Werner Arber

Genetic Bases for Evolutionary Development of Microorganisms[1]

Abstract

A synoptic view is given here to molecular mechanisms promoting and limiting the generation of genetic variation. These mechanisms include (a) the production of conventional spontaneous mutations resulting from DNA replication infidelity and from the action of environmental mutagens, (b) enzymatically mediated repair processes acting on DNA mismatches caused by those processes of mutagenesis listed under (a), (c) various types of DNA rearrangements such as transposition and site-specific recombination at secondary crossing over sites, (d) DNA acquisition upon transformation, conjugation and phage-mediated transduction, and (e) processes modulating the efficiency of DNA acquisition such as effects of restriction-modification systems. Many of these processes are mediated by specific gene products and these often act as generators of genetic variations, hence in a not strictly reproducible way with regard to their site of action on DNA. It is proposed that at least part of the enzymes and organelles involved in these processes function primarily for the evolutionary development of microorganisms and had also been selected for this property.

Introduction

Biological evolution ensures a steady development, long-term maintenance and diversification of life on earth. In this contribution I discuss evidence for the view that the evolutionary process does not only rely on illegitimate chance events, but is importantly influenced by genetically determined biological functions with evolutionary implications. A number of arguments taken from microbial genetics will give support to the following *thesis*:

> Specific genes carried in the genome (and on accessory genetic elements) encode products that fulfil evolutionary functions
> 1. by generating genetic variation or
> 2. by limiting genetic plasticity to tolerable, but evolutionarily useful levels.
>
> In spite of this genetic determination, biological evolution is not directed.

[1] This is a modified and extended version of a contribution to the UNESCO-COGENE Symposium 'Uniqueness and Universality in a Biological World' to be published in 'Biology International' by the International Union of Biological Sciences' (IUBS)

Prokaryotes as objects of studies on mechanisms of biological evolution

Evolutionary processes are known to rely on mutation, selection and isolation. They can ideally be studied with haploid microorganisms which rapidly manifest phenotypic alterations due to mutation events. Bacteria and their viruses have extremely short generation times and are ideally suited for population genetic investigations, e.g. for studies on competition between mixtures of parental and mutant types submitted to various selection pressures. In addition, molecular genetic approaches can reveal the molecular nature of individual mutations, such as nucleotide substitution, small deletions and insertions, and larger DNA rearrangements. For simplicity, spontaneous mutation will here be defined as any alteration occurring to DNA sequences without an intended intervention by an investigator. More often such mutations will be detrimental, sometimes even lethal, than beneficial to the organisms. Therefore, tolerable mutation rates should be smaller than one mutation per genome and per generation. However, we should be aware that generation time is difficult to define for resting bacteria in their stationary phase. In cultures of exponentially growing *E. coli* bacteria the rate of mutagenesis is in the order of 10^{-2} new genetic alterations per cell and per generation. This results in a relatively high degree of genetic polymorphism in colonies grown from a single cell. Natural, spontaneous mutagenesis thus seriously limits the size of clones formed by genetically fully identical individuals upon propagation of bacteria. By mutations occurring in phases of rest, members of pure clones undergo further genetic diversification.

Sources of genetic variation

Many mechanistically different processes contribute in parallel to the generation of mutations. For this discussion, we group these processes into four categories (Arber, 1991 and 1993):

a. reproductive infidelity
b. effects of environmental and internal mutagens
c. DNA rearrangements
d. DNA acquisition

The action of these processes on individuals in large populations generates new genetic diversity (Fig. 1). However, overall genetic diversity is kept in balance by natural selection and, after all, the size of the biosphere, which can hold in the order of 10^{30} living cells. In addition, the efficiency of processes of the first two categories, reproductive infidelity and effects of environmental mutagens, is considerably attenuated by the activity of various enzymatic repair processes, while DNA acquisition (category d) encounters a number of natural limits to DNA transfer.

