Tn917 Transposase. Sequence Correction Reveals a Single Open Reading Frame Corresponding to the *tnpA* Determinant of Tn3-Family Elements

FLORENCE Y. AN* AND DON B. CLEWELL*,†

*Department of Biologic and Materials Sciences, School of Dentistry, and †Department of Microbiology and Immunology, School of Medicine, The University of Michigan, Ann Arbor, Michigan 48109

Received October 9, 1990; revised January 22, 1991

A nucleotide sequence correction on the *Enterococcus faecalis* transposon Tn917 indicates that what was formerly thought to be two open reading frames (ORF5 and ORF6) is actually one reading frame (ORF5). The latter exhibits homology with the Tn3-family transposase determinants known as tnpA. © 1991 Academic Press. Inc.

Tn917 was originally identified on the plasmid pAD2 in the multiply antibiotic resistant *Enterococcus faecalis* strain DS16 (Tomich *et al.*, 1979, 1980). It encodes an inducible MLS (macrolide, lincosamide, and streptogramin B) resistance and is a member of the Tn3-family of transposons (Heffron, 1983; Sherratt, 1989). There is a close general resemblance to the staphylococcal transposon Tn551 (Perkins and Youngman, 1984; Khan and Novick, 1980), except that the latter expresses MLS resistance constitutively.

The complete nucleotide sequence of Tn917 has been previously reported (Shaw and Clewell, 1985). Three open reading frames were found to be associated with transposition functions. One of these, referred to as ORF4, was found to exhibit homology with Tn3-family resolvase determinant (tnpR). The other two open reading frames, ORF5 and ORF6, appeared to be associated with transposase activity (Perkins and Youngman, 1984; Shaw and Clewell, 1985). Although homology of the latter two reading frames with Tn3-family transposase (tnpA) was not noted, subsequent computer comparisons indicated significant homology indeed existed (Murphy, 1989). This prompted a re-

The sequence data from this article has been deposited with the EMBL/GenBank Data Libraries under Accession No. M36722. examination of this region of the transposon, the results of which are reported here. Correction of sequencing errors (mainly frameshiftrelated) has revealed that ORF5 and ORF6 are actually combined and represent a single ORF (now ORF5 or *tnpA*).

Figure 1 shows the corrected sequence of tnpA. The protein product is inferred to consist of 973 amino acid residues and to have a molecular weight of 112,205. This compares favorably with the tnpA products of Tn3 family elements (Sherratt, 1989). The 3' end of the reading frame is 17 bp from an inverted repeat segment (positions 3098 through 3140) which would form a stem consisting of 16 bp with an 11-bp intervening loop (ΔG is -25.4 kcal [Tinoco *et al.*, 1973]). The latter might correspond to a transcription terminator; although subsequent multiple T's, generally associated with ρ -independent transcription terminators, are not present.

The corrections and their positions were as follows: G converted to C at position 687; insert A at 1803; insert G at 1804; delete G between 1807 and 1808; delete T between 1905 and 1906; insert G at 2186; insert G at 2809; insert T at 3079; insert G at 3082.

Figure 2 presents a homology matrix relating to the *tnpA* products of Tn501 (Brown *et al.*, 1985), Tn21 (Ward and Grinsted, 1987), and Tn3 (Heffron *et al.*, 1979); significant homology is clearly evident. The degrees of homology (identical residues) are 32, 32, and

