IS*605*/IS*606* type with *tnpA* and *tnpB* in a divergent orientation: IS*605*, IS*606* (*Helicobacter pylori*); IS*Ljo5* (*Lactobacillus johnsonii*). (iii) IS*8301* type with nonoverlapping *tnpA* and *tnpB* orfs in the same direction: IS*Dra2* (*Deinococcus radiodurans*), IS*H1-8* (*Halobacterium*), IS*Efa4* (*Enterococcus faecium*), IS*1253* (*Dichelobacter nodosus*). (iv) IS*608* with overlapping *tnpA* and *tnpB*: IS*608* (*Helicobacter pylori*). (v) IS*1341* (*Thermophilic bacterium*) group with *tnpB* alone. Left and right are represented as hairpin structures in red and blue, respectively. Orfs are indicated as boxes with arrowheads (showing the direction of translation). *tnpA* is shown in red and *tnpB* in blue.

protein with a possible role in regulation (Pasternak *et al.*, 2013) (see 'IS with Accessory genes' below).

ISs with serine transposases

The third family is represented by IS*607*, which carries a Tpase closely related to serine recombinases such as the resolvases of Tn*3* family elements (see below). Little is known about their transposition mechanism. However, it appears likely, in view of the known activities of resolvases (Grindley, 2002), that IS*607* transposition may involve a circular double-strand DNA intermediate [N.D.F. Grindley cited as pers. commun. in (Filee *et al.*, 2007)]. Members of this family also occur in a group of giant eukaryotic viruses, the nucleocytoplasmic large DNA viruses (NCLDV). These infect protists whose lifestyle involves grazing on various bacteria (Filée *et al.*, 2007). They have presumably been incorporated into the viral genomes together with other bacterial genes by virtue of the highly promiscuous recombination properties of the virus. Interestingly, copies of IS*607*-like elements have more recently been identified in several 'lower' eukaryote genomes (Rolland *et al.*, 2009; Gilbert & Cordaux, 2013).

Most are incomplete although complete copies were identified in the genomes of the protist *Acanthamoeba castellani* and the alga *Ectocarpus siliculosus* (Gilbert & Cordaux, 2013) (Boocock & Rice, 2013). These are closely related to the cyanobacterial IS*Arma1*. In addition to the transposase, members of this family also sometimes include a *tnpB* gene similar to those carried by the IS*200*/IS*605* family (Kersulyte *et al.*, 2000).

Although there are isolated incidences of prokaryotic ISs identified in eukaryote genomes [e.g. IS*5* in Rotifers; (Gladyshev & Arkhipova, 2009)], IS*607* is at present the only example of a prokaryotic IS identified in multiple eukaryotic genomes. However, as is probable in the case of NCLDV, it is possible that these have been inherited as part of a larger DNA segment rather than by transposition (Filée *et al.*, 2007).

Orphan ISs

In addition to these three major IS groupings, there are also several families for which the transposase signature is not clear. Although many have a potential DDE motif (e.g. IS*L3*, IS*As1* or the newly defined IS*Azo13* family), experimental proof will be required to confirm the importance of these residues in catalysis. Another group, IS*NCY* (<u>n</u>ot <u>c</u>lassified <u>y</u>et), is composed of small numbers of unclassified ISs 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.

ISs with accessory genes

Many ISs also include accessory genes involved in regulating their transposition. These are relatively specific for each IS family and thus also serve in definition of the family. However, in many cases, the exact role and activity of the gene product is unclear.

IS*21*

IS*21* family members encode a 'helper' gene, *istB* (Berger & Haas, 2001) 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. The molecular details of IstB activity are not known.

IS*200*/IS*605* and IS*607*

Although IS*200*/IS*605* and IS*607* family members carry very different types of transposase (see above), they often include a second *orf*, *tnpB*, in addition to their transposases (Fig. 2). TnpB is not required for transposition either *in vitro* or *in vivo* (Kersulyte *et al.*, 1998; Ton-Hoang *et al.*, 2005). However, it has been observed to reduce IS*Dra2* (IS*200*/IS*605* family) transposition activity both in its original host, *Deinococcus radiodurans,* and in *E. coli* (Pasternak *et al.*, 2013). The molecular details of TnpB activity are not known.

Full-length TnpB includes three domains, an N-terminal HTH, a central domain and a C-terminal zinc finger (ZF) domain (Pasternak *et al.*, 2013). However, this TnpB configuration is quite variable and there are a large number which appear to be undergoing decay. TnpB analogues have been identified as part of IS*607*-like elements in the Mimi virus and other NCLDV (Filée *et al.*, 2007). They have recently been identified in a variety of eukaryotic genomes sometimes associated with other TEs (Bao & Jurka, 2013; Gilbert & Cordaux, 2013).

IS*66*

In addition to its DDE transposase gene, *tnpC*, IS*66* can include two additional genes, *tnpA* and *tnpB*, whose function is as yet unknown (Han *et al.*, 2001) (Fig. 4a). The three reading frames are disposed in a pattern suggesting translational coupling: *tnpB* is in general in translational reading frame -1 compared to *tnpA* and in most cases the termination codon of *tnpA* and the initiation codon of *tnpB* overlap (ATGA).

IS*91*

While the canonical IS*91* carries only a single *orf*, encoding an HUH Y2 transposase (Chandler *et al.*, 2013), several other family members (e.g. IS*Azo26*; IS*CARN110*; IS*Mno23*; IS*Sde12*; IS*Shvi3*; IS*Sod25* and IS*Wz1*) include a second *orf* located upstream. This *orf* is a Y-recombinase related to the phage integrase family whose role in transposition remains to be determined.

Tn*3* family

The Tn*3* family is an extensive group of transposons which encode large (> 900 aa) DDE transposases. They are included here because certain family members resemble ISs (e.g. IS*1071*) as they encode only the transposase. However, 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 (Grindley *et al.*, 2006). 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
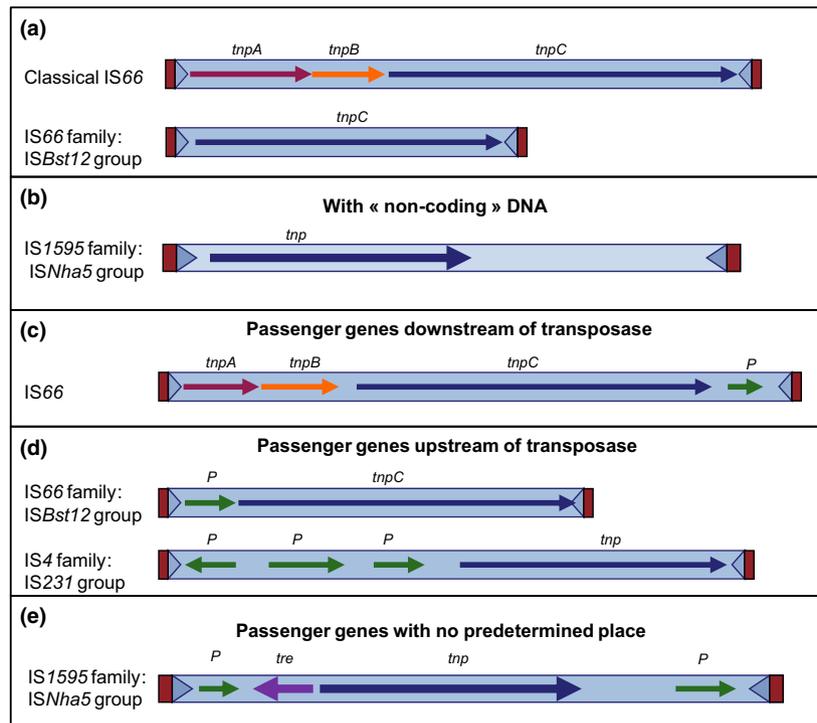
Fig. 4. tIS: IS with passenger genes. The IS is represented as a rectangle with flanking direct repeats (DR) in red and terminal inverted repeats in blue (triangles). The Tpase orfs are shown in dark blue and passenger genes in green. (a) Organisation of a classical IS66 family member and of the ISBst12 group. The 'accessory' genes tnpA and tnpB are shown in red and orange, respectively. (b) An IS1595 family member with noncoding DNA. (c) A tIS66 with a single orf downstream of the Tpase. (d) IS66 and IS4 tIS with passenger gene(s) upstream of the Tpase. (e) An IS1595 family tIS with upstream and downstream passenger genes.

'resolvase', a site-specific recombinase which acts at a unique DNA sequence in the transposon, the Res site. While many Tn3 family members encode a serine recombinase, several are now known to carry a tyrosine site-specific recombinase (TnpI in Tn4655 and Tn4330) (Vanhooff et al., 2006) resembling phage integrases. Moreover, a third group of Tn3 family members include two genes, TnpS and TnpT (Yano et al., 2013). It is possible that, as for the bacteriophage λ which uses a Y site-specific recombinase (Int) for integration together with a second protein, Xis for excision, TnpT is involved in assuring directionality in transposition. There is no evident difference in the Tpases of these two Tn3-like groups.

IS derivatives

We have already indicated that the classical IS model (a single *orf* encoding a DDE transposase, two terminal IRs and flanking DRs) is not universal and that other models such as the IS200/IS605 and IS91 families exist and are also widespread (see 'Major IS groups are defined by transposase type' above). Below, we describe a variety of IS-related TEs which share different levels of similarity with ISs. These include both autonomous (encoding a Tpase) and nonautonomous TEs (lacking a Tpase and whose transposition requires the Tpase of a related element in the same cell) and TEs with passenger genes not implicated in transposition or its regulation (Figs 4 and 5).

