Nomenclature of Transposable Elements in Prokaryotes¹

A. CAMPBELL

Department of Biological Sciences, Stanford University, Stanford California 94305

D. E. BERG

Department of Microbiology and Immunology, Washington University, St. Louis, Missouri 63110

E. M. LEDERBERG

Department of Medical Microbiology, Stanford University School of Medicine, Stanford, California 94305

P. STARLINGER

Institut für Genetik, Universität zu Köln, D-500 Köln, West Germany

D. BOTSTEIN

Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

R. P. NOVICK

Public Health Research Institute of the City of New York, New York, New York 10016

AND

W. SZYBALSKI

McArdle Laboratory for Cancer Research, University of Wisconsin, Madison, Wisconsin 53706

Received March 21, 1979

Transposable elements are defined as specific DNA segments that can repeatedly insert into a few or many sites in a genome. They are classified as simple IS elements, more complex Tn transposons, and self-replicating episomes. Definitions and nomenclature rules for these three classes of prokaryotic transposable elements are specified.

I. INTRODUCTION

At the Cold Spring Harbor meeting on DNA insertions, a committee was elected to draft a set of rules for nomenclature (Campbell *et al.*, 1977). These rules were based on discussions held at the meeting and on previously accepted conventions for the Mu prophage (Howe and Bade, 1975), for bacterial plasmids (Novick *et al.*, 1976), for bacteriophage λ (Hershey, 1971), and for bacterial genetics (Demerec *et al.*, 1966).

This article is a revised version of the original report (Campbell *et al.*, 1977)

modified in light of more recent developments. The most important relevant technical advances have been in the area of nucleotide sequencing. In 1977, it seemed necessary that the primary operational definition for relatedness of two genetic elements should be descent from a common ancestor. Two elements should be identical (or at least very similar) if descended in the laboratory from the same natural source. They are potentially different if derived from different sources. Common descent becomes superfluous when complete nucleotide sequences of both elements are known (except insofar as the sequence of the element used in a given experiment is generally inferred from that of some ancestor on which the sequencing was performed). Elements known to be identical in sequence

¹ In order to obtain wide distribution, this paper is being published concurrently in *Gene*, Vol. 5 (1979) No. 4. Suggestions for changes and amendments in future versions of these rules should be directed to the first author.

clearly should have the same name, even when independently derived.

Because of the range in size of known transposable elements (from 0.8 kb for IS*I* to over 100 kb for some of the conjugative plasmids) we can anticipate an extended interval in which some of the shorter elements (and perhaps numerous isolates thereof) will have been completely sequenced, while sequence data on the larger elements is still incomplete. We therefore retain common descent as the only criterion for relatedness that is directly applicable to all transposable elements recognizing where appropriate the implications of sequence information.

In the following report, the rules are presented in **boldface type**. Adherence to these rules in all publications about inserting elements is highly desirable. Interspersed with the rules are some comments, which include discussion and explanation of the rules, and also suggestions for possible usage in some cases not covered by the rules. The rules are intended to be minimal and applicable to current or clearly imminent situations. We have not tried to anticipate problems that may arise far in the future. We have also avoided prescribing rules for problems that arise currently but rarely. Many of these are best treated on an individual basis relevant to specific needs.

II. TRANSPOSABLE ELEMENTS

II. A. Definition

Transposable elements² are specific DNA segments that can repeatedly insert into a few or many sites in a genome.

II. B. Comments

The definition applies to elements ranging from phage Mu, which seems to insert at random, to phage λ , which usually, but not always, inserts at a single site in the *Escherichia coli* genome. Some (such as λ) have free phases that have been identified (e.g. intracellular replicons or extracellular viruses), whereas others (such as the IS or Tn elements) are presently known only in the inserted state.

III. CLASSES OF TRANSPOSABLE ELEMENTS

III. A-1. IS Elements

IS (Simple Insertion Sequences) elements contain no known genes unrelated to insertion function and are generally shorter than 2 kb.

III. A-2. Symbols

IS1, IS2, IS3, IS4, IS5, etc. with the numbers italicized. (Fiandt *et al.*, 1972; Hirsch *et al.*, 1972; Malamy *et al.*, 1972).