20 Gaggaaaaaagggggggggggggggggggggggggggg	40 CTATCAATTGATTTAGC <u>TTT</u>	60 [AAAAATGTATGACAGCAAAG	80 AG <u>TATTCT</u> ATTCGTCAAAT	100 120 Істтратрестсталатталаласа
140 Ассттттассоттассталаала <u>сс</u> татостта	160 Agatatggctatgaaaagaa	-35 180 ATTTTAACTACTTCACAGCGI	-10 200 GAACAACTTCTTTCTGTAG	220 240 Accacttatcagaagaggattttaaa
слря * S.D. 260	280	I L T T S Q R 300	320	340 360
GCGTATTTTAGTTTTTCTGATTATGATCTGGAGGT A Y F S F S D Y D L E V 380	INQHRG 400	K V N K L G F 420	АІQLСL. 440	A R Y P G C S L S> 460 480
ANTIGECCEATTAAATCAACCAGACTAACTTCTTA N N P I K S T R L T S Y 500 TTAGAAGTATTCAACTATCATTCGGTAGTGC	VSRQLH	LDAIDLN 540	SYDHRN 560	TRANHFNEI> 580 600
TTAGAAGTATTCAACTATCATCGATTGGTAGTGC L E V F N Y H R F G S A 620 GATTTCTTAACTCGAAAAAGAATTATTTTTCCATC	. N T Q K Q L 640	IEYLIEL 660	A L E N D D 680	SIYLМККТ I> 700 720
D F L T R K R I I F P S 740 GATATACAAATTGAAAACTAGAGAGTTTGTTTCA	1 A T L E D 760	I I S R C R D 760	KAENNL 800	FSILLCSLT> 820 840
рідіек L E S L F Q 860) I Y E E T K 880	I T K L A W L 900	К D I P G К. 920	ANPESFMSI> 940 960
TGTAAAAAGTGGAAGTGATTGCTTCCATGGGACT C K K V E V I A S H C I 980 TCCCGTTTTGAGCTTGAAAAAAGATACTCTTTACT	. G T I N V S 1000	HINRNRF 1020	LQLARL 1040	GENYDAYDF> 1060 1080
S R F E L E K R Y S L L 1100 AAAGGGACACGTGATTCACAAGAACAGTTAAAAGA	IAFLVN 1120	HHQYLID		DRILASIKR> 1180 1200
KGTRDSQEQLKE 1220	: К G К L А Т 1240	ККLЕНҮА 1260	S L I D A L 1280	HFAKDNDSN> 1300 1320
CCTTTTGACGAAATTGAACGAATCATGCCTTGGGA P F D E I E R I M P W E 1340	: D L V Q D G 1360	ЕЕАККА I 1380	т G N K N H 1400	GYLEMVRNK> 1420 1440
GCTAATTACCTCCGAAGATACACGCCAATGTTATT A N Y L R R Y T P M L L 1460	, R T L S F K 1480	АТРААNР 1500	V L M A L T 1520	QLTDLHNSG> 1540 1560
AAAAGAAAAAATACCGGCAGATACTTCTACTGATTT K R K I P A D T S T D F 1580 CTAAAGAACAATATTCCATCAGAGAGATATTTCAGT	ч s к к w к 1600	S L V R P E E 1620	G K I D R S 1640	YYELVAFTE> 1660 1680
	E G S M I H			I D S E T I P D T>
		1740	1760	1780 1800
TTTGAGGACTATTTAAAGGATCGGGAAATAATTTT F E D Y L K D R E I I I 1820	AGATTTACAGCTTCAATTT DLQLQF 1840	TATTCGACAGTTGATAAGAG Y S T V D K R 1860	AATTTCAAGAGCAAACCTTA ISRANL 1880	AAAAGTTGGAAAAAGTTACACCTAGC K K L E K V T P S> 1900 1920
TTTGAGGACTATTTAAAGGACGGGAATAATTTT F E D Y L K D R E I I I 1820 GAAGGGAAATATATAGAAAAAACTTTATTGAAT E A E I Y R K K L Y 5 I 1940	AGATITACAGCTTCAATTT D L Q L Q F 1840 AATTCCTAAGATAAGGCTT I P K I R L 1960	TATTCGACAGTTGATAAGAG Y S T V D K R 1860 Agggatcttttaattgaggt S D L L I E V 1980	AATTTCAAGAGCAAACCTTA I S R A N L 1880 5GACAGTTGGACCAACTTTT D S W T N F 2000	ANAGTTGGAANAAGTTACACCTAGC K K L E K V T P S> 1900 1920 CACAAGAATTTAGTCATGATTCTACA S Q E F S H D S T> 2020 2040
TTTGAGGACTATTTAAÄCGATCGGGAATAATTTT F E D Y L K D R E I I I 1820 GAAGGAGAATAATATAGAAAAACTTTATTCAAT E A E I Y R K K L Y S I 1940 GGGAAACCGCCGATGAACAAGAAAATTAT G K P P S E Q E R K I I 2060	AGATTTACAGCTTCAATTT , D L Q L Q F 1840 NATTCCTAAGATAAGGCTT. 1 P K I R L 1960 TTTTGCTGCTTTGCTGGGT : F A A L L G 2080	TATTCCACRGTTGATAAGAG Y S T V D K R 1860 AGTGARCTTTTAATTGAGGT S D L L I E V 1980 TTAGGGATGAATATTGGTCT L G M N I G L 2100	ATTTCAAGAGCAAACCTTA I S R A N L 1880 SGACATTGGGACCAACTTTT D S W T N F 2000 FGAAAAATGGCCCAATCAA E K M A Q S 2120	AAAAGTTGGAAAAAGTTACACCTAGC K K K V T P S 1900 1920 1920 CACAAGAATTAGTCATGATCTACA S C F S H D S T D 2020 2040 CTCCTGGAATTTCTTAGTCATCATCATGATTCTACGTTA T P G I S Y S Q D 2040 21240 2140 2160 2160 2160 2160 2160
ТТТСАGGACTATTTAAÄCGATCGGGAATAATTTT F E D Y L K D R E I I I 1820 GAAGCAGAATATATATAGAAAAACTTTATTCAT E A E I Y R K K L Y S GGGAAACCGCCGATGGACAAGAAAAATTAT G K P P S E Q E R K I I 2060 GCCAATGCCAACGAATGGCGCTTTATAAAAAGC A N A K Q W R F Y K E J 2180	AGATTACAGCTTCANTT: D L Q L Q F 1840 ANTICCTANGATANGGCTT. 