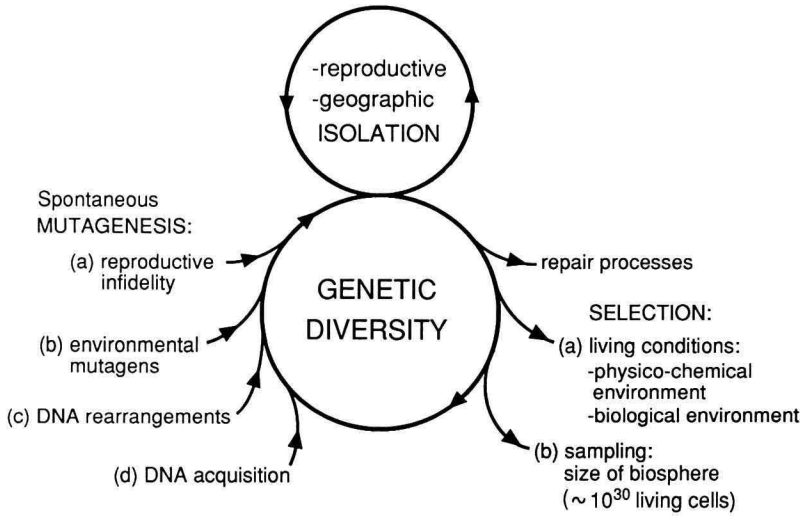

Fig. 1. Genetic diversity is steadily increased by various mutagenesis processes generating genetic variation, while natural selection acting on mixed populations of organisms limits the degree of diversity. Repair processes reduce the probability for some primary alterations on DNA to become fixed as mutations. Geographic and reproductive isolation can modulate genetic diversity. From Arber et al. (1994).

It is generally thought that a major source of nucleotide substitution is the occurrence of short-living tautomeric forms of the nucleotides, that are structural variants of the normal forms presenting different specificity of base pairing. A mispairing which results when an incorporated tautomeric base reassumes its normal form should thus not be qualified as a mistake in the incorporation, it is rather the consequence of a statistically occurring structural variation of a biochemical compound. Many, but not all of such cases of primary infidelity are efficiently repaired before such mutations become fixed.

DNA rearrangements

The following discussion on DNA rearrangements will be limited to bacterial systems. In these haploid organisms, homologous recombination cannot be attributed the same role in the generation of genomic diversity as is done for higher, sexually reproducing organisms with diploid genomes, although bacterial conjugation may sometimes substitute for the lacking recombinational reassociation of alleles from different sets of chromosomes. General recombination can also bring about major alterations in the genome structure and content by unequal crossing over at homologous sequences located at different sites in the genome.

In bacteria, enzymatic systems of site-specific recombination and of transposition are widespread. These and still other processes, often referred to as

illegitimate recombination, widely contribute to genomic plasticity. Most of these processes are catalyzed by specific enzymes and thus result from the action of genetic determinants. Some of the DNA rearrangements mediated by these systems have been studied to great mechanistic details and are thus well understood. This is e.g. the case for site-specific DNA inversion (Glasgow et al., 1989) and transposition of IS elements (Galas and Chandler, 1989).

DNA inversion

Site-specific DNA inversion is a source both of gene fusion and operon fusion. In the well-studied genetic flip-flop systems, one of two - or in some cases more - possible, alternative genomic arrangements is periodically assumed. This process can rapidly result in mixed populations of individuals with different phenotypic properties, if different genomic structures influence gene expression differently. Examples are the connection or disconnection of a promoter with an open reading frame (Silverman et al., 1981) or the fusion of a variable part with a constant part of a gene (Giphart-Gassler et al., 1982; Iida, 1984). The sites of crossing over in this enzymatically mediated process are consensus DNA sequences. Deviations from the consensus can still serve in DNA inversion, although with different efficiencies (Iida and Hiestand-Nauer, 1986 and 1987).

