JOURNAL OF BACTERIOLOGY, Apr. 1991, p. 2724–2728
0021-9193/91/082724-05$02.00/0

Vol. 173, No. 8

Nucleotide Sequence of *xylE* from the TOL pDK1 Plasmid and Structural Comparison with Isofunctional Catechol-2,3-Dioxygenase Genes from TOL pWW0 and NAH7

ROBERT C. BENJAMIN, JOHN A. VOSS, AND DANIEL A. KUNZ*

Department of Biological Sciences, University of North Texas, Denton, Texas 76203

Received 14 September 1990/Accepted 10 January 1991

Detailed restriction and nucleotide sequence analysis of the *Pseudomonas putida* TOL plasmid pDK1 *xylE* gene revealed significant homology with isofunctional *xylE* (81.5%) and *nahH* (78.0%) genes from the TOL pWW0 and NAH7 plasmids. The highest degrees of nucleotide and apparent amino acid conservation (82.2 and 86.4%, respectively) among all three genes were found to exist within a region comprising 264 nucleotides encoding the C terminus. A comparison of localized regions revealed significantly greater homology between $xylE_{pWW0}$ and $xylE_{pDK1}$ within the C-terminal region, whereas $xylE_{pWW0}$ and *nahH* showed greater similarity at the N terminus. The possibility that $xylE_{pWW0}$ may represent a genetic hybrid of $xylE_{pDK1}$ and *nahH* is discussed.

TOL plasmids, which encode for the degradation of toluene and related aromatic hydrocarbons, have been the subject of much recent intensive study. The genes in TOL plasmids are arranged in two operons, one encoding enzymes for initial hydrocarbon oxidation (the upper operon) and the other specifying enzymes for aromatic ring cleavage and the *meta* fission pathway (the lower operon) (for reviews, see references 1, 6, and 22). Despite extensive mapping studies aimed at resolving the genetic fine structure of TOL-encoded operons, little information on the nucleotide sequence of relevant genes has yet to emerge. The one exception to this is *xylE* from TOL plasmid pWW0 ($xylE_{pWW0}$), which encodes the aromatic-ring-cleavage enzyme, catechol 2,3-dioxygenase (C23O) (metapyrocatechase) (EC 1.13.11.2) (13). The *xylE* gene product is a 307-amino-acid polypeptide that shares 74% sequence homology with its isofunctional counterpart *nahH* from plasmid NAH7, conferring naphthalene degradation (4, 5). The two genes are 80% homologous in nucleotide sequence, and therefore the possibility that they evolved from a common ancestor has been proposed (5). A comparison of isofunctional *xylE* genes from separate TOL plasmids has, however, not been undertaken despite the fact that some TOL plasmids may encode for more than a single C23O (2, 10, 14). The present work describes the nucleotide sequence of *xylE* from the TOL plasmid pDK1, which is geographically remote in origin from that of TOL pWW0. In addition, its structure is compared with that of $xylE_{pWW0}$ and *nahH*. A preliminary account of this work appeared previously (20).