1950 TTTTGCTGCTTTGCTGGGT F A A L L G 2080 TCTGGACTGGTGCTCANTCT: L T R A Q S 2200	TATTCGACAGTTGATAAGAG Y S T V D K R 1860 AGTGATATTGAGGT S D L L I E V 1990 TTAGGGATGAATATTGGTCT L G M N I G L 2100 GTTTTGGTTAATTATCAGTT V L V N Y Q L 2220	МТТТСАЛСЛАСАЛАССТТА I S R A N L 1880 3GACARGTTGGACCAACTTTT D S W T N F 2000 ГGAANAAATGGCCCAATCAA E K M A Q S 2120 MAGGTTCCTGTTGCAGACT K L P V A D 2240	$\begin{array}{rrrr} \textbf{AAAAGTTGGAAAAAGTTACACCTAGG} \\ \textbf{K} & \textbf{K} & \textbf{L} & \textbf{E} & \textbf{K} & \textbf{V} & \textbf{T} & \textbf{P} & \textbf{S} \\ \hline 1900 & 1920 \\ \textbf{CACCARGAATTAGTCATAGATCTACA \\ \textbf{S} & \textbf{Q} & \textbf{E} & \textbf{F} & \textbf{S} & \textbf{H} & \textbf{D} & \textbf{S} & \textbf{T} \\ \hline 2020 & 2040 \\ \textbf{CTCCTGGGAATTTCTTAGTCAGTTA \\ \textbf{T} & \textbf{P} & \textbf{G} & \textbf{I} & \textbf{S} & \textbf{Y} & \textbf{S} & \textbf{Q} & \textbf{L} \\ \hline 2140 & 2160 \\ \textbf{TTTGGGGGGTGAAGGAAAACCACTGCT \\ \textbf{F} & \textbf{K} & \textbf{G} & \textbf{E} & \textbf{K} & \textbf{T} & \textbf{A} \\ 2260 & 2280 \end{array}$
$\begin{array}{cccc} \texttt{TTGAGGACTATTTAAAGCATCAGGAAATAATTTT}\\ \texttt{F} \texttt{E} \texttt{D} \texttt{Y} \texttt{L} \texttt{K} \texttt{D} \texttt{R} \texttt{E} \texttt{I} \texttt{I} \texttt{I} \texttt{I}\\ \texttt{I} \texttt{S1} \texttt{C1}\\ \texttt{GAAGCAGAAATATATTATAGAAAAAACTTTATTCAAT\\ \texttt{E} \texttt{A} \texttt{E} \texttt{I} \texttt{Y} \texttt{R} \texttt{K} \texttt{K} \texttt{L} \texttt{Y} \texttt{S1}\\ \texttt{GGAAACCGCCGATGGAACAAGAAAAATTAT\\ \texttt{G} \texttt{G} \texttt{P} \texttt{P} \texttt{S} \texttt{E} \texttt{Q} \texttt{E} \texttt{R} \texttt{K} \texttt{I} \texttt{I}\\ \texttt{2060}\\ \texttt{GCCAATGCCAAACAATGGCCGTTTATAAAGAAGA\\ \texttt{A} \texttt{N} \texttt{K} \texttt{Q} \texttt{W} \texttt{R} \texttt{P} \texttt{Y} \texttt{K} \texttt{E} \texttt{J}\\ \texttt{2180}\\ \texttt{TCAGACGGATGCGCGTCCCCCGCTCCCAATGGGCGCTCTCACG\\ \texttt{S} \texttt{D} \texttt{G} \texttt{M} \texttt{R} \texttt{V} \texttt{P} \texttt{V} \texttt{G} \texttt{V} \texttt{S} \texttt{J}\\ \texttt{2300} \end{array}$	AGATTACAGCTTCANATT D L Q L Q F 1840 AATTCCTANGATANGGCTT. I P K I R L 1950 TTTTGCTGCTTTGCTGGGT C F A A L L G 2080 ICTCGACTCGTGCTCANTCT. L T R A Q S 2200 ICTCANATCCGATGTTAAT L K S D V N 2320	TATTCACAGTTCATARAGAG Y S T V D K R 1860 ACTCATCTTTTATTCAGTTAGGT S D L L I E V 1990 TTAIGGATGATATTIGGTCT L G M N I G L 2100 GTTTTGGTTATTATCAGTT V L V N Y Q L 2220 CCACATTACCAAAGTATGGA P H Y K S M E 2340	МТТТСАЛСЛЯСАЛАССТАК I S R A N L 1880 3GACENETIGGACCAACTITT D S W T N F 2000 IGAANAATGGCUCAATCAA E K M A Q S 2120 MAGCTTUCTGTTGCAGACT K L P V A D 2240 MAGGAGCTACAATGATTC K G A T M I 2360	$\begin{array}{rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr$
$\begin{array}{ccccc} TTTGAGGACTATTTAAAGCATCAGGGAAATAATTTTF E D Y L K D R E I I I I1820GAAGCAGAAATAATATATAGAAAAAACTTAATCAATE A E I Y R K K L Y S IGGGAAACCGCCGAGTGGAACAAGAAAAATTATG K P P S E Q E R K I I2060GCCAATGCCAAATGGCGCTTTATAAAGAAGA N A K Q W R F Y K E J2180TCAGACGGAATGCGCGGTCCCAGTGGGGGCTCTCAGS D G M R V P V G V S J2300CATCATATCAAGTTGCTTCAATAAAAACGCJH H I E V A S T N T R E2420$	AGATTACAGCTTCANTT: D L Q L Q F 1840 MATTCCTANGATANGGCTT. I P K I R L 1950 TTTTGCTGCTTTGCTGGGT TTTTGCTGCTTTGCTGGGT 2080 ICTCGAGCTGGTCCANTCT L T R A Q S 2200 ICTCANATCGATGTTAAT L K S D V N 2320 MCGTACTCATACCCTTCAT C A T H T L D 2440	TATTCACAGTTCATAGAGG Y S T V D K R 1860 ACTCATCTTTTATTCATTCAGTT S D L L I E V 1990 TTAGGGATGATATTGGTCT L G M N I G L 2100 GTTTTGGTTATTATCAGTT V L V N Y Q L 2220 CCACATTACCAAAGTATGGA P H Y K S M E 2340 CCCCACATTACCAAAAGTATGAAAC G L L Y H E T 2460	MITICAAGAGCAAACCITA I S R A N L 1880 JGACAGITGGACCAACTITT D S W T N F 2000 IGAAAAAATGGGCCCAATCAA E K M A Q S 2120 MAGGTTCCTGTTGGAGACT K L P V A D 2240 MAGAGAGCTACAATGATCA K G A T M I 2360 AGAATGAAATGACGAAC D L D I E E 2480	$\begin{array}{cccc} \text{AAAAGTTGGAAAAAGTTACACCTAGG \\ \text{K} & \text{K} & \text{L} & \text{E} & \text{K} & \text{V} & \text{T} & \text{P} & \text{S>} \\ \hline 1900 & & & & & & & & & \\ 1900 & & & & & & & & & & \\ 1900 & & & & & & & & & & & \\ 1900 & & & & & & & & & & & \\ 1900 & & & & & & & & & & \\ 1900 & & & & & & & & & & \\ 1900 & & & & & & & & & & \\ 1900 & & & & & & & & & & \\ 2020 & & & & & & & & & \\ 2020 & & & & & & & & & \\ 2260 & & & & & & & & & & \\ 2380 & & & & & & & & & & \\ 2380 & & & & & & & & & & \\ 2380 & & & & & & & & & & \\ 2380 & & & & & & & & & & \\ 2380 & & & & & & & & & & \\ 2380 & & & & & & & & & & \\ 2380 & & & & & & & & & & \\ 2390 & & & & & & & & & \\ 2200 & & & & & & & & \\ 2200 & & & & & & & \\ 2200 & & & & & & & \\ 2200 & & & & & & & \\ \end{array}$