	gix*	TTCctgTAAACC	AA	GGTTTTgGATAA	
a)	gix*Q391	Tatcacctctgg	AA	aGcaTacGgagt	(1)
b)	gix*Q390	aTagatgAttgC	AA	ccTgcTtacggg	(1)
c)	gix*Q389	cgCctgatAcag	AA	cGaaTTgctTgc	(2)
d)	gix*Q387	TTgggcTtcgga	At	cGTTTTccgggA	(5)
e)	gix*Q386	TTCTatgAAAgg	tt	GGgcTTcGgaAt	(1)
f)	gix*Q385	ccgcCtTctAtg	AA	aGgTTggGcTtc	(2)
g)	gix*Q384	actcttTttcCg	AA	GGTaactGgctt	(2)
h)	gix*Q388B	aTCTaggtgAag	At	ccTTTTtGATAA	(1)
i)	gix*Q393	aaaTtaaAAAtg	AA	GtTTTaaatcAA	(6)
k)	gix*Q392	aagTttTAAAtC	AA	tcTaaagtATAt	(1)

Fig. 2. Nucleotide sequences used as secondary crossing over sites in Cin-mediated site-specific DNA inversion. On plasmid pSHI383 rare DNA inversion between the natural crossing over site *gix** and a secondary crossing over site brought the expression of the kanamycin resistance gene *kan* under the control of either promoter lacUV5 (sites a to f) or promoter PI (sites g and h). The plasmid also underwent unequal cointegration using *gix** and either site i or k and resulting in the fusion of *kan* with the operon under control of promoter P*bla*. Nucleotides corresponding to the *dix* consensus sequence of efficient crossing over sites are shown as underlined capital letters. Numbers in parenthesis refer to the number of independent isolates having used the crossing over site in question. The data were pooled from Iida and Hiestand-Nauer (1987).

Interestingly, the reaction can still take place with very low probability on DNA sequences widely diverting from the consensus (Fig. 2). It is possible that short-term structural variations of the interacting partners, recombinase and its substrate DNA, thereby play a critical role. Many different DNA sequences can thus occasionally serve for DNA inversion. This rare use of secondary crossing over sites is thought to represent an important natural source of novel gene fusions and novel operon fusions with evolutionary relevance (Arber, 1990).

Transposition

Another source of genomic rearrangements is the transpositional activity of mobile genetic elements. A number of different such 'inserted sequence' elements, IS elements, reside each in a number of copies in bacterial genomes. Once in a while they undergo enzymatically mediated DNA rearrangements, which include simple transposition, the formation of an adjacent deletion and DNA inversion, as well as the cointegration of plasmids or of a plasmid with the chromosome.

Interestingly, transposition also occurs in resting bacteria. E.g., bacterial subclones kept alive for decades in a stab culture accumulate genetic polymorphism due to transposition, and the resulting diversity increases linearly with the time of storage of the bacteria (Naas et al., 1994 and 1995). The degree of genetic diversity thus obtained largely depends on the target selection criteria of the participating IS elements. Depending on the IS element these criteria show different levels of sequence specificities.

A relatively strong target specificity is e.g. exerted by IS*30*, although other target sequences than the preferred ones are also used occasionally (Caspers et al., 1984; Stalder and Arber, 1989). The 1221-bp IS*30* element is a resident of *E. coli* K12. A constitutively expressed open reading frame covers practically the entire length of the element and encodes the transposase (Dalrymple et al., 1984). This is a bifunctional DNA recombinase and it efficiently mediates site-specific recombination and less efficiently transpositional DNA rearrangements (Olasz et al., 1993). The level of these activities are controlled by a leaky transcription terminator located within the reading frame of the transposase (Dalrymple and Arber, 1986), a constitutively expressed anti-sense RNA complementary to the middle part of the transposase gene (A. Arini, M. Keller and W. Arber, to be published) and possibly by an autorepression exerted by the transposase itself (Stalder et al., 1990).

Evidence has been obtained that IS*30* transposition goes through a structural intermediate, $(IS30)_2$ formed by two directly repeated IS*30* elements separated by normally a 2-bp spacer (Olasz et al., 1993). This intermediate structure originates from the action of the site-specific recombinase, and its resolution is usually also brought about by site-specific deletion formation, which occurs upon standard growth conditions with a frequency in the order of 10^{-1} per cell

and per generation. However, $(IS30)_2$ also gives rise to different kinds of transpositional DNA rearrangements such as intramolecular DNA inversion and deletion formation, as well as intermolecular inverse transposition (Fig. 3). In the presence of efficient target sequences for IS*30* transposition, these DNA rearrangements occur with frequencies in the range of 10^{-3} per cell and per generation. For these reasons, IS*30*, which is in principle remarkably stable, can occasionally give rise to a burst of transposition, once an intermediate form has been produced. In such a burst different subclones with various DNA rearrangements can arise with high frequency, thereby providing a population of bacteria with different genetic variants (Naas et al., 1994).