III. A-3. Comments

At present, the IS elements listed above are not known to contain any genes, either related or unrelated to insertion function. Our intention in specifying "genes unrelated to insertion function" is that even if insertion genes should be discovered in some IS elements, they still would be classified as simple insertion sequences. The term "gene" as used here means a DNA segment coding for a functional product and does not include, e.g., the "stop signals" responsible for polar effects of IS elements or promoters allowing IS elements to turn on adjacent genes.

Operationally, the failure to detect other genes in IS elements is closely tied to the absence of a recognizable phenotype conferred by their presence. The only observed consequences of IS insertion are *cis*-specific effects, such as changes in gene expression due to insertion within an operon, or the rate of incidence of adjacent deletions. Effects other than *cis*-specific ones

² The term "transposable element" is derived from the work of McClintock (1952) on mobile genetic elements in maize. In prokaryotes, an acceptable synonym is "translocation sequence" (Novick *et* al., 1976).

are expected to be unobservable because of the endogenous origin of IS elements. They are always present in the cell, not only at the site under examination, but at other, generally unidentified sites as well. One or more copy of each of the common IS elements is present in the *E. coli* genome. The phenotype of *E. coli* lacking, for example, IS1 is unknown.

Because several copies of each element are present in the genome, it is currently impossible to trace the lineage of a given IS element. When IS2 appears at a new site, we cannot say which of the several preexisting IS2s is ancestral to it. It must therefore clearly be borne in mind that a name such as IS2 is a generic term referring to all insertions that appear indistinguishable by hybridization, heteroduplexing, or restriction analysis, but which could have individual differences in base sequence and which might have undergone mutations or rearrangements in the laboratory (see IV. E).

There is presently no advantage to giving each IS2 in a different location a specific as well as a generic name. Such a designation would provide no additional information not already given by the mutation number (see below) that identifies the insertion event. This situation may change in the future for two reasons: (1) IS2 elements may be introduced into species from which they are naturally absent. IS2, for example, is said to be absent from Salmonella typhimurium. If a single copy of IS2 is introduced into S. typhimurium, then all subsequent transpositions within that species must be derived from that specific IS2. If and when such experiments are performed, all IS2s derived from an identified IS2 insertion should be assigned a specific designation (e.g., IS2.1) to distinguish them from IS2 insertions derived from another, potentially different source. (2) Laboratory mutants of IS elements recognizably different from wild type have been isolated. For example, Tomich and Friedman (1977) reported a mutant of IS2 (IS2 rip2) that has lost its polarity phenotype. (3) When the

complete base sequence of an IS insert has been determined, either it will prove identical to some previously sequenced insert or else minor variations in base sequence may come to light. Here again it is desirable to assign specific designations so that all inserts with identical sequences have the same designation and are distinguished from those known to differ from them.

III. B-1. Tn Elements (transposons)

Tn are more complex transposable elements, often containing two IS elements. They behave formally like IS elements but are generally larger than 2 kb and contain additional genes unrelated to insertion function.

III. B-2. Symbols

Tn1, Tn2, Tn3, etc. with the numbers italicized. A different number is assigned to each independent isolate from nature even if it is apparently identical to some previous isolate by criteria other than direct determination of nucleotide sequence. If the complete nucleotide sequence of a new isolate is shown to be identical to that of some previous isolate, the element should be designated by the symbol assigned to the first isolate with that sequence. If useful in a particular context, genes carried by a Tn element may be indicated following the name; e.g., Tn9 cam^+ [Tn9 is a transposable element that confers chloramphenicol resistance (cam^+)]. Mutations in Tn elements are designated by standard genetic symbols following the name of the element, e.g., Tn9 cam4 is a particular mutant of Tn9 that no longer confers chloramphenicol resistance.

III. B-3. Comments

Unlike IS elements, the lineage of Tn elements generally can be followed unambiguously. Hence Tn1, Tn2, etc. are specific rather than generic terms. Every isolate labeled Tn2 either has a pedigree in the laboratory that traces back to the original

Tn2 isolate or has been shown to be identical to that isolate in base sequence.

Note that the criterion is base sequence identity, not just extensive similarity. One might question whether this condition is perhaps too stringent, as it requires a new isolation number for a transposon that differs from Tn2 even by one or two base pairs. Such an element would clearly be related to Tn2 and might have arisen from Tn2 by mutation in the laboratory. However, it is precisely in such cases that it is important to keep straight which of the various Tn2like elements was used in a particular experiment.