$\begin{array}{ccccc} \mbox{ttransc} \mbox{transc} \mbo$	$\begin{array}{cccccccccccccccccccccccccccccccccccc$	$\begin{array}{cccccccccccccccccccccccccccccccccccc$	MITICAAGAGCAAACCITA I S R A N L 1880 30ACASTIGGACCAACTITI D S W T N F 2000 100 C C C C C C AATCAA E K M A Q S 2120 C C C C C C C C C A C C C C C C C C C C	$\begin{array}{cccc} \text{AAAAGTTGGAAAAAGTTACACCTAGG \\ & K & L & E & V & T & P & S \\ \hline 1900 & 1920 \\ \hline CACAAGAATTAGTCATGATTCTACA \\ & S & Q & E & F & S & N & D & S & T \\ \hline 2020 & 2040 \\ \hline CTCCTGGAATTCTTATTCTCAGTTA \\ & T & G & I & S & Y & S & Q & L \\ \hline 2140 & 2160 \\ \hline TTGGGGTGAAGGAAAACCACTGCT \\ F & W & G & E & G & K & T & T & \\ \hline 2260 & 2280 \\ \hline 2260 & 2280 \\ \hline 2260 & 2400 \\ \hline CATCATAAATGATAGGAAAACCACTGCT \\ R & S & I & N & D & R & H & T & \\ \hline 2300 & R & H & R & T & \\ \hline 2300 & R & H & R & M & \\ \hline 2310 & T & N & G & Y & S \\ \hline CTTCCTAGTAGATACAAATCGATACGACTTATCT \\ H & F & T & D & T & N & G & Y & S \\ \hline CTTCCTAGTACGATACCACTTATCAGAAATGATACACTTATCAGAAATGATACACTTATCAGAAATGATACACTTATCAGAATCGATACTATCAGAAATGACTACTTATCAGAAA \\ P & S & Y & Y P & N & L & S \\ \hline 2620 & 2640 \\ \hline \end{array}$
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	$\begin{array}{cccccccccccccccccccccccccccccccccccc$	$\begin{array}{cccccccccccccccccccccattcachages of the test of the test of the test of te$	$ \begin{array}{ccccc} MATTECAAGAGCAAACCTTAI S R A N L \\ 1800 \\ 30ACASTTOGACCAACTTTT D S W T N F \\ 2000 \\ 1$	$\begin{array}{cccc} \text{AAAAGTTGGAAAAAGTTACACCTAGG \\ & K & L & E & K & V & T & P & S \\ \hline 1900 & 1920 \\ \hline caccaaGaAATTAGTCATGATCTACA \\ & S & G & E & F & S & H & D & S & T \\ \hline 2020 & 2040 \\ \hline crccTGGGAATTTCTTATTCTCAGTTA \\ & T & G & I & S & Y & S & L \\ \hline 2140 & 2160 \\ \hline TTGGGGTGAAGGAAAACCACTGCT \\ F & G & G & G & K & T & X \\ \hline 2260 & 2280 \\ \hline CatCataGAAGGAAAACCACTGCT \\ R & S & I & N & D & R & H & T & T \\ \hline 2260 & 2280 \\ \hline CatCatAGAAATCATAGGCATACCACTT \\ R & S & I & N & D & R & H & T & T \\ \hline 2390 & 2400 \\ \hline ATTTACAGTACACAACTGATTCT \\ H & F & T & D & T & N & G & Y & S \\ \hline 2200 & 2200 \\ \hline CTCCTGTACCCATACACAAATCCATTACTAACGACT \\ P & S & I & N & D & R & H & T & T \\ \hline 2390 & 2400 \\ \hline CTCCTTACTACCCACTACTTATTCT \\ H & F & T & D & T & N & G & Y & S \\ \hline 2500 & 2520 \\ \hline CTTCCTTACTACCCACTACTTATCAGAAA \\ P & S & Y & Y & P & N & L & S \\ \hline CTCCTTTACTATTAGGAAAGCTAGGC \\ \hline S & S & L & L & G & K & L & G \\ \hline 2740 & 2760 \\ \hline \end{array}$
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	$\begin{array}{cccccccccccccccccccccccccccccccccccc$	$\begin{array}{cccccccccccccccccccccccccccccccccccc$	$ \begin{array}{ccccc} \text{MATTECAAGAGCAAACCTTAI S R A N L \\ 1800 \\ \text{SGACAGATTGGACCAACTATTD S W T N F \\ 2000 \\ \text{FGAAAAAATGGGCCCAATCAAT } \\ \text{SGACAGTTCCTGTTGCAGACT } \\ \text{SGACAGCTTCCTGTTGCAGACT } \\ \text{K L P V A D } \\ \text{SGACAGCAATGATACT } \\ \text{K G A T M I } \\ \text{2360 } \\ \text{AAAGGAGCAATATTGAGGAAC } \\ \text{D L D I E E } \\ \text{2480 } \\ \text{AATCTAGATATTGAAGAATACTCA } \\ \text{L F S I K S } \\ \text{2600 } \\ \text{CAAAGGAAAAGTATCAAATCAAC } \\ \text{L F S I K S } \\ \text{2600 } \\ \text{CAAAGGAAAAGTATCAAATCAAC } \\ \text{L F S I K S } \\ \text{2600 } \\ \text{CAAAAGGAAAAGTATCTA } \\ \text{CAAACGAAAAGTATCTA } \\ \text{CAACGAAAAGTATCTA } \\ \text{CACGAAAGGAAAAGTAGTATCTA } \\ \text{CACGAACAGGAAAAGTATCTA } \\ \text{CACGAACAGGAAAAGTATCTA } \\ \text{CACGAACAGAAAAGTATCTA } \\ \text{CACGAACTATATTATCAGAATAGTA } \\ \text{CACGAACAGAAAAGTATCTA } \\ \text{CACGAACAGAAAAGTATCTA } \\ \text{CACGAACAGAAAAGTATCTA } \\ \text{CACGAACAGAAAAGTATCTA } \\ \text{CACGAACAGGAAAAGTATCTA } \\ \text{CACGAACGAAAAGTATCTA } \\ \\ \text{CACGAACGAACGAAAAGTATCTA } \\ \\ \ \text{CACGAACGAAAAGTATCTA } \\ \\ \ \text{CACGAACGAAAAGTAACTA } \\ \\ \ \text{CACGAACGAAAAGTAACTA } \\ \\ \ CACGAACTAACGAAAAGTA A \\ \\ \ \text{CACGAACGAAAAGTAACTAACTAACTAACTAACTAACTAA$	$\begin{array}{cccc} AAAAGTTGGAAAAAGTTACACCTACCR K L E R V T P S1900 1920CACAAGAATTAGTCATCATCTACAS Q E F S H D S T>2020 2040CTCCTGGGAATTTCTTATTCTCACTAT P G I S Y S Q L>2140 2160TTTGGGGTGAAGGAAAAACCACTGCTF W G E C K T T A>2260 2280CATCATAATGATAGTAGCCATACCACTR S I N D R H T T>2380 2400CATCATAATGATAGCATACCACTCATR S I N D R H T T>2380 25002500 2500CTTCTGATCATACACACACTCACTACTACACACTR S Y Y P N L S E>2600 2500CATCATACTATAGGAACACACTGCTB S Y Y P N L S E>2600 2500CATCATACTATAGGAACACACTGCTS S L L L G K L G2740 2760CATCACGGCGAAGGATCACTCATGCAE L R R R I T H C$
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	$\begin{array}{cccccccccccccccccccccccccccccccccccc$	$\begin{array}{c} {\rm TrattcGacaGattGathAGAG}\\ {\rm Y} & {\rm S} & {\rm T} & {\rm V} & {\rm D} & {\rm K} & {\rm R} \\ {\rm 1660}\\ {\rm ActGatctTtTTAATGAGGT}\\ {\rm S} & {\rm D} & {\rm L} & {\rm I} & {\rm E} & {\rm V} \\ {\rm 1980}\\ {\rm TTAGGGATGAATATGGTCT}\\ {\rm L} & {\rm G} & {\rm M} & {\rm I} & {\rm G} & {\rm L} \\ {\rm C} & {\rm I} & {\rm N} & {\rm I} & {\rm G} & {\rm L} \\ {\rm C} & {\rm C} & {\rm N} & {\rm I} & {\rm G} & {\rm L} \\ {\rm C} & {\rm C} & {\rm N} & {\rm I} & {\rm G} & {\rm L} \\ {\rm C} & {\rm C} & {\rm N} & {\rm I} & {\rm G} & {\rm L} \\ {\rm C} & {\rm C} & {\rm N} & {\rm V} & {\rm V} & {\rm V} & {\rm V} \\ {\rm C} & {\rm C} & {\rm V} & {\rm N} & {\rm Y} & {\rm Q} & {\rm L} \\ {\rm C} & {\rm C} & {\rm C} & {\rm C} \\ {\rm C} & {\rm C} & {\rm C} & {\rm C} \\ {\rm C} & {\rm C} & {\rm C} & {\rm N} & {\rm Y} & {\rm N} & {\rm Y} & {\rm Q} \\ {\rm C} & {\rm C} & {\rm C} & {\rm C} \\ {\rm C} & {\rm C} & {\rm C} & {\rm C} \\ {\rm C} & {\rm C} & {\rm C} & {\rm C} \\ {\rm C} & {\rm C} & {\rm C} & {\rm C} \\ {\rm C} & {\rm C} & {\rm C} & {\rm C} \\ {\rm C} & {\rm C} & {\rm C} & {\rm C} \\ {\rm C} & {\rm C} & {\rm C} & {\rm C} \\ {\rm C} & {\rm C} & {\rm C} & {\rm C} \\ {\rm C} & {\rm C} & {\rm C} & {\rm C} \\ {\rm C} & {\rm C} & {\rm C} & {\rm C} \\ {\rm C} & {\rm C} & {\rm C} \\ {\rm C} & {\rm C} & {\rm C} \\ {\rm C} & {\rm C} & {\rm C} \\ {\rm C} & {\rm C} & {\rm C} \\ {\rm C} & {\rm C} \\ {\rm C} & {\rm C} & {\rm C} \\ {\rm C} \\ {\rm C} & {\rm C} \\ {\rm C} & {\rm C} \\ {\rm C} \\ {\rm C} & {\rm C} \\ {\rm C} \\ {\rm C} & {\rm C} \\ {\rm C$	$ \begin{array}{ccccc} \text{MATTCAACAGCAAACCTTAI S R A N L \\ 1800 \\ \text{SGACAGTTGGACCAACTTTT } \\ 2000 \\ \text{FGAAAAAATGGCCCAATCAAT } \\ \text{SGACAGTTCGTTCGTTGCAGACT } \\ \text{SGACAGTTCGTTGCAGACT } \\ \text{SGACAGTTCGTTGCAGACT } \\ \text{SGACAGCAACAATGATCC } \\ \text{SGACAGACCTACAATGATCC } \\ \text{SGACTACAATGAATACACL } \\ \text{SGACTACAAATGAATCACACL } \\ \text{SGACTACAAATGAATCACACL } \\ \text{SGACTACAAATACACAATGATCC } \\ \text{SGACTATACACAAATACACC } \\ \text{SGACTACACACACAATGATCC } \\ \text{SGACTACACACACACACACTCC } \\ \text{SGACTATCCGCCACAACTCC } \\ \text{SGACTATCCGCCACACCTCC } \\ \text{SGACTATCCGCCACACCTCC } \\ \text{SGACTACTCCCCACACCTCC } \\ \text{SGACTACTCGCCACACCTCC } \\ \text{SGACTACTCGCCACACCTCC } \\ \text{SGACTACTCGCCACACCTCC } \\ \text{SGACTACTCGCCCACACCTCC } \\ \\ \text{SGACTACTCGCCCACACCTCC } \\ \text{SGACTACTCGCCCACACCTCC } \\ \\ \text{SGACTACTCGCCCACCCCACACCTCC } \\ \\ \text{SGACTACTCGCCCACACCTCC } \\ \\ \text{SGACTACTCGCCCACACCTCC } \\ \\ \text{SGACTACTCGCCCACACCTCC } \\ \\ \ \\ \text{SGACTACTCGCCCACACCTCC } \\ \\ \\ \ \\ \ \\ \ \\ \ \\ \ \\ \ \\ \ \\ \ \\$	$\begin{array}{ccccc} AAAAGTTGGAAAAAGTTACACCTACCK K L E K V T P S1900 1920CACAAGAATTAGTACATCATACTTCACAS Q E F S H D S T>2020 2040CTCCTGGGAATTTCTTACTACATTAT P G I S Y S Q L>2140 2160TTTGGGGGGAAGGAAAAACCATGGTTT G G L S Y S Q L>2140 2160CATCATAAATGATAGCATACTGGTF W G E G K T T A>2260 2280 2280CATCATAAATGATAGCATACGATCATR S I N D R H T T>2380 2400CATCATAAATGATAGCATACGATCATTCTH F T D T N G Y S>2500 2500CTTCCTACTACACTACATCAGACP S Y Y P N L S E>2602 260CTTCCTACCTAACTTATCAGAAP S Y Y P N L S E>2602 260CATCATACTATAGGAACACCATGCGS S L L L G K L G2740 ACTACGGCGAAGGATCACTCATGCAE L R R I T H C2460 R L N V D2460 R L N V D2480 R L N V D2480 3000$