That a large fraction of lethal mutations is also caused by transposition of IS elements had been demonstrated by a study of bacteriophage P1 mutations affecting the vegetative reproduction of the phage but not its maintenance as prophage in the lysogenic condition (Arber et al., 1981; Sengstag and Arber, 1983). In these studies about 95% of the detected lethal mutations were caused by the insertion of an IS element originating from the bacterial chromosome. A study of the integration sites of these IS elements on the P1 prophage showed that most IS elements did not randomly select their target sites. Rather, some elements such as IS*30* highly preferred a particular site as already mentioned,

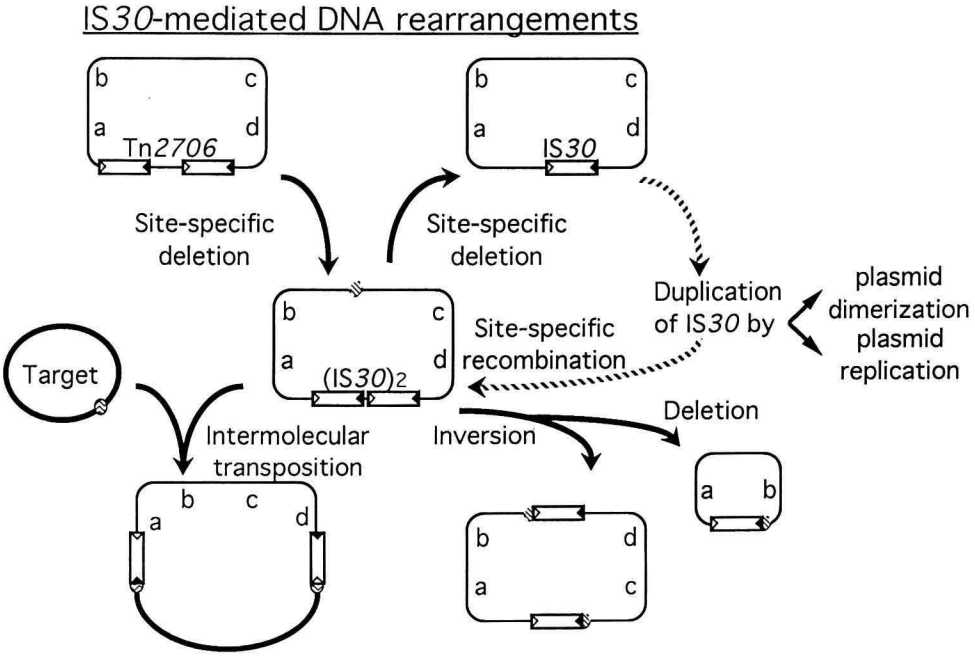

Fig. 3. Pathways of IS*30*-mediated DNA rearrangements. The recombinogenic $(IS30)_2$ structure shown in the center results from site-specific recombination involving the two IS*30* ends. The $(IS30)_2$ structure can resolve by site-specific deletion, but it also gives rise to transpositional DNA rearrangements shown in the lower part of the figure. Drawn after Olasz et al. (1993) and reproduced from Arber et al. (1994).

while IS*2*, the element most active under the experimental conditions used, preferred particular regions of the phage genome for its insertion, but within these regions, many different target sequences were used (Fig. 4). This illustrates how some genome regions can more often be affected by spontaneous mutation than others and how recombination processes can occur at a large number of possible crossing over sites, each of which may have its characteristic probability to serve for transposition. We thus count transposable genetic elements to genetically determined variation generators.