ISs and relatives with passenger genes

An increasing number of ISs carrying passenger genes are being identified. These include genes for transcription regulators, methyltransferases and antibiotic resistance (see Fig. 4). They can be located upstream, downstream or on both sides of the transposase gene (Fig. 4c, d and e). These elements, simpler than known Tns, are called transporter ISs (tISs) (Siguier et al., 2009). They can be significantly longer than typical ISs (e.g. ISCausp2, 7915 bp). Several of these 'extended' ISs include a significant length of DNA with no clear coding capacity (e.g. ISBse1, ISSpo3, ISSpo8) (Fig. 4b). At least some incomplete ISs presently identified in sequenced genomes may be of this type because the second IS end would occur at an unexpectedly distant position and would not necessarily have been identified.

tISs are generally present in low copy number. Many occur only in single copy in a given genome raising the question of whether they are active. Moreover, more than one closely related but nonidentical derivative can be found in a single genome (e.g. ISSpo3, and ISSpo8 in Silicibacter pomeroyi DSS-3 and ISNwi4 and ISNha5 in Nitrobacter winograski Nb-255). Others are present in more than one copy. ISPre3, an IS66 relative from Pseudomonas resinovorans plasmid pCAR1 includes a hypothetical protein and is present as 2 copies with different insertion sites as judged by their typical 8 bp DRs (E. Gourbeyre, unpublished).
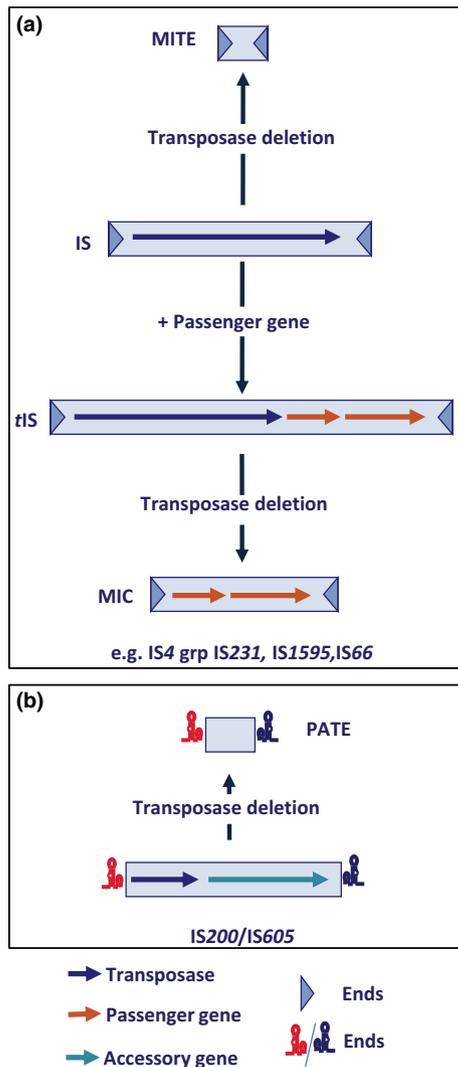
Fig. 5. Relationship between IS, tIS and MITES. (a) The (hypothetical) relationship between different IS derivatives (shown as light blue boxes). Horizontal arrows indicate open reading frames encoding the Tpase (dark blue), passenger genes (orange). The terminal inverted repeats are shown as darker blue triangles. Examples of IS families which include such derivatives are indicated at the bottom of the panel. (b) The particular case of MITES derived from IS200/IS605 family members. The IS ends with their essential secondary structures are shown in red (left end) and blue (right end). The colour scheme is as described for (a). The TnpB accessory gene is shown as a green horizontal arrow.

The mechanisms involved in acquisition of additional genes to generate tISs are at present unclear. They do not appear to carry programmed recombination systems as do some members of the Tn3 family (see 'ISs with accessory genes: Tn3 family' above). One possibility is that tISs are derived by deletion from ancestral compound transposons. These are composed of two ISs flanking any DNA segment either in direct or inverted orientation (Craig *et al.*, 2002). The flanking ISs are able to mobilise the intervening DNA segment. They were among the first types of transposon described and include the models Tn5 (flanking IS50) and Tn10 (flanking IS10) as well as Tn9 (flanking IS1). As these early examples, many other such composite transposons have been identified either by experiment or from genome sequencing (Tn number registry: http://www.ucl.ac.uk/eastman/research/departments/microbial-diseases/tn).

Some early observations concerning Tn5 and Tn10 suggest that the flanking ISs can undergo mutation rendering them less autonomous [for example mutations within one IS which inactivate its transposase; see (Mahillon & Chandler, 1998)]. Furthermore, studies on IS101 from the pSC101 plasmid clearly indicated that transposition can occur using one established IS end and a second surrogate end located at some distance from the IS (Machida *et al.*, 1982). If, as in this case, the intervening DNA includes a passenger gene, this creates a novel transposon. It has also been observed that other ISs such as IS911 can use surrogate ends during transposition (Polard *et al.*, 1994). Moreover, isolated individual IS ends are often observed in sequenced genomes and could provide a source of surrogate ends.

The Tn3 transposon family derivatives

Tn3 family members are quite variable: several examples lack passenger genes and therefore do not fall into the strict definition of a transposon, while others lack both passenger and resolvase genes (e.g. ISVsa19, ISShfr9, ISBusp1, IS1071...) and therefore closely resemble ISs. Many family members carry a number of passenger genes. These can represent entire operons, notably mercury resistance, or individual genes involved in antibiotic resistance, breakdown of halogenated aromatics or virulence [e.g. (Liebert *et al.*, 1999)]. They often carry integron recombination platforms enabling them to incorporate additional resistance genes by recruiting integron cassettes (Mazel, 2006).

IS-related ICEs

Other structures which obscure the definition of an IS have been identified among various TEs. For example, ICEs (integrative conjugative elements) are found integrated into the host genome but can excise and transfer from cell to cell. Their insertion and excision are generally catalysed by enzymes related to site-specific recombinases, whereas their transfer depends on a second set of proteins, which includes a 'relaxase', often a single-strand endonuclease of the HUH superfamily (Burrus *et al.*, 2002) (Chandler *et al.*, 2013). However, ISSag10, a tIS

member of the IS*1595* family from *Streptococcus agalactiae* which includes an O-lincosamide nucleotidyltransferase gene, encodes a DDE transposase and undergoes cell-to-cell transfer when complemented with an autonomous ICE, Tn*916* (Achard & Leclercq, 2007). In this case, a cryptic origin of transfer is located within the 3′ end of the resistance gene. These nonautonomous ICEs have been called IMEs [integrative mobilisable elements; (Adams *et al.*, 2002)] or cis-mobilisable elements [CIMEs; (Pavlovic *et al.*, 2004)].

More recent studies have identified a new ICE family, transposon of Group B Streptococcus (TnGBS), in which the enzyme catalysing their integration and excision belongs to another DDE-group Tpase (Brochet *et al.*, 2009; Guerillot *et al.*, 2013). Indeed, this has led to the identification of an entirely new family of classic ISs carrying DDE Tpases, the IS*Lre2* family (Guerillot *et al.*, 2014). In addition to the TnGBS family, other ICEs have been identified which include a DDE Tpase closely related to that of the IS*30* family (Smyth & Robinson, 2009). It seems likely that examples of ICEs with other IS family Tpases are awaiting identification. Moreover, in addition to a variety of transfer functions, certain ICEs carry plasmid-related replication genes important in ensuring sufficient stability of the transposition intermediates to enable their subsequent integration (Lee *et al.*, 2010; Guerillot *et al.*, 2013). Indeed, early examples of ICEs (Murphy & Pembroke, 1995) were initially thought to be resistance plasmids and assigned an incompatibility group, incJ. These are maintained as an integrated copy in the host chromosome but can nevertheless give rise to circular copies (Pembroke & Murphy, 2000). This is yet another example of the increasingly indistinct frontiers between phage, plasmids and transposons (Bi *et al.*, 2012) (http://db-mml.sjtu.edu.cn/ICEberg/).

IS*91*-related ISCRs

These MGE include a putative transposase of the HUH family similar to that of the IS*91* family, the so-called 'common region' (CR). Although it has yet to be demonstrated that such structures indeed transpose, ISCRs are associated with multiple flanking antibiotic resistance genes. It is thought that these genes are transmitted during the rolling circle type of transposition mechanism postulated to occur in IS*91* transposition. This involves an initiation event at one IS end, polarised transfer of the IS strand into a target molecule and termination at the second end (Garcillan-Barcia *et al.*, 2002). Flanking gene acquisition is thought to occur when the termination mechanism fails and rolling circle transposition extends into neighbouring DNA where it may encounter a second surrogate end (Garcillan-Barcia *et al.*, 2002). This type of

mobile element may prove to play an important role in the assembly and transmission of multiple antibiotic resistance (Toleman *et al.*, 2006) (Toleman & Walsh, 2011).