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	$\begin{array}{cccccccccccccccccccccccccccccccccccc$	$\begin{array}{cccc} \texttt{rtattcackactattakacac} \\ \texttt{rtatcackactattakacac} \\ \texttt{s} & \texttt{s} & \texttt{v} & \texttt{v} & \texttt{k} & \texttt{k} \\ \texttt{ls60} \\ \texttt{actaatcttttaattaacat} \\ \texttt{s} & \texttt{b} & \texttt{l} & \texttt{l} & \texttt{l} & \texttt{v} \\ \texttt{ls70} & \texttt{l} & \texttt{l} & \texttt{l} & \texttt{v} \\ \texttt{ls70} & \texttt{l} & \texttt{l} & \texttt{l} & \texttt{v} \\ \texttt{ls70} & \texttt{l} & \texttt{l} & \texttt{l} & \texttt{v} \\ \texttt{ls70} & \texttt{l} & \texttt{l} & \texttt{l} & \texttt{v} \\ \texttt{ls70} & \texttt{l} & \texttt{l} & \texttt{l} & \texttt{v} \\ \texttt{ls70} & \texttt{cactattattactttact} \\ \texttt{l} & \texttt{cactattattact} \\ \texttt{l} & \texttt{l} & \texttt{n} & \texttt{l} & \texttt{s} \\ \texttt{l} & \texttt{l} & \texttt{v} \\ \texttt{l} & \texttt{l} \\ l$	$ \begin{array}{ccccc} \text{MitticAacagcallectrate} \\ \text{I} & \text{S} & \text{R} & \text{A} & \text{N} & \text{L} \\ \text{I} & \text{S} & \text{R} & \text{A} & \text{N} & \text{L} \\ \text{I} & \text{S} & \text{R} & \text{A} & \text{N} & \text{L} \\ \text{I} & \text{S} & \text{N} & \text{I} & \text{N} & \text{F} \\ \hline & \text{2000} \\ \text{Isocatation} \\ \text{Isocatation} \\ Constant of the second s$	$\begin{array}{ccccc} \text{AAAAGTTGGAAAAAGTTACACCTAGG \\ & k & L & E & k & V & T & P & S \\ \hline 1900 & 1920 \\ \hline CACAAGAATTAGTCATGATCTACA \\ & S & Q & E & F & H & D & S & T \\ \hline 2020 & 2040 \\ \hline CTCCTGGAATTTCTTATTCTCAGTA \\ & T & G & I & S & Y & S & Q & L \\ \hline 2140 & 2160 \\ \hline TTGGGGTGAAGGAAAACCACTGCT \\ F & G & E & G & K & T & A \\ \hline 2260 & 2280 \\ \hline CTCCTGCTAGTAAATGATAGGCATACCACTT \\ R & S & I & N & D & R & H & T & T \\ \hline 2300 & CTCCTACTACTAATGCATCACTATTCT \\ H & F & T & D & T & N & G & Y \\ CATCATAAATGATAGCACACTATCACAATTCC \\ H & F & T & D & T & N & G & Y \\ CATCATAAATGATAGCACACTATCAGAAA \\ F & S & Y & Y & P & N & L & S \\ CTCCTACTACTACCACTAATGCACACTATCAGAAA \\ F & S & Y & Y & P & N & L & S \\ CTCCTACTACGGCGAGGAAAACCACTCATGGAAAAGCACTAGCACACTATGGAAAGCACTCATGCACAGGCAGAGAATCACTCATGGAAAGCTAGGCAGAGAAGCACCTCATGGAAAGCACAGCACACTCATGGAAAGCACACGCATAGTGATAGGAAGCAATGCAATGGAAAGCAATGCAATGGAAAGCAATGCAATGGAAAGCAATGCAATGGAAAGCAATGCAATGGAAAGCAATGCAATGGAAAGCAATGGACAT \\ H & V & S & P & I & N & W & E \\ 3100 & 3120 \\ \hline \end{array}$
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	$\begin{array}{cccccccccccccccccccccccccccccccccccc$	$\begin{array}{cccccc} {\rm ratic cack cort cat had a cack of the case of the c$	$\begin{array}{cccc} \text{ANTTECAAGAGCAAACCTTAI S R A N L \\ 1880 \\ 1880 \\ \text{SGACAGTTGGACCAACTTTT } \\ 0 S W T N F \\ 2000 \\ \text{ISGACAGTTGGACCAACTTT } \\ 2000 \\ \text{ISGACAGTTGGGCCAATCAA } \\ \text{SGACATCAGACAATGATCA } \\ \text{SGACAGCAATGATCAATGATCA } \\ \text{K G A T M I } \\ 2360 \\ \text{AAGGGACTACAATGATACAC } \\ \text{AAGGACATATGAGCAACAATGATCA } \\ \text{AAGGACATATGAGCAATGATCA } \\ \text{AAGGACATATGAGAAATGATCA } \\ \text{AAGGACATATGAGAAATGAACAC } \\ \text{L D I E E } \\ 2480 \\ \text{AATGTAGATATGAGAAAGAATATCAA } \\ \text{CAATGTAGATATTGAGAAACAATGATCA } \\ \text{CAATGTAGAAAAGAAAGAAATATCAA } \\ \text{CAATGTAGAAAAGAAAGAAATATCAA } \\ \text{CAATGTAGAAAAGAAAGAAATATCAA } \\ \text{CAATGTAGAAAAGAAAGAAAGAATATCA } \\ \text{CAATATATTATACAGAAAGAAAGAACTATCA } \\ \text{CAATATATATTACAGAATAGTG } \\ \text{CAATATATATATACAGAATAGTG } \\ \text{CAATATATATACAGAATAGTA } \\ CAAAAAGAAAAAAGAAATATGAAGAA $	$\begin{array}{c} \text{AAAAGTTGGAAAAAGTTACACCTAGC } \\ K K L E K V T P S \\ 1900 1920 \\ \text{CACAAGAATTAGTACATCTACA \\ S G E F S H D S T \\ 2020 2040 \\ \text{CTCCTGGAATTTCTTATTCTCACTA \\ T F G I S Y S Q L \\ 2140 2040 \\ \text{CTCCTGGAATTTCTTATTCTCACTA \\ F G I S Y S Q L \\ 2140 2160 \\ \text{CTCCTGGAATTTCTTATTCTCACTA \\ 2260 2280 2280 \\ 2260 2280 \\ 2280 240 \\ \text{CTCCTACAATGATAGTAGCACTACCACT \\ R S I N D R H T T \\ 2380 2400 2400 \\ \text{CTCCTGACTACACACACTTCT \\ R S I N D R H T T \\ 2380 2400 2400 \\ 2500 2500 2500 \\ 2500 2500 \\ 2500 2500$