Composite transposons are defined as two identical IS elements flanking one or several genes unrelated to the transposition process. Composite transposons can originate when two copies of the same IS element subsequently insert into different sites of a DNA segment (Iida et al., 1981). Although the two participating IS elements can still transpose alone, they sometimes transpose together as a unit with the DNA segment carried between them. This can happen intramolecularly as well as intermolecularly, e.g. to a natural gene vector such as a conjugative plasmid or a phage genome. This then opens the possibility of horizontal transfer not only of the IS element involved, but also of the gene(s) carried between the two IS elements.

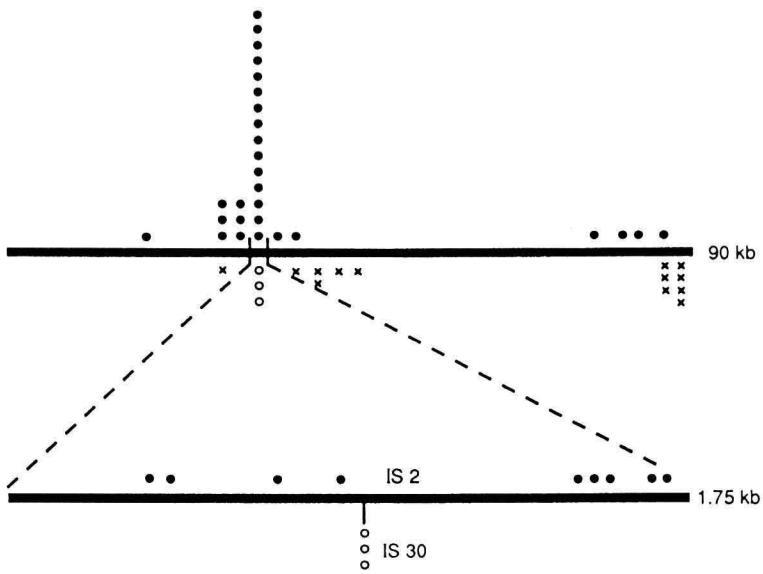

Fig. 4. Location of independent IS insertions into the genome of bacteriophage P1 and resulting in mutants affected in the vegetative reproduction of the phage. The circular 90-kb genome of P1 is shown linearized in the upper part. Dots shown above the genome identify independent IS*2* insertions, crosses shown below the genome refer to insertions of IS*1*, IS*3*, IS*5*, Tn*1000* and circles to IS*30* insertions. As shown in the lower part, each of nine sequenced IS*2* insertions into the hot region for IS*2* transposition had occurred into a different sequence, while the 3 independent IS*30* insertions had occurred between the same base pairs and both orientations had been used. After Sengstag and Arber (1983).

Gene acquisition

The process of horizontal gene transfer has been widely documented, e.g. by the horizontal spreading of genetic determinants for antibiotic resistances. This latter example also nicely illustrates the important role played by selection for genetically altered derivatives in dependence of changes in the environmental conditions.

The acquisition of genetic information from a donor by a receptor strain is at the basis of classical microbial genetics, i.e. (1) transformation of a receptor strain by the uptake of free DNA originating from a donor strain, (2) conjugation, in which donor and receptor bacteria enter in close contact and in which a conjugative plasmid serves as vector for the transfer of donor genes to recipient bacteria, and (3) phage-mediated transduction, in which a viral genome serves as gene vector. In all of these processes gene transfer is followed by the establishment of the acquired genes in the receptor cell. This can be brought about by a recombination process or else by the establishment of the transferred vector together with its passenger DNA as an autonomous replicon.

As was already mentioned, various limitations reduce the efficiency of gene acquisition. These limits include (a) the requirement of surface compatibilities for the DNA uptake in transformation, for interaction between donor and receptor cells in conjugation and for phage infection in transduction, (b) the action of restriction-modification systems on penetrating DNA molecules, (c) the requirements for the already described establishment step, and finally (d) the responses given by the receptor cell to the expression of the acquired functions, which may risk to perturb the functional harmony of the host cell. This latter limitation is less severe for the acquisition of only small portions of genetic information, a condition strongly favored by the action of restriction endonucleases.