The TOL plasmid pDK1 was originally isolated from *Pseudomonas putida* HS1 in Minnesota by Kunz and Chapman (11). The catabolic genes were subsequently subcloned by Shaw and Williams (18) into plasmid RP1 and shown to exist in two separate operons similar to the situation in TOL pWW0 (22). We acquired the RP1-TOL pDK1 recombinant, designated pDKR1, from the above investigators and subjected it to further subcloning procedures to isolate the *xylL-E* region of the lower operon. pBR325 and pUC19 cloning vectors and pDKR1 DNA from *P. putida* PaW630 (Trp^- Str^r) were prepared by established methods (8, 19). DNA from an *Eco*RI digest of pDKR1 was subcloned into pBR325, generating plasmid pBK188 containing a 5.1-kb insert spanning the *xylL-E* region. This plasmid was then digested with *Xho*I, and DNA was further subcloned into the *Sal*I cloning site of pUC19 to generate pBK189 carrying the *xylE* gene on a 2.2-kb fragment. Cells carrying the latter plasmid expressed C23O activity as evidenced by the formation of yellow ring-fission product (2-hydroxymuconic semialdehyde) when suspended in catechol-containing (1 mM) phosphate buffer (pH 7.0). The 2.2-kb *Xho*I fragment from pBK189 was isolated and digested with *Taq*I, *Bst*EII, *Rsa*I, *Sal*I, or multiple digests thereof, and individual fragments were purified before being subjected to Maxam and Gilbert (12) or M13 (16, 17, 23) sequencing techniques. The $xylE_{pDK1}$ gene was identified by comparing preliminary sequence data with published sequences for $xylE_{pWW0}$ and *nahH* (4, 5, 13). Figure 1 shows a detailed restriction map of the $xylE_{pDK1}$ region and the sequencing strategy used.

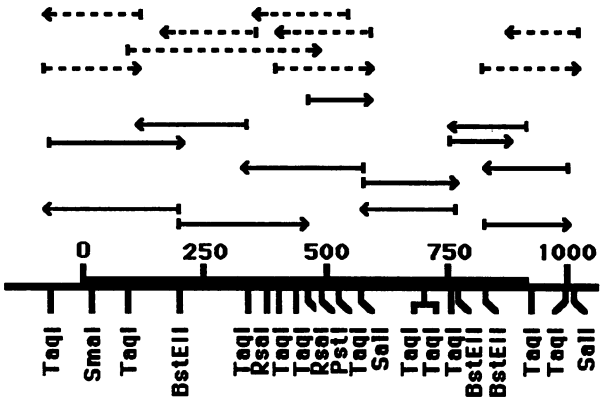

FIG. 1. Restriction map and DNA sequencing strategy used for the TOL plasmid pDK1 *xylE* gene. Solid and broken arrows represent DNA sequenced by the methods of Maxam and Gilbert (12) and Sanger et al. (16), respectively. The location of $xylE_{pDK1}$ is shown superimposed on the bottom line (thick line), and relative distances in base pairs are indicated.

* Corresponding author.