In certain contexts, it may even be convenient to refer to a set of very similar transposons by some generic term. Since the publication of our original report, some workers have in fact preferred to designate any or all of the ampicillin transposons TnI - Tn2 as "TnA." We prescribe no rules for such generic terms. They are not an appropriate substitute for Tn numbers, as they provide no indication either of base sequence identity or of recent common ancestry. Their use may be appropriate and convenient in the text of a scientific paper, provided that the specific Tn number is indicated when the element is first mentioned.

Designations for some Tn elements described in the published literature are given in Table 1. In Table 1, "plasmid origin" indicates the natural plasmid from which the Tn element was derived. For example, Gottesman and Rosner (1975) studied transpositions of the chloramphenicol resistance determinant Tn9 from P1Cm to λ . Phage P1Cm originated from growth of phage P1 in a bacterium whose chloramphenicol resistance was derived from plasmid R14 (Kondo and Mitsuhashi, 1964). The R14 plasmid is now called pSM14 (Novick, 1974). The Tn3 element came either from pSC50, a derivative of R1 (Kopecko and Cohen, 1975) or has been transferred directly out of R1 (Heffron et al., 1975a).

The designations assigned in Table 1 should be used in all publications referring to elements listed here.

III. C. Episomes

Episomes are complex, self-replicating elements, often containing IS and Tn elements. Examples include bacteriophage λ and the F plasmid of *E. coli*. No changes in existing designations for these elements are proposed.

Transposable element	Plasmid origin	Resistance markers ^a	Reference
Tn/	RP4	Ар	Hedges and Jacob (1974)
Tn2	RSF1030	Ap	Heffron et al. (1975b)
Tn <i>3</i>	R1	Ap	Kopecko and Cohen (1975)
Tn4	R1	Ap,Sm,Su	Kopecko and Cohen (1974)
Tn <i>5</i>	JR 67	Km	Berg et al. (1975)
Tno	JR72	Km	Berg et al. (1975)
Tn7	R483	Tp,Sm	Barth et al. (1975)
Tn9	pSM14	Cm	Gottesman and Rosner (1975)
Tn10	R100	Tc	Foster et al. (1975); Kleckner et al. (1975)
Tn551	pI258	Em	Pattee et al. (1977)

TABLE 1

Tn Elements

^{*a*} Ap = amp = bla = ampicillin (β -lactamase); Sm = streptomycin; Su = sulfonamide; Km = kan = kanamycin; Tp = trimethoprim; Tc = tet = tetracycline; Cm = cam = chloramphenicol; Em = erythromycin.

IV. ORGANISMS AND GENOMES WITH INSERTED ELEMENTS

IV. A. Insertion in a Particular Genome

When an inserting element has been introduced into the genome of a previously described bacterial strain, or has been transposed within a strain, the new strain should be given an isolation number. The genotype of a strain that has acquired an inserting element can be denoted by the genotype of the parent strain, followed by the name of the element in parentheses:

$$R126 = W3350(\lambda) = F^{-}galK2galT1(\lambda).$$

Plasmids bearing IS or Tn elements should be designated according to the rules developed for plasmid nomenclature (Novick etal., 1976). In particular, each plasmid line derived from an independent insertion event should be assigned a new plasmid number.

IV. B. Insertion at a Particular Site

If location is known, specify gene or region in which element is inserted, followed by a number designating the particular insertion mutation, then by a double colon, and finally by name of inserted element.

- galT236::IS1 Mutation 236 in the galT gene is an IS1 insertion.
- hisG34::Tn10 Mutation 34 in the hisG gene is a Tn10 insertion.
- galPO-E490::IS2 Mutation 490 between galPO and galE is an IS2 insertion.
- λP -Qb1::IS2 Mutation b1 between genes P and Q is an IS2 insertion.
- F8 42kb-7::Tn2 Mutation 7 at 42 kb on F8 plasmid is a Tn2 insertion.

Designation of new insertion mutations should be by number. Previously published symbols may be used for mutations already named in the literature, e.g., MS348, r32, bi2, crg, etc. Old symbols should be changed only when it is generally agreed that they violate the rules "excessively" (Demerec et al., 1966). As in other areas of prokaryotic genetics, each published allele number must be unique. Individual laboratories engaged in parallel mutant isolations should endeavor by private communication to avoid assignment of the same number to different mutations prior to publication. (Numbers 1 and 0 should generally be avoided because they often are confused with letters L and O).