Nonautonomous derivatives: Miniature Inverted repeat TEs (MITEs), Mobile Insertion Cassette (MICs), Palindrome-associated TEs (PATEs)

Structures called MITEs (Feschotte *et al.*, 2002) are related to ISs with DDE Tpases (Fig. 4a). They are composed of two appropriately oriented left and right ends similar to those of ISs. They are generally < 300 bp long and are probably derived from ISs by internal deletion. Some carry short noncoding sequences between these IRs which may or may not be IS-derived. Others carry various coding sequences and have been called MICs (De Palmenaer *et al.*, 2004). MITEs are considered to be nonautonomous TEs mobilisable *in trans* by Tpases of full-length parental genomic copies. They were first identified in plants (Feschotte *et al.*, 2002) and are related to Tc/mariner elements (distantly related to bacterial IS*630* family). IS*630*-related MITEs were also the first described bacterial examples (Correia *et al.*, 1988; Oggioni & Claverys, 1999; Buisine *et al.*, 2002). MITEs showing similarities to other IS families have now been observed in bacteria and archaea (Brugger *et al.*, 2002; Filee *et al.*, 2007). Among those derived from ISs with DDE transposases, representatives of the IS*1*, IS*4*, IS*5*, IS*6*, and even Tn*3* family members such as IS*Rf1* from *Sinorhizobium fredii* have been identified.

MITE-like structures related to elements with other types of transposase have also been identified. Among these are IS*200*/IS*605* family derivatives (Filee *et al.*, 2007); (P. Siguier, unpublished) now called PATEs (Dyall-Smith *et al.*, 2011) (Fig. 4b) reflecting the subterminal secondary structures which constitute the ends of these ISs. Although originally observed in the Archaea as derivatives of known IS*200*/IS*605* family members, they have also been observed in certain cyanobacteria and *Salmonella* (P. Siguier, unpublished). They presumably represent decay products that appear quite frequently for this IS family.

Full-length copies of the parental IS may not be available or may be so divergent as to escape detection in standard BLAST analysis. This is the case for certain MITEs from the archaea (Brugger *et al.*, 2002) and is probably also true for the bacteria. Their detection and analysis are therefore arduous.

A good example of an IS group that includes examples of many of these IS derivatives (canonical ISs, MITEs, MICs and tISs) is the IS*231* subgroup of the large IS*4* family [Fig. 4; (De Palmenaer *et al.*, 2004)].

Characteristics of family life

Having presented an overview of different characteristics that contribute to assignment of an IS to a specific family, we detail below some of the more specific features which are used to group ISs into families and which impact on their behaviour.

Transposases and mechanism

Transposition requires a set of DNA cleavages at the ends of a TE and a set of strand transfer reactions which move these ends into a suitable target DNA molecule catalysed by the TE-encoded Tpase. Transposases are often multidomain proteins. In addition to catalysis, they must also recognise specific DNA sequences at the IS ends and engage with the DNA target to form multimeric nucleo-protein assemblies, transpososomes. These nucleo-protein assemblies provide a precise architecture within which the chemical steps of transposition are carried out. They are composed of two or more Tpase monomers and, in some cases, accessory proteins such as DNA-architectural proteins (for example IHF, HU and HNS; reviewed in [Mizuuchi, 1992a, b; Haniford, 2006]) or possibly protein chaperones [e.g. GroEL in the case of IS1; (Ton-Hoang *et al.*, 2004)]. They are dynamic and undergo conformational changes to coordinate DNA cleavages and strand transfers and ensure that, once started, transposition goes to completion (Dyda *et al.*, 2012; Montano *et al.*, 2012).

DDE transposases

Many of the presently identified Tpases are members of the DDE family (Fig. 1; Table 1). These are structurally and catalytically related to RNaseH and other nucleic acid processing enzymes (Rice *et al.*, 1996) and are said to have an RNaseH fold. The highly conserved DDE triad serves to coordinate one or two divalent cations such as Mg^{2+} which in turn assist polarisation of the phosphate group belonging to the target phosphodiester bond, facilitating cleavage. DDE enzymes use hydroxyl groups as nucleophiles for cleavage and for strand transfer: H_2O for initial cleavage and a 3′OH, generated by cleavage of the TE ends, for the strand transfer reaction [reviewed in (Mizuuchi, 1992a, b; Hickman *et al.*, 2010a, b)].

The DDE triad is often followed by a basic amino acid residue, generally K or R, 7 amino acids downstream and on the same face of the α-helix that also carries the final conserved E residue. Differences in the DDE spacing as well as the presence or absence of specific submotifs (for instance IS4 family transposases include the conserved N2 N3 C1 signatures which carry the D, D and E residues

and a motif YREK [Y-(2)-R-(3)-E-(6)-(K)]; (Rezsohazy *et al.*, 1993; De Palmenaer *et al.*, 2008) are used in distinguishing different groups and families (Table 1). The earliest DDE motifs from transposases, those of the IS3 family, were identified by their similarities with retroviral integrases (IN) (Fayet *et al.*, 1990) and, like IN, have a D35E spacing. Other DDE-group enzymes exhibit a spacing of 33–35 residues between the conserved D and E. However, the DDE transposases of several families carry relatively long distinctive insertion domains of either α-helical or β-strand (Hickman *et al.*, 2010a, b).

Certain IS groups appear to carry variants of the conserved DDE motif with N or H residues replacing the E. The role of these alternative residues in catalysis has yet to be tested. Some groups, for example the IS1595 family, include several of these variant motifs (Table 1) (Siguier *et al.*, 2009) and the fact that these IS exist in several copies suggests that they are active.

DDE enzymes catalyse cleavage of only one DNA strand generally generating a 3′OH at the IS end. This is known as the transferred strand because the 3′OH is used as a nucleophile in the integration step to attack the target phophodiester bond and complete strand transfer. However, transposition of these ISs occurs via double-strand DNA intermediates and therefore requires processing of the second strand (called the nontransferred strand) to liberate the IS from flanking DNA sequences in the donor molecule. This can occur in several different ways and is also IS family-specific (Turlan & Chandler, 2000; Curcio & Derbyshire, 2003) (Table 1) and serves to reinforce groupings derived from bioinformatics comparison by providing a mechanistic coherence. Thus, for ISs of the IS4 family (IS50, IS10), the initial 3′OH is used to attack the opposite strand forming a transient hairpin bridge at the IS end. This is then cleaved using H_2O as a nucleophile to liberate the IS and regenerate the 3′OH on the transferred strand ready for strand transfer and integration. This is known as a cut-and-paste mechanism. Other ISs such as the IS630 family, related to the eukaryote mariner/Tc superfamily, employ a mechanism in which the initial H_2O-catalysed cleavage of the nontransferred strand occurs with a small offset of 2 bases into the IS prior to cleavage of the transferred strand (Feng & Colloms, 2007) as do their eukaryotic cousins (Richardson *et al.*, 2009).

An additional mechanism adopted by certain elements with DDE Tpases is cointegrate formation. Here, the transposon inserts into a target replicon in a process accompanied by TE replication. This results in fusion of the donor and target replicons with a directly repeated TE copy at each junction known as a Shapiro intermediate (Shapiro, 1979). Tn3 and IS6 family members

transpose using this pathway as does bacteriophage Mu and its relatives (Chaconas & Harshey, 2002).

However, by far the most common mechanism is the so-called copy–paste mechanism (Curcio & Derbyshire, 2003), which generates a transient double-strand circular DNA intermediate. This has been adopted by a significant number of IS families including IS3, IS30, IS110, IS256, ISLre2 and possibly others. IS1 uses this as one of several transposition pathways (Turlan & Chandler, 1995). For IS3 family members, circle formation occurs in an asymmetric manner. One IS end is cleaved to generate the characteristic 3′OH of the transferred strand. This then serves to attack several nucleotides exterior to the second end to generate a single-strand bridge leaving a free 3′OH on the IS flank. The 3′OH can act as a replication primer. IS replication would regenerate an intact copy reconstituting the donor plasmid and produce a double-strand circular DNA intermediate. Due to low basal Tpase levels, this initial step may occur in a stochastic manner. However, formation of the circular intermediate results in the assembly of a transient strong promoter composed of a -35 promoter element in the right IS end oriented outwards and a -10 promoter element in the left end oriented inwards (Ton-Hoang et al., 1997). This promoter serves to drive transposase synthesis and consequent integration and disassembly of the promoter. Thus, the circular intermediate once generated is committed to terminate transposition.

DEDD transposases

Another motif, DEDD (Buchner et al., 2005) is characteristic of the four-way Holliday junction (HJ) resolvase, RuvC. RuvC also has an RNaseH fold (Ariyoshi et al., 1994). A DEDD motif has also been identified in transposases of the IS110 family (Tobiason et al., 2001), closely related to the Piv and MooV invertases from Moraxella lacunata/M. bovis (Fulks et al., 1990; Rozsa et al., 1997) and N. gonorrhoeae (Choi et al., 2003; Skaar et al., 2005), respectively. Piv catalyses inversion of a DNA segment permitting expression of a type IV pilin. However, the organisation of IS110 family members and the inversion systems are different. In the IS, the recombinase is located within the element, whereas in the inversion systems, it is located outside the invertible segment (Buchner et al., 2005).

Although it has proved difficult to determine the activity of these transposases in detail in vitro, transposition of ISs with DEDD Tpases may be unusual and involve HJ intermediates that must be resolved using a RuvC-like mechanism. This type of recombination would be consistent with the close relationship between DEDD transposases and the Piv/MooV invertases that presumably resolve HJ structures during inversion (Tobiason et al., 1999).

HUH transposases

The second major group of transposases are the HUH enzymes [for a review see (Chandler et al., 2013)]. HUH refers to a pair of His residues (H) separated by a bulky hydrophobic residue (U) (Ilyina & Koonin, 1992). Together with a third residue, the two His coordinate a divalent metal ion cofactor required for catalysis. These enzymes also include one or two tyrosine (Y) residues, which act as nucleophiles and form transient 5′ phosphotyrosine enzyme–DNA transposition intermediates. Like DDE enzymes, HUH enzymes are widespread and assume other roles in the cell. In addition to transposition, they are involved in rolling circle plasmid and phage replication, in rolling hairpin replication of eukaryotic viruses (Rep proteins) and conjugative plasmid transfer (relaxases or Mob proteins).