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	AGATTACAGCTTCANATT: , D Q L Q F 1840 MATTCCTARGATAAGGCTT. : I P K I R L 1960 TTTTGCTGCTTTGCTGGGT: : F A A L L G 2080 TCTGACTCGTGCTCATTGCTGGGT : F A A L L G 2080 TCTGACTGTGGTCGTCATTCT , L K S D V N 2320 NGCTACTCATACCGTGTCAT : A T H T L D 2440 CGCAGGACACTGATGCTCAT : A T H T L D 2440 CGCAGGACACTGATGCTCGCATC. : D F E P R I 2560 CGAGGACACTGGTCGCATT. : R N Y D E I 2680 CGAGGACACTAGGTCGCCATT. : R N Y D E I 2680 CGAGGACACTAGGTCGCCATT. : R N Y D E I 2680 CGAGGACTATTTTTTGCGCGCA R E L F F G R 2920 CCACAGCACTATATATTTTTCGCCCATTA. : A Y N Y 3040 CACTATTCCTAACACTTACATTAC	TATTCACACAGTTCATAGAGA TATTCACACAGATCATTGACAGA Y S T V D K R 1660 AGTCATACTTTTAATTGAGGT S D L I E V 1980 TTAGGGATGAATATTGGTCT L G M N I G L 22100 GTTTIGGTTAATTATCAGTT. V L V N Y Q L 2220 CACACTTACAAAAGTATGAC G L U Y H E T 2340 CCCCGACTTTACACAAGTATCACA R N I K K S Q 2380 CACAGAATTTAATCACAAACAT R N I K K S Q 2380 CCCCGAAAATTTAATCACAACA R N I K K S Q 2380 CCCCGAAAATTTAATCACACA R N I K K S Q 2380 CCCCGAAAATTTAATCACACA R S I F M I 2400 CCCCGAAAATTTAATCACACCCAA R C K F M E R 2400 CCCCGAAAATTTAATCACACCCAA CCCCGAAAATTTAATCACACCCAA L V K I D P E 3060 ACAGAATTGAATATAAAAAAA	$ \begin{array}{ccccc} \text{AMTTECAAGAGCAAACCTTAI S R A N L \\ 1800 \\ \text{SGACAGTTGGACCAACTTTT } \\ 1800 \\ \text{SGACAGTTGGACCAACTTTT } \\ 2000 \\ \text{FGAAAAATGGCCCAATCAA } \\ \text{SGACATTCGGTTCGTTGCAGACT } \\ \text{AAGGAGCTACAATGATTCC } \\ \text{K G A T M I } \\ 22400 \\ \text{CAATGTAGATATTGACGAAC } \\ \text{CAATGTAGATATTGACGAACATTCTA } \\ \text{CAATGTAGATATTACAGATAGTAC } \\ \text{CAATGTAGATATTACAGAAAGGAACGAACAGCAAATGATCTA } \\ \text{CAATGTAGACAGGAAAAGTATCTA } \\ \text{CAATGTAGACAGGAAAAGTATCTA } \\ \text{CAATGTAGACAGGAAAAGTATCTA } \\ \text{CAATGTAGACAGGAAAAGTATCTA } \\ \text{CAATATTGACGCAACAGAAAGGAACATTC } \\ \text{CAATATTGAGCACAGAAAAGTATCAACTC } \\ \text{D I R R Q L } \\ 2960 \\ \text{CAATATAGGCCTTGAAACATT } \\ \text{ACGGCCTTGAAACATT } \\ \text{CAATATGGCCTTGAAACATT } \\ \text{CAATATAGGCCTTGAAACATT } \\ \text{CAATATAGGCCTTGAAAATAGCATT } \\ \text{CAATATAGGCCTTGAAACATT } \\ \\ \text{CAATATAGGCCTTGAAACATT A \\ \\ \text{CAATATAGGCCTTGAAACATT A \\ \\ \text{CAATATAGGCCTTGAAATAATAGCT } \\ \\ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \$	AAAAGTTGGAAAAAGTTACACCTÀGC AAAAGTTGGAAAAAGTTACACCTÀGC 1900 1920 1900 1920 CACAAGAATTAGATCTACATCTACATCTACATTAGAATTAGAAATTAGTACATATCTACATCAATGATAGCAATGAAAGAAA
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	$x_{0}ATTTACAGCTTCAAATTT, $	TATTCACACAGTTCATAAGAG Y S T V D K R 1860 AGTCATATTTAATTGAGGT S D L L I E V 1980 TTAGGGATGAATATTGGTCT L G M N I G L 2100 GTTTTGGTTAATTATCAGTT V L V N Y Q L 2220 CCACATTAACAAAGTATGGA P H Y K S M E 2340 CCCCACATTACACAAGTATGAAACA G L L Y H E T 2460 AGAAATTATAAAAAGTATGAAACA R N I K K S Q 2380 AAACGAATCGCCTATTGATCAAACA R N I K K S Q 2380 AAACGAATGCCTATTGATCAAACA R N I K K S Q 2380 AAACGAATGCCTATTGATCAAACA R N I K K S Q 2380 CACAAGAACGCATTTTATGGAGGGG R C K F M E R 2940 CCCGGAAAATTATAGATCCCGA L V K I D P E 3060 AGAGAATTGAATATAAAAAA R B L N I K N 3180	$ \begin{array}{ccccc} Antical carge all controls of the second sec$	AAAAGTTGGAAAAAGTTACACCTÀGC AAAAGTTGGAAAAATTACACCTÀGC 1900 1920 1900 1920 CACAAGAATTAGATCAATCTACA S G E F S H D S T 2020 2040 CTCCTGGGATTTCTTATTGTCAGTA T F G I S Y S Q L> 2140 2160 TTGGGGATGGAGGAAAACCACTGCT F W G E G K T T A> 2260 22800 CATCATAGTAATGATACGACT ATTTAGCGATAACCACTGCT R S I N D R H T TD 2300 2400 CATCATACATACAATACGATTACACACT P S Y Y P N L S E CTOCTACTACCCTAACTTATCAGAA CATCATACACACCGGCAGAGAAACCTACTCATGGA S S Y Y P N L S E 2400 CTCTCTACTACCCTAACTTATCAGAA CTTCTACTACTACTACTACTACTACATAGAA S S Y Y P N L S E 2400 CTCTCTACTACTACTACTACTATCAGAA S L L G K L G 2740 2740 S A L N V L 2900 3100 3100 3120 3120