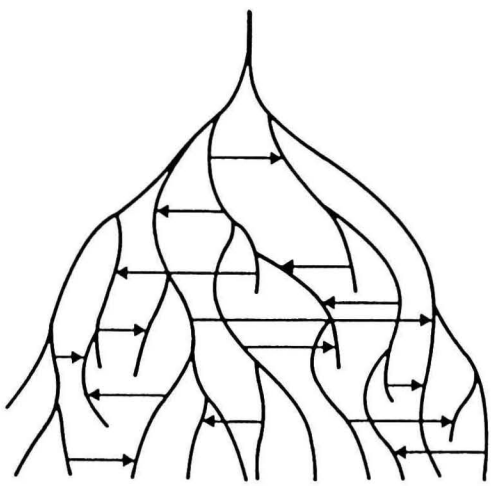

Fig. 5. The tree of microbial evolution drawn with horizontal connectors to symbolize the occasional occurrence of horizontal gene transfer between different branches of the tree.

These enzymes cleave large DNA molecules into small fragments, the free ends of which are recombinogenic. DNA acquisition thus follows a strategy of acquisition in small steps.

In view of the multitude of processes contributing not only to vertical biological evolution by alterations occurring within the genome of an organism, but also to horizontal biological evolution by the occasional acquisition of small portions of foreign genetic information, the evolutionary tree should schematically be drawn with horizontal connections allowing for a gene flux between different branches (Fig. 5).

Conclusions

It should be emphasised that the mechanistically different processes providing genetic diversity can only partially substitute for each other. Rather, they often fulfil different evolutionary functions. This can be seen in a comparison of the processes of (1) nucleotide substitution resulting from infidelity upon DNA replication, (2) intragenomic DNA rearrangements and (3) DNA acquisition. The first of these processes serves in the strategy to stepwise develop new biological functions and to improve and adjust available biological functions. The second process, DNA rearrangement by any of the described enzyme-mediated recombinations, can lead to an improvement of available capacities, particularly by the fusion of different functional domains and by the fusion of expression control signals with coding sequences thereby leading to different expression controls. Finally, the acquisition of sequence domains and motifs, of functional genes, and of clusters of genes offers the receptor cell a chance to profit of a successful development made by others.

The attribution of primarily evolutionary biological functions to DNA recombination systems acting as generators of genetic variations, to systems providing means for horizontal gene transfer, and also to natural limiters of genetic plasticity, such as mismatch repair systems or restriction-modification systems, is largely a matter of attitude of an investigator towards the object of his investigations, nature and life. Is it reasonable to assume that genetically encoded biological functions serve only to meet the needs of individual lives, by providing housekeeping and accessory functions required during the life span of the organisms? Alternatively, one can consider the process of a steady biological evolution as sufficiently important to justify the presence of genetic determinants specified and mainly responsible for both the past and the future development of a multitude of life forms able to adapt to changing living conditions and to withstand some of the contra-selective forces such as those found under extreme life conditions.

What has been described here for bacteria with a few selected examples might well have a more general validity. Analogous genetic variation generators act also in the development of variations in the immune response of higher animals. I am aware that a very strict subdivision of biological functions into (1) those

serving to maintain intact the cellular physiology, (2) those others serving for developmental purposes of multicellular organisms and (3) still others serving for the biological evolution of populations would not correspond to reality. Indeed, some genetically determined products serve for more than one of these purposes. But some specific gene products may very well primarily be used for the needs of biological evolution and they may also have been in the long term selected for this purpose. Since the selection of genetic variation generators is made mostly on variations concerning gene products which are not determined by the evolutionary genes in question, this kind of group selection - or selection between bacterial clones - must be exerted at the population level and must depend on the presence of enough appropriate variants to fulfil the selective needs. This can help to explain why in spite of the genetic determination of the evolutionary process, the direction of evolution remains undetermined and is a matter of interplay between the aleatoric occurrence of particular mutations and the action of natural selection exerted on the sustained mixed populations of organisms.

Affiliation of author

Department of Microbiology
Biozentrum, University of Basel
Klingelbertstr. 70
CH-4056 Basel, Switzerland
phone 061 267 2130
Fax 061 267 2118

References

Arber, W., 1990 - Mechanisms in microbial evolution. *J. Struct. Biol.* **104**, 107–111.