IS200/IS605-family transposases

These Y1 Tpases are among the smallest identified to date with c. 150 amino acids. These promote single-strand transposition (Ronning et al., 2005; Ton-Hoang et al., 2005) in contrast to DDE enzymes, which catalyse double-strand transposition. They form obligatory dimers in which the catalytic site is composed of an HUH motif from one monomer and a single Y residue together with the third residue necessary for coordinating the essential divalent metal ion contributed by the second monomer (Ronning et al., 2005). Although the founding family member, IS200, was identified in Salmonella 30 years ago (Lam & Roth, 1983), their activities have only recently been unravelled (Ronning et al., 2005; Ton-Hoang et al., 2005, 2010; Barabas et al., 2008; Guynet et al., 2008, 2009). Two model ISs, IS608 from Helicobacter pylori and ISDra2 (Hickman et al., 2010a, b; Pasternak et al., 2010; Ton-Hoang et al., 2010) from D. radiodurans [IS8301 in (Islam et al., 2003)] have been studied in detail. Both use obligatory circular ssDNA intermediates for their mobility. They excise as ssDNA circles with abutted left and right ends and insert 3′ to a conserved element-specific penta- or tetra-nucleotide target (Guynet et al., 2009). As the transposon is 'peeled' from its donor site as a single strand, this mechanism might be called 'peel-and-paste' (B. Ton-Hoang pers. commun.). The target sequence is essential for further transposition (Ton-Hoang et al., 2005). The transposase, TnpA, recognises small, subterminal hairpin structures in a strand-specific manner and catalyses single-strand cleavage at both ends on the 'top' strand. Surprisingly, the left and right cleavage sites are not directly recognised by TnpA. Instead, they form a network of particular base interactions with short 'guide' sequences located 5′ to the TnpA-bound subterminal

secondary structures at each end. This network includes both canonical (Watson-Crick) and noncanonical base interactions (Barabas *et al.*, 2008; He *et al.*, 2011) (Hickman *et al.*, 2010a, b). These are also essential to stabilise the nucleoprotein complex, the transpososome, within which the DNA strand cleavages and transfers occur. Indeed, changes in the guide sequences result in predictable changes in insertion site specificity (Guynet *et al.*, 2009).

Family members excise from, and insert preferentially into, the lagging strand template of chromosome and plasmid replication forks (Ton-Hoang *et al.*, 2010). Their lagging strand preference generates an insertion bias reflecting the mode of replication (uni- or bi-directional) of the target replicon. Moreover, transposition of IS*Dra2* is strongly induced upon recovery of the highly radiation-resistant *D. radiodurans* host from irradiation (Pasternak *et al.*, 2010) as a result of the large amounts of ssDNA generated during the reassembly of the shattered *D. radiodurans* genome (Zahradka *et al.*, 2006).

IS*91*/ISCR-family transposases

The second major HUH Tpase group, those of the IS*91* family, has been known for some time (Garcillan-Barcia & de la Cruz, 2002; Garcillan-Barcia *et al.*, 2002). They were recognised as relatives of the rolling circle plasmid replication Rep proteins. They are larger than the simple Y1 transposases of the IS*200*/IS*605* family, carry a pair of tyrosine residues that are both necessary for transposition (Garcillan-Barcia *et al.*, 2002) together with an N-terminal ZF motif and are known as Y2 transposases. Tpases of the related ISCR elements are similar to IS*91* transposases but include only a single Y residue. Although there is no information concerning the transposition of ISCRs, IS*91* is thought to transpose by a rolling circle replication mechanism (RCR) initiating at one end (ori, 3′ to the transposase) and terminating at the other (ter, 5′ to the transposase) (Garcillan-Barcia *et al.*, 2002). Relatively, frequent failure of correct termination (1%) results in transposition of additional genes that flank the transposon in the donor molecule. Indeed, *ori* is essential for activity while removal of *ter* reduces but does not eliminate transposition.

Insertion of IS*91*, like that of members of the IS*200*/IS*605* family, is oriented with ori adjacent to the 3′ of a specific tetranucleotide target (5′-CTTG or 5′-GTTC) and, like that of the IS*200*/IS*605* family, this target sequence is essential for further transposition (del Pilar Garcillan-Barcia *et al.*, 2001; Garcillan-Barcia *et al.*, 2002). In the proposed RCR mechanism, displacement of an IS*91*/ISCR active transposon strand would be driven by leading strand replication of the donor replicon from a 3′OH generated by cleavage at ori. Although the original model proposed that the cleaved IS end is transferred to the target DNA and the IS is replicated 'into' the target replicon, single-strand and double-strand IS*91* circles have been observed and it is difficult to explain these as intermediates in the RCR model.

Serine transposases

IS*607* family members encode a Tpase closely resembling serine site-specific recombinases that use serine as a nucleophile for cleavage of the DNA strand (Grindley, 2002). At present, little is known about transposition of this IS family although it is thought that these elements generate circular intermediates [N.D.F. Grindley pers. commun. cited in (Filee *et al.*, 2007)]. Presumably, the enzyme catalyses similar cleavages and strand transfers as its site-specific serine recombinase cousins using a transitory 5′ phosphoserine covalent intermediate. Based on transposase structures from structural genomics studies and detailed knowledge of the general serine recombinase mechanism, (Boocock & Rice, 2013) have proposed a model for the transposition mechanism. This includes a synaptic transposase tetramer (as for classical serine recombinases). The model explains the lack of target specificity exhibited in IS*607* transposition (Kersulyte *et al.*, 2000), behaviour which is unusual for this type of recombinase.

Transposase organisation and its role in regulation: domain structure

Transposases are composed of a combination of domains with recognisable secondary structures. In many cases, the presence and order of these domains defines subgroups within an IS family.

Although there are only a limited number of Tpase structures available, these secondary structures can be predicted from the primary amino acid sequence and include ZF (Zn), helix-turn-helix (HTH), leucine zipper (LZ), RNaseH fold or HUH domains (see references in (Montano & Rice, 2011; Dyda *et al.*, 2012).

DDE transposases

For DDE transposases, the domains of the protein which ensure sequence-specific DNA binding (i.e. to the IS ends) are typically located towards the N-terminal (N-terminal) end and the catalytic domain more towards the C-terminal (C-terminal) end. From a functional point of view, this arrangement facilitates a phenomenon called cotranslational DNA binding in which the protein folds in the course of translation allowing the nascent polypeptide chain to

initiate binding directly to the closest IS end. Indeed, for members of a number of IS families, including IS*1* (Zerbib *et al.*, 1987), IS*3* (Haren *et al.*, 1998), IS*30* (Stalder *et al.*, 1990), the isolated DNA-binding domain bind more avidly than the entire protein, suggesting that the C-terminal end may actually inhibit specific binding by the N-terminal domain, possibly by steric masking. Once bound, the N-terminal domain would be insensitive to C-terminal domain masking. Thus, specific Tpase binding would only occur transitorily early during translation providing an attractive explanation for the preferential cis activity (the preference to act on the TE from which the Tpase is expressed) of many Tpases (Duval-Valentin & Chandler, 2011). Both Helix-turn-Helix (HTH) and ZF domains have been implicated as DNA-binding domains (e.g. IS*911* (Rousseau *et al.*, 2004); IS*1* (Ohta *et al.*, 2004; Ton-Hoang *et al.*, 2004). These can be found individually or together as illustrated in Fig. 6.

The ability to multimerise is another important Tpase property. This is because the TE ends are typically both involved simultaneously in the transposition process within the transpososome. Multimerisation can serve as an important regulatory mechanism because a Tpase monomer bound to one TE end often carries out its catalytic activities on the other end. This assures that catalysis only occurs once the transpososome complex has been correctly assembled and is primed to complete the transposition process.

Although multimerisation can be complex and can involve several different Tpase regions (Braam *et al.*, 1999), in the case of IS*3* family members, this is accomplished by a coiled-coil LZ structure (Haren *et al.*, 2000).

Several examples of variations in domain organisation found in related DDE transposases are shown in Fig. 6. For IS*1*, the transposase is generally expressed as a fusion protein by programmed translational frameshifting (Fig. 6a; see '*Recoding and domain organisation*' below). However, derivatives are found in which the transposase is produced from one continuous frame. In these cases, derivatives which have been truncated from the N-terminal ZF or with additional N-terminal extensions have been identified (Fig. 6b). Similar truncated derivatives are also found in the case of the IS*1595* family distantly related to IS*1* (Fig. 6c) and the large closely related IS*3* and IS*481* families (Fig. 6d and e).