FIG. 1. The nucleotide sequence of tnpA (ORF5) of Tn917. The underlined regions correspond to the potential promoter, Shine Dalgarno (S. D.) site, and the inverted repeats that may represent a transcription terminator. The inferred amino acids of TnpA are shown in single-letter designations and are encoded from basepairs 160 to 3078. The designation " $tnpR^*$ " indicates the 3' terminus of tnpR, which is 16 bp

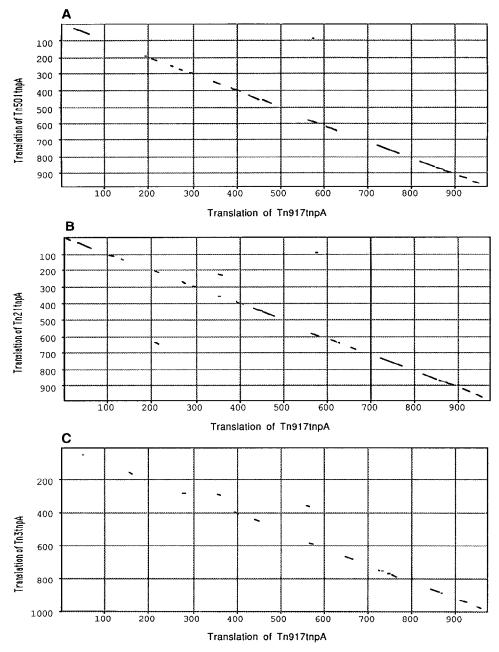


FIG. 2. Homology matrix plots comparing of tnpA of Tn917 with Tn501 (A), Tn21 (B), and Tn3 (C). A MacVector (IBI) program was used to generate the comparisons. Each point reflects at least 50% homology within a 14-amino acid alignment with the corresponding protein.

upstream of the start site of *tnpA*. The dashed line designates the right-end 38-bp inverted repeat. Sequencing was performed utilizing a pBluescript (Stratagene, Inc.) clone (pAM3917R) of a segment from the pAD1::Tn917 derivative pAM938 (Ehrenfeld and Clewell, 1987). The clone contained the 3' half of the transposon and adjacent pAD1 DNA. Overlapping nested deletions [*ExoIII*/mung bean nuclease protocol (Stratagene, Inc.)] generated from both orientations were sequenced using the double-strand DNA procedure as described by Ausubel *et al.* (1987). A Sequenase kit from U.S. Biochemical Corp. was utilized. In some cases, specific synthetic primers made at the University of Michigan Biomedical Research Core Facility were used.

17%, respectively. Homologies based on similarity are 53, 50, and 30%, respectively.

The similarity of Tn917 with other Tn3 family transposons, with respect to both *tnpR* and *tnpA*, is consistent with a close evolutionary relationship. In this regard it is noteworthy that Tn917, which originated in the gram-positive *E. faecalis*, has been shown to transpose in a gram-negative (*Escherichia coli*) background (Kuramitsu and Casadaban, 1986).

ACKNOWLEDGMENTS

We thank N. Brown, J. Mahillon, and E. Murphy for pointing out the homologies between the ORF5 and ORF6 reading frames and tnpA of other Tn3-family elements, and N. Brown and D. Rouch for information on the likely position of certain frameshifts. This work was supported by USPHS Grants GM33956, AI10318, and DE02731.

REFERENCES

- AUSUBEL, F. M., BRENT, R., KINGSTON, R. E., MOORE, D. D., SEIDMAN, J. G., SMITH, J. A., AND STRUHL, K. (1987). "Current Protocols in Molecular Biology." Wiley, New York.
- BROWN, N. L., WINNIE, J. N., FRITZINGER, D., AND PRIDMORE, R. D. (1985). The nucleotide sequence of the *tnpA* gene completes the sequence of the *Pseudomonas* transposon Tn501. Nucleic Acids Res. 13, 5657–5669.
- EHRENFELD, E. E., AND CLEWELL, D. B. (1987). Transfer functions of the *Streptococcus faecalis* plasmid pAD1: Organization of plasmid DNA encoding response to sex pheromone. *J. Bacteriol.* **169**, 3473–3481.
- HEFFRON, F. (1983). Tn3 and its relatives. In "Mobile Genetic Elements" (J. A. Shapiro, Ed.), pp. 223–260. Academic Press, New York.

- HEFFRON, F., MCCARTHY, B. J., OHTSUBO, H., AND OHTSUBO, E. (1979). DNA sequence analysis of the transposon Tn3. Three genes and three sites involved in transposition of Tn3. Cell 18, 1153-1164.
- KHAN, S. A., AND NOVICK, R. P. (1980). Terminal nucleotide sequence of Tn551, a transposon specifying resistance in *Staphylococcus aureus*: Homology with Tn3. *Plasmid* 4, 148–154.
- KURAMITSU, H., AND CASADABAN, M. (1986). Transposition of the gram-positive transposon Tn917 in *Esche* richia coli. J. Bacteriol. **167**, 711–712.
- MURPHY, E. (1989). Transposable elements in grampositive bacteria. *In* "Mobile DNA" (D. E. Berg and M. M. Howe, Eds.), pp. 269–288. Am. Soc. Microbiol., Washington, DC.
- PERKINS, J., AND YOUNGMAN, P. J. (1984). A physical and functional analysis of Tn917; a Streptococcus transposon in the Tn3 family that functions in Bacillus. Plasmid 12, 119–138.
- SHAW, J., AND CLEWELL, D. B. (1985). Complete nucleotide sequence of macrolide-lincosamide-streptogramin B-resistance transposon Tn917 in Streptococcus faecalis. J. Bacteriol. 164, 782–796.
- SHERRATT, D. (1989). Tn3 and related transposable elements: Site-specific recombination and transposition. *In* "Mobile DNA" (D. Berg and M. M. Howe, Eds.), pp. 163–184. Am. Soc. Microbiol., Washington, DC.
- TINOCO, I., BORER, P., DENGLER, B., LEVINE, M., UH-LENBECK, O., CROTHERS, D., AND GRALLA, J. (1973). Improved estimation of secondary structure in ribonucleic acids. *Nature (London). New Biol.* 246, 40–41.
- TOMICH, P., AN, F., AND CLEWELL, D. B. (1979). A transposon (Tn917) in *Streptococcus faecalis* that exhibits enhanced transposition during induction of drug resistance. *Cold Spring Harbor Symp. Quant. Biol.* **43**, 1217–1221.
- TOMICH, P., AN, F., AND CLEWELL, D. B. (1980). Properties of erythromycin-inducible Tn917 in *Streptococcus faecalis. J. Bacteriol.* **141**, 1366–1374.
- WARD, E., AND GRINSTED, J. (1987). The nucleotide sequence of the *tnpA* gene of Tn21. Nucleic Acids Res. 15, 1799–1806.

Communicated by Francis L. Macrina