Arber, W., 1991 - Elements in microbial evolution. *J. Mol. Evol.* **33**, 4–12.

Arber, W., 1993 - Evolution of prokaryotic genomes. *Gene* **135**, 49–56.

Arber, W., Hümbelin, M., Caspers, P., Reif, H.J., Iida, S. and Meyer, J., 1981 - Spontaneous mutations in the *Escherichia coli* prophage P1 and IS-mediated processes. *Cold Spring Harbor Symp. Quant. Biol.* **45**, 38–40.

Arber, W., Naas, T. and Blot, M., 1994 - Generation of genetic diversity by DNA rearrangements in resting bacteria. *FEMS Microbiol. Ecol.* **15**, 5–14.

Caspers, P., Dalrymple, B., Iida, S. and Arber, W., 1984 - IS*30*, a new insertion sequence of *Escherichia coli* K12. *Mol. Gen. Genet.* **196**, 68–73.

Dalrymple, B. and Arber, W., 1986 - The characterization of terminators of RNA transcription on IS*30* and an analysis of their role in IS element-mediated polarity. *Gene* **44**, 1–10.

Dalrymple, B., Caspers, P. and Arber, W., 1984 - Nucleotide sequence of the prokaryotic mobile genetic element IS*30*. *EMBO J.* **3**, 2145–2149.
Galas, D.J. and Chandler, M., 1989 - Bacterial insertion sequences. In: Mobile DNA, Berg, D.E. and Howe, M.M. (Eds.), *Am. Soc. Microbiol.*, Washington, DC, pp. 109–162.
Giphart-Gassler, M., Plasterk, R.H.A. and van de Putte, P., 1982 - G inversion in bacteriophage Mu: a novel way of gene splicing. *Nature* **297**, 339–342.
Glasgow, A.C., Hughes, K.T. and Simon, M.I., 1989 - Bacterial DNA inversion systems. In: Mobile DNA, Berg, D.E. and Howe, M.M. (Eds.), *Am. Soc. Microbiol.*, Washington, DC, pp. 637–659.
Iida, S. and Hiestand-Nauer, R., 1986 - Localized conversion at the crossover sequences in the site-specific DNA inversion system of bacteriophage P1. *Cell* **45**, 71–79.
Iida, S. and Hiestand-Nauer, R., 1987 - Role of the central dinucleotide at the crossover sites for the selection of quasi sites in DNA inversion mediated by the site-specific Cin recombinase of phage P1. *Mol. Gen. Genet.* **208**, 464–468.
Iida, S., 1984 - Bacteriophage P1 carries two related sets of genes determining its host range in the invertible C segment of its genome. *Virology* **134**, 421–434.
Iida, S., Meyer, J. and Arber, W., 1981 - Genesis and natural history of IS-mediated transposons. *Cold Spring Harbor Symp. Quant. Biol.* **45**, 27–37.
Naas, T., Blot, M., Fitch, W. and Arber, W., 1995 - Dynamics of IS-related genetic rearrangements in resting *Escherichia coli* K12. *Mol. Biol. Evol.* **12**, 198–207.
Naas, T., Blot, M., Fitch, W.M. and Arber, W., 1994 - Insertion sequence-related genetic variation in resting *Escherichia coli* K-12. *Genetics* **136**, 721–730.
Olasz, F., Stalder, R. and Arber, W., 1993 - Formation of the tandem repeat $(IS30)_2$ and its role in IS*30*-mediated transpositional DNA rearrangements. *Mol. Gen. Genet.* **239**, 177–187.
Sengstag, C. and Arber, W., 1983 - IS*2* insertion is a major cause of spontaneous mutagenesis of the bacteriophage P1: non-random distribution of target sites. *EMBO J.* **2**, 67–71.
Silverman, M., Zieg, J., Mandel, G. and Simon, M., 1981 - Analysis of the functional components of the phase variation system. *Cold Spring Harbor Symp. Quant. Biol.* **45**, 17–26.
Stalder, R. and Arber, W., 1989 - Characterization of *in vitro* constructed IS*30*-flanked transposons, *Gene* **76**, 187–193.
Stalder, R., Caspers, P., Olasz, F. and 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.