DEDD Tpases

It is interesting to note that in the related DEDD Tpases that are related to the Piv and MooV invertases and limited at present to IS*110* family members, the potential DNA-binding domain is located downstream of the
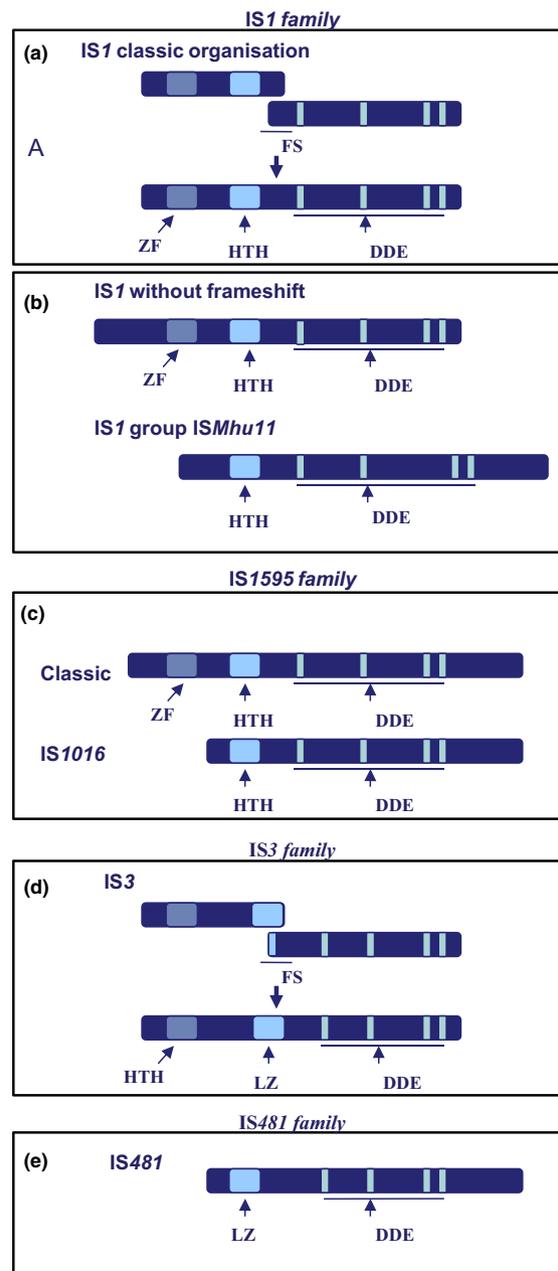
Fig. 6. Domain organisation of transposases of the DDE family. The relative positions of the potential ZF, HTH, LZ and the 'DDEK/R' catalytic motif are indicated from left to right as light blue boxes. The figure illustrates the N-terminal and C-terminal extension of the different transposase examples. (a) Classical IS*1* with frameshift. The position of the frameshift window which is used to generate InsAB is indicated. (b) IS*1* without frameshift and the IS*Mhu11* group showing the deletion of the ZF, the C-terminal extension and the increased spacing between the second (d) and (e) residues. (c) The IS*1595* family showing the classical IS*1595* group and the IS*1016* group which does not carry the N-terminal ZF. (d) The IS*3* family including members with and without the translational frameshift. (e) The closely related IS*481* family which lack the N-terminal HTH domain and exhibit an additional C-terminal domain.

catalytic domain (Choi *et al.*, 2003). This again reinforces the idea that the DEDD ISs possess a different transposition mechanism to those with DDE Tpases.

HUH transposases

The smaller Y1 HUH Tpases of the IS*200*/IS*605* family are very compact with DNA binding, cleavage site and target recognition functions closely interwoven. The active Tpase is dimeric and, in the published protein structures, the active site is composed of the HUH motif of one monomer and the active site tyrosine of the other. The tyrosine residue is located downstream of the HUH on an α-helix joined to the body of the protein by a flexible arm. The α-helix also carries a third residue which, with the His pair, completes the amino acid triad essential for divalent metal ion binding (Ronning *et al.*, 2005; Barabas *et al.*, 2008), (Hickman *et al.*, 2010a, b). Each Y residue in this 'trans' position cleaves one IS end generating a 5′ phosphotyrosine bond. Strand transfer is then thought to occur by rotation of the flexible arms so that the Y residue covalently linked to DNA engages with the HUH of the same monomer, a 'cis' configuration. The cleavage reaction is then reversed resulting in strand transfer and the enzyme then appears to be reset in the trans configuration (He *et al.*, 2013).

The overall organisation of the much larger Y2 enzymes is less well determined although there is a ZF motif which possibly functions as a DNA recognition domain located in the N-terminal region (Chandler *et al.*, 2013).

Serine transposases

The family of serine recombinases is composed of three groups: the resolvase/integrase group; the large serine recombinases; and the serine transposases (Boocock & Rice, 2013). For the two former groups, the catalytic domain is invariably located at the N-terminus and is followed by a 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 Rutherford [Van Duyne & Rutherford, 2013]).

It is interesting to note that the IS*607* family serine transposases carry their DNA-binding domain N-terminal to the catalytic domain (Gilbert & Cordaux, 2013) (Boocock & Rice, 2013) in a similar way to other (DDE) transposases and may reflect a similar function.

Recoding and domain organisation

Certain TEs can express combinations of different functional protein domains, providing a way of encoding two proteins of different function in one DNA segment and resulting in highly compact genetic structures. These types of noncanonical readout of the genetic code are collectively known as 'Recoding' (Gesteland *et al.*, 1992). Recoding is significantly more frequent and well documented in MGE-encoded genes of both prokaryotes and eukaryotes than in any other gene family (Baranov *et al.*, 2005; Sharma *et al.*, 2011).

Programmed ribosomal frameshifting

The most commonly encountered recoding event is probably programmed ribosomal frameshifting (PRF) involving ribosome slippage of 1 nt. This often occurs in the 5′ direction to generate a -1 frameshift and requires specific signals in the mRNA such as an upstream RBS, a sequence facilitating slippage (slippery codons) and downstream secondary structures (hairpins and pseudoknots) (Sharma *et al.*, 2011).

PRF occurs in retroviruses, coronaviruses and some plant viruses. It is also very common in prokaryotic ISs. Initially, PRF signals were identified in the transposase genes of IS*1* and IS*3* family members (Sekine & Ohtsubo, 1989; Escoubas *et al.*, 1991; Polard *et al.*, 1991; Sekine *et al.*, 1994) where they serve to fuse the product of the upstream frame, generally a DNA-binding protein which, on its own, acts as a regulator, to the downstream catalytic domain (Zerbib *et al.*, 1990; Ton-Hoang *et al.*, 1998; Rousseau *et al.*, 2007).

Programmed transcriptional realignment

A less well-characterised phenomenon is programmed transcriptional realignment (PTR) (Sharma *et al.*, 2011). This is mechanistically distinct from PRF and is due to insertion or deletion of a few nucleotides by transcription 'slippage' leading to different types of frameshift (+/−: 1, 2. . .). Although PTR needs signals such as a run of A, T or A_nG_n, these are not as distinct as those involved in PRF and the phenomenon is less well studied principally because it is more difficult to address experimentally.

Two additional families (IS*5* sgr IS*427* and certain IS*630*) whose transposases are also distributed over two consecutive reading phases have been identified although at present it is unclear whether transposase expression involves PRF or PTR.

Pyrrolysine- and selenocysteine-mediated stop codon read-through

Less common than PRF and PTR are recoding events involving stop codon (UAG) read-through by insertion of the atypical amino acids pyrrolysine and selenocysteine (Söll, 2007). Insertion sites are characterised by a secondary

structure element located downstream of the stop codon respectively called PYLIS (Namy *et al.*, 2004) (Ibba & Soll, 2004) and SECIS (Liu *et al.*, 1998). Certain ISs probably use stop codon read-through to control transposase expression. These have been identified in organisms that encode these read-through mechanisms such as *Methanosarcina* (IS*Mac17*, IS*Mba9* and IS*Mma13* in the case of pyrrolysine) and *Desulfovibrio vulgaris* (IS*Dvu3*, in the case of selenocysteine). As UAG read-through has clearly evolved as a more general regulatory mechanism in these organisms, this implies that the IS in question have taken advantage of these host read-through regulatory pathways and have evolved with their hosts to limit their transposition activity.

Domestication

One important aspect of the impact of ISs which has not been addressed in detail is the capacity of the host to harness their properties to perform various cellular functions. Such domestication has been described repeatedly in eukaryote systems (Sinzelle *et al.*, 2009) but remains relatively rare in the prokaryotes. However, in addition to their activities in modulating gene expression (see 'Impact of ISs on Genome Expression' above), there are two interesting examples of prokaryotic IS domestication. Both concern HUH enzymes related to those of the IS*200*/IS*605* family (see "Major IS groups are defined by transposase type' above).

The REP system

In one case, a TnpA-like enzyme, TnpA$_{REP}$ (Ton-Hoang *et al.*, 2012) or RAYT (Rep Associated Tyrosine Transposase, (Nunvar *et al.*, 2010) is associated with generally high copy number sequences for REP originally identified in the enterobacteria (Higgins *et al.*, 1982). These resemble the ends of IS*200*/IS*605* family members in structure and can occur in clusters localised around the chromosome. Some are grouped into pairs called bacterial interspersed mosaic elements (BIMES). They are present in many bacteria (Tobes & Ramos, 2005; Nunvar *et al.*, 2010, 2013) and have several key roles in aspects of host physiology, including organisation of genome structure (Gilson *et al.*, 1990; Boccard & Prentki, 1993; Espeli & Boccard, 1997), regulation of gene expression (Espeli *et al.*, 2001; Khemici & Carpousis, 2004; Moulin *et al.*, 2005; Aguena *et al.*, 2009) and genome plasticity (Clement *et al.*, 1999; Wilde *et al.*, 2003; Tobes & Pareja, 2006). TnpA$_{REP}$ from *E. coli* cleaves and rejoins REP sequences (Ton-Hoang *et al.*, 2012) and is very similar in structure to TnpA of IS*200*/IS*605* family members (Messing *et al.*, 2012). Little is known at present about the mechanisms involved in their invasion of and spread throughout their host genomes.