IV. C. Comments

The rules are intentionally nonspecific as to designations for insertions of known elements at unknown sites or unknown elements at known sites. For many purposes, strains carrying known elements at unknown sites are most conveniently treated as described above for insertions into a particular genome, postponing a designation of the specific insertion event until information on location is available.

Some laboratories may find it convenient to serialize insertions using the Greek capital omega to indicate that the number following refers to an insertion. This number would then be followed by a set of square brackets containing the available information about the nature and location of the insertion, thus:

- XY1000 = C600Ω100[K-12::Tn5] Tn5 somewhere in the K-12 genome, including chromosome and plasmids.
- $XY1100 = 8325\Omega 101[chr::Tn551] Tn551$ somewhere in the host chromosome.
- $XY1200 = C600\Omega102[galT236::IS1] IS1$ within chromosomal gene galT.
- $XY1300 = C600[pSC101\Omega103[4.5 kb::Tn1]]$ Tn1 at 4.5 kb on the pSC101plasmid.

Note that the assignment of *both* a strain number and an insertion number may in some cases be redundant; however, if the DNA segment carrying the insertion is transferred to a new strain, the insertion number should be retained since there has not been a new insertion, but the resulting *strain* would be a new one. Note also that as more information about the insertion becomes available, this can simply be included inside the square brackets; retention of the insertion number will provide assurance of the identity of the genetic segment in question.

Insertion of an uncharacterized element at a known site should be indicated simply by a gene designation and mutant number. The insertional nature of the mutation should not be part of its name unless or until the insertion has been identified; e.g., $biop_B131$ (insertion of unknown DNA into the biotin promoter). When the insertion and its location are better characterized, the designation should be modified, accordingly (e.g., $biop_B131::ISI$).

IV. D. Orientation

Where the orientation of the insertion is known, it may be desired to include that orientation in the strain description. Orientation may be specific either (a) with respect to direction along the genome or (b) with respect to the polarity of the operon.

Following (with a slight modification) the conventions adopted for phage Mu, orientation with respect to the genome is indicated by a plus (+) or minus (-) sign following the symbol of the element. For example, proA::Mu + . (A plus orientation means that if one moves around the *E. coli* map in a clockwise motion, the immunity end of the Mu prophage will be reached first and the *S* end last [Howe and Bade, 1975]). For IS and Tn elements, the plus orientation is arbitrary and the minus orientation is opposite to it.

Orientation with respect to the operon may be indicated with roman numerals I and II. Orientation I is assigned arbitrarily, usually when a new IS element is first characterized in an operon with known polarity:

For IS1, both orientations are polar. IS2 in orientation II is frequently constitutive (antipolar) because of readthrough from an IS2 promoter (Fiandt *et al.*, 1972; Saedler *et al.*, 1974).

IV. E. Mutations in Inserted Elements

Mutations are listed after the symbol for the element:

IV. F. Complex Genome Designations

Genetic symbols for a strain having two insertions and other mutations should be written in the same order as mapped on the genome:

-

If polarity is indicated, the above strain would be

 λ Wam43 b2 xis int-c60::IS2 · + cI857

P-Qb1::IS2 + Sam7 (genome polarity)

 λ Wam43 b2 xis int-c60::IS2 · II cI857

P-Qb1::IS2 · I Sam7 (operon polarity).

When both the inserted element and the genome into which it is inserted carry mutations, the symbols for the inserted element and its mutations may be set off by parentheses:

$$\lambda$$
Wam43 y r32::(IS2 rip2)·I Sam7.

V. CENTRAL REGISTRY

To avoid duplication of numbers, all new IS and Tn elements should be checked with a central registry before numerals are assigned to them in publications.

A registry for Tn elements is maintained by Dr. Esther Lederberg, Department of Medical Microbiology, Stanford University Medical School, Stanford, California 94305, as part of the Plasmid Reference Center.

CAMPBELL ET AL.