Group I introns

The second example is the group I bacterial intron or IStron sometimes found in the Clostridia and Bacilli, which includes TnpA- and TnpB-like genes presumably as homing endonucleases (Braun *et al.*, 2000, Hasselmayer, *et al.*, 2004a, b). However, in a large number of IStrons, these genes are clearly undergoing decay and exist as truncated derivatives as are many of their IS cousins (P. Siguier, unpublished). The IStron TnpA is quite similar to that of IS*Cpe2*, itself from *Clostridium perfringens* and to that of IS*Dra2*. Moreover, all IStron copies are inserted 3′ to a pentanucleotide, TTGAT, the target sequence for IS*Dra2*. TTGAT is also complementary to the intron internal guide sequence presumably required at the RNA level in the splicing reaction. Little is known about IStron behaviour.

Conclusion

Diversity and the boundaries between different TE types

We have tried here to provide an overview of the contribution of prokaryotic IS to the life of their host. We have underlined their diversity in sequence, organisation and behaviour and have attempted to describe their importance in generating genomic modifications and their contribution to modifying gene expression. We have also highlighted the growing realisation that their classification as genetic objects such as Tn, ISs and ICEs is in some ways artificial because there appears to be a large degree of diversity in these MGEs such that the border between certain types is becoming increasingly ambiguous. This is clearly illustrated by the large spectrum of interrelated mobile elements which are vehicles for antibiotic resistance. In addition to the numerous 'classical' composite transposons with flanking ISs and unit transposons such as those of the large Tn*3* family, we have described autonomous (tISs) and nonautonomous (MICs) IS derivatives, integrative conjugative elements (ICEs) with IS-like Tpases (Brochet *et al.*, 2009) and ISCR with IS*91*-like HUH Tpases and multiple resistance genes both upstream and downstream of the Tpase (Toleman *et al.*, 2006). It is tempting to suggest that ISs may be major building blocks used in assembling other more complex TEs. This clearly is the case in the domesticated IStron and widely spread TnpA$_{REP}$ systems.

Relationship with eukaryotic TE

Although often treated separately, it is worthwhile underlining that many prokaryotic IS families (at least those with DDE transposases) have relatives in TEs identified in eukaryotes (e.g. IS*1595*/*merlin*; IS*256*/*mutator*; IS*630*/Tc1/

mariner; IS*1380*/*piggyBac*). They have similar sizes and, where known, similar transposition mechanisms (Hickman *et al.*, 2010a, b). Although certain elements have received detailed attention (Hua-Van & Capy, 2008), the phylogenetic relation between most has yet to be analysed.

Future directions

One view of the prokaryote genome is as an ecological niche in which TEs and, in particular ISs, form part of a genomic landscape continuously modified by their own activity. This involves both endogenous elements and those accumulated by horizontal transfer. Our present understanding of the way in which ISs shape genomes is based largely on descriptive studies which often employ inaccurately annotated genomes. To obtain a more dynamic picture of the forces involved and the processes shaping a genome, it is necessary to understand the detailed mechanism(s) of IS movement; their target sites, that is, whether they prefer replication forks, actively transcribed regions, plasmids, chromosomes, specific DNA sequences or secondary structures and in what combinations; the interactions (interference) between elements; involvement of host factors etc. There have been few attempts to understand these factors and even fewer attempts to integrate them.

There is a need to understand TE behaviour not only on a genomic scale but within populations and communities to explain their dynamics. The role of genome isolation in IS expansion is an excellent example. To understand TE behaviour, there is a need to build quantitative multiscale models (Dada & Mendes, 2011) rather than descriptive models. This is a challenging task because this type of systems biology approach will require incorporation of a large number of very diverse parameters. At one level, this involves considerations such as how the DNA spreads through communities, the regulatory circuitry including the signals involved in DNA transfer between cells, the nature and function of the acquired mobile DNA, the way in which cells deal with the 'intrusion' of newly introduced genes and how they accommodate the associated changes in fitness. On a second level, it requires knowledge of the fine regulatory circuits determining TE movement itself, both biochemical and temporal, how this depends on various aspects of host physiology (interaction with DNA structural proteins, with replication repair and the cell cycle, with different chromosomal domains) and to what extent there may be interaction/interference between different TEs in the same cell.

As a final word, there are clearly more potential TE types awaiting identification and characterisation [e.g.

(Ricker *et al.*, 2013)] and these will also need to be factored into the final picture.

Acknowledgements

References

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

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

Aguena M, Ferreira GM & Spira B (2009) Stability of the pstS transcript of *Escherichia coli*. *Arch Microbiol* **191**: 105–112.

Allewalt JP, Bateson MM, Revsbech NP, Slack K & Ward DM (2006) Effect of temperature and light on growth of and photosynthesis by *Synechococcus* isolates typical of those predominating in the octopus spring microbial mat community of Yellowstone National Park. *Appl Environ Microbiol* **72**: 544–550.

Altschul SF, Gish W, Miller W, Myers EW & Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* **215**: 403–410.

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

Andersson SG & Kurland CG (1998) Reductive evolution of resident genomes. *Trends Microbiol* **6**: 263–268.

Ariyoshi M, Vassylyev DG, Iwasaki H, Nakamura H, Shinagawa H & Morikawa K (1994) Atomic structure of the RuvC resolvase: a Holliday junction-specific endonuclease from *E. coli*. *Cell* **78**: 1063–1072.

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.

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

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

Bao W, Jurka MG, Kapitonov VV & Jurka J (2009) New superfamilies of eukaryotic DNA transposons and their internal divisions. *Mol Biol Evol* **26**: 983–993.

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.

Baranov PV, Hammer AW, Zhou J, Gesteland RF & Atkins JF (2005) Transcriptional slippage in bacteria: distribution in sequenced genomes and utilization in IS element gene expression. *Genome Biol* **6**: R25.

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.

Bi D, Xu Z, Harrison EM *et al.* (2012) ICEberg: a web-based resource for integrative and conjugative elements found in Bacteria. *Nucleic Acids Res* **40**: D621–D626.

Blanc G, Ogata H, Robert C, Audic S, Claverie J-M & Raoult D (2007) Lateral gene transfer between obligate intracellular bacteria: evidence from the *Rickettsia massiliae* genome. *Genome Res* **17**: 1657–1664.

Boccard F & Prentki P (1993) Specific interaction of IHF with RIBs, a class of bacterial repetitive DNA elements located at the 3′ end of transcription units. *EMBO J* **12**: 5019–5027.

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

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

Braam LM, Goryshin IY & Reznikoff WS (1999) A mechanism for Tn5 inhibition. Carboxyl-terminal dimerization. *J Biol Chem* **274**: 86–92.

Braun V, Mehlig M, Moos M, Rupnik M, Kalt B, Mahony DE & von Eichel-Streiber C (2000) A chimeric ribozyme in *Clostridium difficile* combines features of group I introns and insertion elements. *Mol Microbiol* **36**: 1447–1459.

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.

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

Brugger K, Torarinsson E, Redder P, Chen L & Garrett RA (2004) Shuffling of *Sulfolobus* genomes by autonomous and non-autonomous mobile elements. *Biochem Soc Trans* **32**: 179–183.

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.

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.

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

Burrus V, Pavlovic G, Decaris B & Guedon G (2002) Conjugative transposons: the tip of the iceberg. *Mol Microbiol* **46**: 601–610.

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.

Chaconas G & Harshey RM (2002) Transposition of phage Mu DNA. *Mobile DNA*, Vol. II (Craig NL, Craigie R, Gellert M & Lambowitz A, eds), pp. 384–402. ASM press, Washington, DC.

Chandler M & Mahillon J (2002) Insertion sequences revisited. *Mobile DNA*, Vol. II (Craig NL, Craigie R, Gellert M & Lambowitz A, eds), pp. 305–366. ASM press, Washington, DC.

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.

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.

Clamp M, Cuff J, Searle SM & Barton GJ (2004) The Jalview Java alignment editor. *Bioinformatics* **20**: 426–427.

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.

Cordaux R, Udit S, Batzer MA & Feschotte C (2006) Birth of a chimeric primate gene by capture of the transposase gene from a mobile element. *P Natl Acad Sci USA* **103**: 8101–8106.

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.

Craig NL (2002) Tn7. *Mobile DNA*, Vol. II (Craig NL, Craigie R, Gellert M & Lambowitz A, eds), pp. 423–456. ASM press, Washington, DC.

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

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

Dada JO & Mendes P (2011) Multi-scale modelling and simulation in systems biology. *Integr Biol (Camb)* **3**: 86–96.

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

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

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.

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.

Dobrindt U, Hochhut B, Hentschel U & Hacker J (2004) Genomic islands in pathogenic and environmental microorganisms. *Nat Rev Microbiol* **2**: 414–424.

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

Dyall-Smith ML, Pfeiffer F, Klee K *et al.* (2011) *Haloquadratum walsbyi*: limited diversity in a global pond. *PLoS One* **6**: e20968.

Dyda F, Chandler M & Hickman AB (2012) The emerging diversity of transpososome architectures. *Q Rev Biophys* **45**: 493–521.

Enright AJ, Van Dongen S & Ouzounis CA (2002) An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res* **30**: 1575–1584.

Escoubas JM, Prere MF, Fayet O, Salvignol I, Galas D, Zerbib D & Chandler M (1991) Translational control of transposition activity of the bacterial insertion sequence IS1. *EMBO J* **10**: 705–712.

Espeli O & Boccard F (1997) *In vivo* cleavage of *Escherichia coli* BIME-2 repeats by DNA gyrase: genetic characterization of the target and identification of the cut site. *Mol Microbiol* **26**: 767–777.

Espeli O, Moulin L & Boccard F (2001) Transcription attenuation associated with bacterial repetitive extragenic BIME elements. *J Mol Biol* **314**: 375–386.

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.

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.

Feschotte C, Zhang X & Wessler S (2002) Miniature inverted repeat transposable elements and their relationship to established DNA transposons. *Mobile DNA*, Vol. II (Craig NL, Craigie R, Gellert M & Lambowitz A, eds), pp. 1147–1158. ASM press, Washington, DC.

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

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.

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.

Garcillan-Barcia MP & de la Cruz F (2002) Distribution of IS91 family insertion sequences in bacterial genomes: evolutionary implications. *FEMS Microbiol Ecol* **42**: 303–313.

Garcillan-Barcia MP, Bernales I, Mendiola MV & De la Cruz F (2002) IS91 rolling circle transposition. *Mobile DNA*, Vol. II (Craig NL, Craigie R, Gellert M & Lambowitz A, eds), pp. 891–904. ASM press, Washington, DC.

Gesteland RF, Weiss RB & Atkins JF (1992) Recoding: reprogrammed genetic decoding. *Science* **257**: 1640–1641.

Gil R, Belda E, Gosalbes MJ *et al.* (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.

Gilbert C & Cordaux R (2013) Horizontal transfer and evolution of prokaryote transposable elements in eukaryotes. *Genome Biol Evol* **5**: 822–832.

Gilson E, Perrin D & Hofnung M (1990) DNA polymerase I and a protein complex bind specifically to *E. coli* palindromic unit highly repetitive DNA: implications for bacterial chromosome organization. *Nucleic Acids Res* **18**: 3941–3952.

Gladyshev EA & Arkhipova IR (2009) A single-copy IS5-like transposon in the genome of a bdelloid rotifer. *Mol Biol Evol* **26**: 1921–1929.

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.

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

Grindley ND, Whiteson KL & Rice PA (2006) Mechanisms of site-specific recombination. *Annu Rev Biochem* **75**: 567–605.

Gueguen E, Rousseau P, Duval-Valentin G & Chandler M (2006) Truncated forms of IS911 transposase downregulate transposition. *Mol Microbiol* **62**: 1102–1116.

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.

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.

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.

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.

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.

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.

Haniford DB (2006) Transpososome dynamics and regulation in Tn10 transposition. *Crit Rev Biochem Mol Biol* **41**: 407–424.

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.

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

Haren L, Normand C, Polard P, Alazard R & Chandler M (2000) IS911 transposition is regulated by protein-protein interactions via a leucine zipper motif. *J Mol Biol* **296**: 757–768.

Hasselmayer O, Nitsche C, Braun V & von Eichel-Streiber C (2004a) The IStron CdISt1 of *Clostridium difficile*: molecular symbiosis of a group I intron and an insertion element. *Anaerobe* **10**: 85–92.

Hasselmayer O, Braun V, Nitsche C, Moos M, Rupnik M & von Eichel-Streiber C (2004b) *Clostridium difficile* IStron CdISt1: discovery of a variant encoding two complete transposase-like proteins. *J Bacteriol* **186**: 2508–2510.

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.

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.

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

Hickman AB, James JA, Barabas O *et al.* (2010b) DNA recognition and the precleavage state during single-stranded DNA transposition in *D. radiodurans*. *EMBO J* **29**: 3840–3852.

Higgins CF, Ames GF, Barnes WM, Clement JM & Hofnung M (1982) A novel intercistronic regulatory element of prokaryotic operons. *Nature* **298**: 760–762.

Hsia R, Ohayon H, Gounon P, Dautry-Varsat A & Bavoil PM (2000) Phage infection of the obligate intracellular bacterium, *Chlamydia psittaci* strain guinea pig inclusion conjunctivitis. *Microbes Infect* **2**: 761–772.

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

Hua-Van Al & Capy P (2008) Analysis of the DDE motif in the Mutator superfamily. *J Mol Evol* **67**: 670–681.

Ibba M & Soll D (2004) Aminoacyl-tRNAs: setting the limits of the genetic code. *Genes Dev* **18**: 731–738.

Ilyina TV & Koonin EV (1992) Conserved sequence motifs in the initiator proteins for rolling circle DNA replication encoded by diverse replicons from eubacteria, eucaryotes and archaebacteria. *Nucleic Acids Res* **20**: 3279–3285.

Islam SM, Hua Y, Ohba H, Satoh K, Kikuchi M, Yanagisawa T & Narumi I (2003) Characterization and distribution of IS8301 in the radioresistant bacterium *Deinococcus radiodurans*. *Genes Genet Syst* **78**: 319–327.

Jellen-Ritter AS & Kern WV (2001) Enhanced expression of the multidrug efflux pumps AcrAB and AcrEF associated with insertion element transposition in *Escherichia coli* mutants Selected with a fluoroquinolone. *Antimicrob Agents Chemother* **45**: 1467–1472.

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

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.

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.

Kersulyte D, Velapatino B, Dailide G *et al.* (2002) Transposable element ISHp608 of *Helicobacter pylori*: nonrandom geographic distribution, functional organization, and insertion specificity. *J Bacteriol* **184**: 992–1002.

Khemici V & Carpousis AJ (2004) The RNA degradosome and poly(A) polymerase of *Escherichia coli* are required *in vivo* for the degradation of small mRNA decay intermediates containing REP-stabilizers. *Mol Microbiol* **51**: 777–790.

Kieny M-P (2012) The evolving threat of antimicrobial resistance: Options for action. World Health Organization.

Kleiner M, Young JC, Shah M, VerBerkmoes NC & Dubilier N (2013) Metaproteomics reveals abundant transposase expression in mutualistic endosymbionts. *MBio* **4**: e00223–00213.

Kobayashi K, Tsukagoshi N & Aono R (2001) Suppression of hypersensitivity of *Escherichia coli* acrB mutant to organic solvents by integrational activation of the acrEF operon with the IS1 or IS2 element. *J Bacteriol* **183**: 2646–2653.

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

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

Lederberg EM (1981) Plasmid reference center registry of transposon (Tn) allocations through July 1981. *Gene* **16**: 59–61.

Lee CA, Babic A & Grossman AD (2010) Autonomous plasmid-like replication of a conjugative transposon. *Mol Microbiol* **75**: 268–279.

Liebert CA, Hall RM & Summers AO (1999) Transposon Tn21, flagship of the floating genome. *Microbiol Mol Biol Rev* **63**: 507–522.

Liu Z, Reches M, Groisman I & Engelberg-Kulka H (1998) The nature of the minimal 'selenocysteine insertion sequence' (SECIS) in *Escherichia coli*. *Nucleic Acids Res* **26**: 896–902.

Liu D, Bischerour J, Siddique A, Buisine N, Bigot Y & Chalmers R (2007) The human SETMAR protein preserves

most of the activities of the ancestral Hsmar1 transposase. *Mol Cell Biol* **27**: 1125–1132.

Machida Y, Machida C & Ohtsubo E (1982) A novel type of transposon generated by insertion element IS102 present in a pSC101 derivative. *Cell* **30**: 29–36.

Mahillon J & Chandler M (1998) Insertion sequences. *Microbiol Mol Biol Rev* **62**: 725–774.

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

McKenna M (2013) The last resort. *Nature* **499**: 394–396.

Medini D, Donati C, Tettelin H, Masignani V & Rappuoli R (2005) The microbial pan-genome. *Curr Opin Genet Dev* **15**: 589–594.

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.

Messing SA, Ton-Hoang B, Hickman AB et al. (2012) The processing of repetitive extragenic palindromes: the structure of a repetitive extragenic palindrome bound to its associated nuclease. *Nucleic Acids Res* **40**: 9964–9979.

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

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

Mizuuchi K (1992b) Polynucleotidyl transfer reactions in transpositional DNA recombination. *J Biol Chem* **267**: 21273–21276.

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

Montano SP & Rice PA (2011) Moving DNA around: DNA transposition and retroviral integration. *Curr Opin Struct Biol* **21**: 370–378.

Montano SP, Pigli YZ & Rice PA (2012) The mu transpososome structure sheds light on DDE recombinase evolution. *Nature* **491**: 413–417.

Moran NA (1996) Accelerated evolution and Muller's rachet in endosymbiotic bacteria. *P Natl Acad Sci USA* **93**: 2873–2878.

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

Moulin L, Rahmouni AR & Boccard F (2005) Topological insulators inhibit diffusion of transcription-induced positive supercoils in the chromosome of *Escherichia coli*. *Mol Microbiol* **55**: 601–610.

Moya A, Pereto J, Gil R & Latorre A (2008) Learning how to live together: genomic insights into prokaryote-animal symbioses. *Nat Rev Genet* **9**: 218–229.

Murphy DB & Pembroke JT (1995) Transfer of the IncJ plasmid R391 to recombination deficient *Escherichia coli* K12: evidence that R391 behaves as a conjugal transposon. *FEMS Microbiol Lett* **134**: 153–158.

Naas T, Blot M, Fitch WM & Arber W (1994) Insertion sequence-related genetic variation in resting *Escherichia coli* K-12. *Genetics* **136**: 721–730.

Naas T, Blot M, Fitch WM & Arber W (1995) Dynamics of IS-related genetic rearrangements in resting *Escherichia coli* K-12. *Mol Biol Evol* **12**: 198–207.

Namy O, Rousset JP, Napthine S & Brierley I (2004) Reprogrammed genetic decoding in cellular gene expression. *Mol Cell* **13**: 157–168.

Nunvar J, Huckova T & Licha I (2010) Identification and characterization of repetitive extragenic palindromes (REP)-associated tyrosine transposases: implications for REP evolution and dynamics in bacterial genomes. *BMC Genomics* **11**: 44.