TABLE 2

Tn Nos.	Laboratory	Location
11-20	D. Botstein	Cambridge, Mass., U. S. A.
21-70	S. N. Cohen	Stanford, Calif., U. S. A.
71-100	N. Datta	London, England
101-200	D. Botstein	Cambridge, Mass., U. S. A.
201-300	S. N. Cohen	Stanford, Calif., U. S. A.
301-400	L. Rosner	Bethesda, Md., U. S. A.
401-500	J. Shapiro	Chicago, Ill., U. S. A.
501-520	V. Stanisich	Melbourne, Australia
521-550	D. Taylor, R. Grant	Toronto, Ontario, Canada
551-600	R. P. Novick	New York, N. Y., U. S. A.
601-700	J. Davies, D. E. Berg	Madison, Wisc. and St. Louis, Mo., U. S. A
701-750	N. Datta	London, England
751-800	Hammersmith Hospital	London, England
801-850	P. Bennett	Bristol, England
851-900	M. Syvanen	Cambridge, Mass., U. S. A.
901	E. Veltkamp	Amsterdam, Holland
902	J. Scott	Atlanta, Ga., U. S. A.
903	A. Oka	Kyoto, Japan
904-915	R. Olsen	Ann Arbor, Mich., U. S. A.
926-950	D. Kopecko, J. Wohlheiter	Washington, D. C., U. S. A.
951-1000	H. Saedler	Freiburg, West Germany
1001-1099	M. Guyer	Bethesda, Md., U. S. A.
1100-1400	S. Falkow	Seattle, Wash., U. S. A.
1401-1500	A. Summers, G. Jacoby	Boston, Mass., U. S. A.
1501-1520	A. I. Bukhari	Cold Spring Harbor, N. Y., U. S. A.
1521-1600	P. Courvalin	Paris, France
1601–1640	P. Kontomichalou	Athens, Greece
1641-1660	M. Van Montagu, J. Schell	Ghent, Belgium
1681–1695	F. Heffron	San Francisco, Calif., U. S. A.
1696–1700	C. Rubens	Seattle, Wash., U. S. A.
1701-1716	D. Nakada	Pittsburgh, Pa., U. S. A.
1721-1750	R. Schmitt	Regensburg, West Germany
1751-1770	G. Smirnov	Moscow, U. S. S. R.
1771-1800	F. Schöffl	Erlangen, West Germany
1801-1820	O. Marquardt	Hamburg, West Germany
1821-1830	H. Tschäpe	Wernigerode (Harz), East Germany
1831–1840	G. Lebek	Berne, Switzerland
1841-1860	B. Davey	Melbourne, Australia

REGISTRY OF TN NUMBER ALLOCATIONS THROUGH FEBRUARY 1979

Blocks of numbers have been assigned to investigators on request. Some of these assignments are listed in Table 2.

ACKNOWLEDGMENTS

The authors are grateful to Stanley Cohen and to many other individuals for suggestions and information. The activities of the Plasmid Reference Center are supported by NIH Grant AI-13093 and Contract NO1-AI-72531.

REFERENCES

- BARTH, P. T., DATTA, N., HEDGES, R. W., AND GRINTER, N. J. (1976). Transposition of a deoxyribonucleic acid sequence encoding trimethoprim and streptomycin resistances from R483 to other replicons. J. Bacteriol. 125, 800-810.
- BERG, D. E., DAVIES, J., ALLET, B., AND ROCHAIX, J. (1975). Transposition of R factor genes to bacteriophage λ . *Proc. Nat. Acad. Sci. USA* 72, 3628–3632.