Nunvar J, Licha I & Schneider B (2013) Evolution of REP diversity: a comparative study. *BMC Genomics* **14**: 385.

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

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.

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.

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.

Olliver A, Valle M, Chaslus-Dancla E & Cloeckaert A (2005) Overexpression of the multidrug efflux operon acrEF by insertional activation with IS1 or IS10 elements in *Salmonella enterica* serovar typhimurium DT204 acrB mutants selected with fluoroquinolones. *Antimicrob Agents Chemother* **49**: 289–301.

Papke RT, Ramsing NB, Bateson MM & Ward DM (2003) Geographical isolation in hot spring cyanobacteria. *Environ Microbiol* **5**: 650–659.

Parkhill J & Thomson N (2003) Evolutionary strategies of human pathogens. *Cold Spring Harb Symp Quant Biol* **68**: 151–158.

Parkhill J, Sebaihia M, Preston A et al. (2003) Comparative analysis of the genome sequences of *Bordetella pertussis*, *Bordetella parapertussis* and *Bordetella bronchiseptica*. *Nat Genet* **35**: 32–40.

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.

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.

Pasternak C, Ton-Hoang B, Coste G, Bailone A, Chandler M & Sommer S (2010) Irradiation-induced *Deinococcus radiodurans* genome fragmentation triggers transposition of a single resident insertion sequence. *PLoS Genet* **6**: e1000799.

Pasternak C, Dulermo R, Ton-Hoang B *et al.* (2013) ISDra2 transposition in *Deinococcus radiodurans* is downregulated by TnpB. *Mol Microbiol* **88**: 443–455.

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.

Pembroke JT & Murphy DB (2000) Isolation and analysis of a circular form of the IncJ conjugative transposon-like elements, R391 and R997: implications for IncJ incompatibility. *FEMS Microbiol Lett* **187**: 133–138.

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

Polard P, Prere MF, Chandler M & Fayet O (1991) Programmed translational frameshifting and initiation at an AUU codon in gene expression of bacterial insertion sequence IS911. *J Mol Biol* **222**: 465–477.

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

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.

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.

Preston A, Parkhill J & Maskell DJ (2004) The *Bordetellae*: lessons from genomics. *Nat Rev Micro* **2**: 379–390.

Qi X, Daily K, Nguyen K *et al.* (2012) Retrotransposon profiling of RNA polymerase III initiation sites. *Genome Res* **22**: 681–692.

Read TD, Fraser CM, Hsia RC & Bavoil PM (2000) Comparative analysis of *Chlamydia* bacteriophages reveals variation localized to a putative receptor binding domain. *Microb Comp Genomics* **5**: 223–231.

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.

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.

Rice P, Craigie R & Davies DR (1996) Retroviral integrases and their cousins. *Curr Opin Struct Biol* **6**: 76–83.

Richardson JM, Colloms SD, Finnegan DJ & Walkinshaw MD (2009) Molecular architecture of the Mos1 paired-end complex: the structural basis of DNA transposition in a eukaryote. *Cell* **138**: 1096–1108.

Ricker N, Qian H & Fulthorpe RR (2013) Phylogeny and organization of recombinase in trio (RIT) elements. *Plasmid* **70**: 226–239.

Rio DC (1991) Regulation of *Drosophila* P element transposition. *Trends Genet* **7**: 282–287.

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.

Rolland T, Neuveglise C, Sacerdot C & Dujon B (2009) Insertion of horizontally transferred genes within conserved syntenic regions of yeast genomes. *PLoS One* **4**: e6515.

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.

Rousseau P, Gueguen E, Duval-Valentin G & Chandler M (2004) The helix-turn-helix motif of bacterial insertion sequence IS911 transposase is required for DNA binding. *Nucleic Acids Res* **32**: 1335–1344.

Rousseau P, Loot C, Guynet C, Ah-Seng Y, Ton-Hoang B & Chandler M (2007) Control of IS911 target selection: how OrfA may ensure IS dispersion. *Mol Microbiol* **63**: 1701–1709.

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.

Salvatore P, Pagliarulo C, Colicchio R *et al.* (2001) Identification, characterization, and variable expression of a naturally occurring inhibitor protein of IS1106 transposase in clinical isolates of *Neisseria meningitidis*. *Infect Immun* **69**: 7425–7436.

Schmitz-Esser S, Penz T, Spang A & Horn M (2011) A bacterial genome in transition–an exceptional enrichment of IS elements but lack of evidence for recent transposition in the symbiont *Amoebophilus asiaticus*. *BMC Evol Biol* **11**: 270.

Sekine Y & Ohtsubo E (1989) Frameshifting is required for production of the transposase encoded by insertion sequence 1. *P Natl Acad Sci USA* **86**: 4609–4613.

Sekine Y, Eisaki N & Ohtsubo E (1994) Translational control in production of transposase and in transposition of insertion sequence IS3. *J Mol Biol* **235**: 1406–1420.

Shaheen M, Williamson E, Nickoloff J, Lee SH & Hromas R (2010) Metnase/SETMAR: a domesticated primate transposase that enhances DNA repair, replication, and decatenation. *Genetica* **138**: 559–566.

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

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.

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.

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.

Silva FJ, Latorre A & Moya A (2003) Why are the genomes of endosymbiotic bacteria so stable? *Trends Genet* **19**: 176–180.

Sinzelle L, Izsvak Z & Ivics Z (2009) Molecular domestication of transposable elements: from detrimental parasites to useful host genes. *Cell Mol Life Sci* **66**: 1073–1093.

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.

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

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.

Söll D (2007) The genetic code revisited — four decades after Francis Crick. *Nucleic Acids Symp Ser* **51**: 13–14.

Spradling AC, Bellen HJ & Hoskins RA (2011) *Drosophila* P elements preferentially transpose to replication origins. *P Natl Acad Sci USA* **108**: 15948–15953.

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.

Szeverenyi I, Hodel A, Arber W & Olasz F (1996) Vector for IS element entrapment and functional characterization based on turning on expression of distal promoterless genes. *Gene* **174**: 103–110.

Tettelin H, Riley D, Cattuto C & Medini D (2008) Comparative genomics: the bacterial pan-genome. *Curr Opin Microbiol* **11**: 472–477.

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.

Thompson JD, Higgins DG & Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**: 4673–4680.

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

Tobes R & Ramos J-L (2005) REP code: defining bacterial identity in extragenic space. *Environ Microbiol* **7**: 225–228.

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.

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.

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

Toleman MA, Bennett PM & Walsh TR (2006) ISCR elements: novel gene-capturing systems of the 21st century? *Microbiol Mol Biol Rev* **70**: 296–316.

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.

Ton-Hoang B, Polard P & Chandler M (1998) Efficient transposition of IS911 circles *in vitro*. *EMBO J* **17**: 1169–1181.

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.

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.

Ton-Hoang B, Pasternak C, Siguier P *et al.* (2010) Single-stranded DNA transposition is coupled to host replication. *Cell* **142**: 398–408.

Ton-Hoang B, Siguier P, Quentin Y, Onillon S, Marty B, Fichant G & Chandler M (2012) Structuring the bacterial genome: Y1-transposases associated with REP-BIME sequences. *Nucleic Acids Res* **40**: 3596–3609.

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

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

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

Van Dongen S (2000) A cluster algorithm for graphs. *Technical Report INS-R0010, National Research Institute for Mathematics and Computer Science in the Netherlands, Amsterdam.*

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

Vanhooff V, Galloy C, Agaisse H, Lereclus D, Revet B & Hallet B (2006) Self-control in DNA site-specific recombination mediated by the tyrosine recombinase TnpI. *Mol Microbiol* **60**: 617–629.

Varani A, Siguier P, Gourbeyre E, Charneau V, & Chandler M (2011) ISsaga is an ensemble of web-based methods for high throughput identification and semi-automatic annotation of insertion sequences in prokaryotic genomes. *Genome Biol* **12**: R30.

Vos JC, van Luenen HG & Plasterk RH (1993) Characterization of the *Caenorhabditis elegans*

Tc1 transposase *in vivo* and *in vitro*. *Genes Dev* **7**: 1244–1253.

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

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.

Wolter DJ, Smith-Moland E, Goering RV, Hanson ND & Lister PD (2004) Multidrug resistance associated with mexXY expression in clinical isolates of Pseudomonas aeruginosa from a Texas hospital. *Diagn Microbiol Infect Dis* **50**: 43–50.

Yano H, Genka H, Ohtsubo Y, Nagata Y, Top EM & Tsuda M (2013) Cointegrate-resolution of toluene-catabolic transposon Tn4651: determination of crossover site and the segment required for full resolution activity. *Plasmid* **69**: 24–35.

Zabala JC, de la Cruz F & Ortiz JM (1982) Several copies of the same insertion sequence are present in alpha-hemolytic plasmids belonging to four different incompatibility groups. *J Bacteriol* **151**: 472–476.

Zahradka K, Slade D, Bailone A *et al.* (2006) Reassembly of shattered chromosomes in *Deinococcus radiodurans*. *Nature* **443**: 569–573.

Zerbib D, Jakowec M, Prentki P, Galas DJ & Chandler M (1987) Expression of proteins essential for IS1 transposition: specific binding of InsA to the ends of IS1. *EMBO J* **6**: 3163–3169.

Zerbib D, Polard P, Escoubas JM, Galas D & Chandler M (1990) The regulatory role of the IS1-encoded InsA protein in transposition. *Mol Microbiol* **4**: 471–477.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1. ISs in genomes.
Table S2. Gene activation by IS insertion.