- CAMPBELL, A., BERG, D., BOTSTEIN, D., LEDER-BERG, E., NOVICK, R., STARLINGER, P., AND SZYBALSKI, W. (1977). Nomenclature of transposable elements in prokaryotes. *In* DNA Insertion Elements, Plasmids and Episomes, (A. I. Bukhari, J. S. Shapiro, and S. L. Adhya, S. L. eds.), pp. 15-22. Cold Spring Harbor Laboratory, Cold Spring Harbor, N. Y.
- DEMEREC, M., ADELBERG, E. A., CLARK, A. J., AND HARTMAN, P. E. (1966). A proposal for a uniform nomenclature in bacterial genetics. *Genetics* 54, 61-76.
- FIANDT, M., SZYBALSKI, W., AND MALAMY, M. H. (1972). Polar mutations in *lac*, *gal*, and phage λ consist of a few IS DNA sequences inserted with either orientation. *Mol. Gen. Genet.* **119**, 223-231.
- FOSTER, T. J., HOWE, T. G. B., AND RICHMOND, M. H. (1975). Translocation of the tetracycline resistance determinant from R100-1 to the Escherichia coli K-12 chromosome. J. Bacteriol. 124, 1153-1158.
- GOTTESMAN, M., AND ROSNER, J. L. (1975). Acquisition of a determinant for chloramphenicol resistance by coliphage λ . *Proc. Nat. Acad. Sci. USA* 72, 5041-5045.
- HEDGES, R. W., AND JACOB, A. E. (1974). Transposition of ampicillin resistance from RP4 to other replicons. *Mol. Gen. Genet.* 132, 31-40.
- HEFFRON, F., RUBENS, C., AND FALKOW, S. (1975a). Translocation of a plasmid DNA sequence which mediates ampicillin resistance: Molecular nature and specificity of insertion. *Proc. Nat. Acad. Sci. USA* 72, 3623-3627.
- HEFFRON, F., SUBLETT, R., HEDGES, R. W., JACOB, A., AND FALKOW, S. (1975b). Origin of the TEM beta lactamase gene found on plasmids. J. Bacteriol. 122, 250-256.
- HERSHEY, A. D. (Ed.) (1971). "The Bacteriophage Lambda." Cold Spring Harbor Laboratory, Cold Spring Harbor, N. Y.
- HIRSCH, H.-J., STARLINGER, P., AND BRACHET, P. (1972). Two kinds of insertions in bacterial genes. *Mol. Gen. Genet.* 119, 191-206.
- Howe, M., AND BADE, E. (1975). Molecular biology of bacteriophage Mu. Science 190, 624-632.

- KLECKNER, N., CHAN, R., TYE, B., AND BOTSTEIN, D. (1975). Mutagenesis by insertion of a drug-resistance element carrying an inverted repetition. J. Mol. Biol. 97, 561-575.
- KONDO, E., AND MITSUHASHI, S. (1964). Drug resistance of enteric bacteria. IV. Active transducing bacteriophage PlCM produced by the combination of R factor with bacteriophage Pl. J. Bacteriol. 88, 1266-1276.
- KOPECKO, D. J., AND COHEN, S. N. (1975). Sitespecific recA-independent recombination between bacterial plasmids: Involvement of palindromes at the recombinational loci. Proc. Nat. Acad. Sci. USA 72, 1373-1377.
- MALAMY, M. H., FIANDT, M., AND SZYBALSKI, W. (1972). Electron microscopy of polar insertions in the *lac* operon of *Escherichia coli*. Mol. Gen. Genet. 119, 207-222.
- MCCLINTOCK, B. (1952). Chromosome organization and gene expression. Cold Spring Harbor Symp. Quant. Biol. 16, 13-47.
- NOVICK, R. P. (1974). Bacterial plasmids. In "Handbook of Microbiology" (A. I. Laskin and H. A. Lechevalier, eds.), Vol. IV, pp. 537-586. Chem. Rubber Co., Cleveland.
- NOVICK, R. P., CLOWES, R. C., COHEN, S. N., CURTISS, R., III, DATTA, N., AND FALKOW, S. (1976). Uniform nomenclature for bacterial plasmids: A proposal. *Bacteriol. Rev.* 40, 168-189.
- PATTEE, P. A., THOMPSON, N. E., HAUBRICH, D., AND NOVICK, R. P. (1977). Chromosomal map locations of integrated plasmids and related elements in *Staphylococcus aureus*. *Plasmid* 1, 38-51.
- SAEDLER, H., REIF, H. J., HU, S., AND DAVIDSON, N. (1974). IS2, a genetic element for turn-off and turn-on of gene activity in *E. coli. Mol. Gen. Genet.* 132, 265-289.
- TOMICH, P. K., AND FRIEDMAN, D. I. (1977). Isolation of mutations in insertion sequences that relieve IS-induced polarity. In "DNA Insertion Elements, Plasmids, and Episomes" (A. I. Bukhari, J. A. Shapiro, and S. L. Adhya, eds.), pp. 99-107. Cold Spring Harbor Laboratory, Cold Spring Harbor, N. Y.