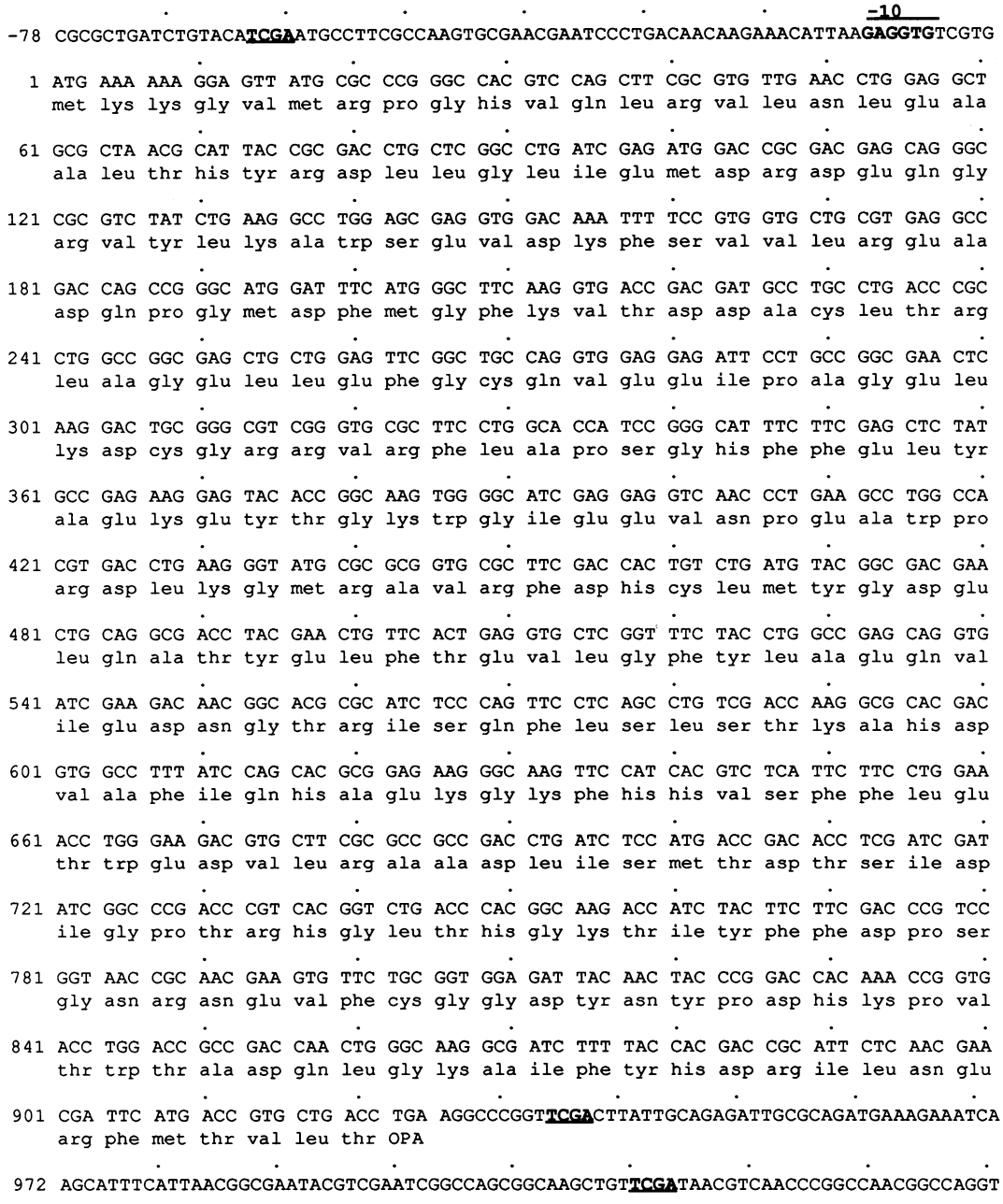

FIG. 2. Nucleotide sequence and deduced amino acid sequence of the *xylE* gene from the TOL plasmid pDK1. The sequence shown includes the 921-bp open reading frame and flanking regions. The putative ribosome binding site at −10 is shown in boldface type. The *Taq*I sites immediately flanking the gene (see Fig. 1) are underlined.

A comparison of restriction sites for *Sma*I, *Bst*EII, *Sal*I, and *Pst*I with those published for other C23O genes (2, 10) indicates that $xylE_{pDK1}$ encodes a class Ib type C23O. It should be noted, however, that two *Bst*EII sites were detected 50 bp apart at the 3′ end of the gene (Fig. 1) instead of the single site reported for other class Ib C23Os. The nucleotide and predicted amino acid sequences of $xylE_{pDK1}$ are illustrated in Fig. 2. As expected, the gene was found to have an identical size and a similar restriction pattern, G+C content (55%), and nucleotide sequence as $xylE_{pWW0}$ and *nahH* had. A comparison of the sequence with those of $xylE_{pWW0}$ and *nahH* (Fig. 3 and Table 1) shows that 72% of the nucleotides and 78% of the amino acids have been conserved for all three genes. A total of 258 variable nucleotides, affecting 199 codons, were found in the three genes. Most of the changes are silent (131 of 199) or result in neutral amino acid substitutions (eight are glutamate-to-aspartate switches and eight involve valine-isoleucine-leucine exchanges), a phenomenon not unlike that observed for other isofunctional bacterial genes (3, 9). The DNA homology for any pair of genes ranges from 78 to 82%. Comparisons made within arbitrarily subdivided localized regions (I to VI)

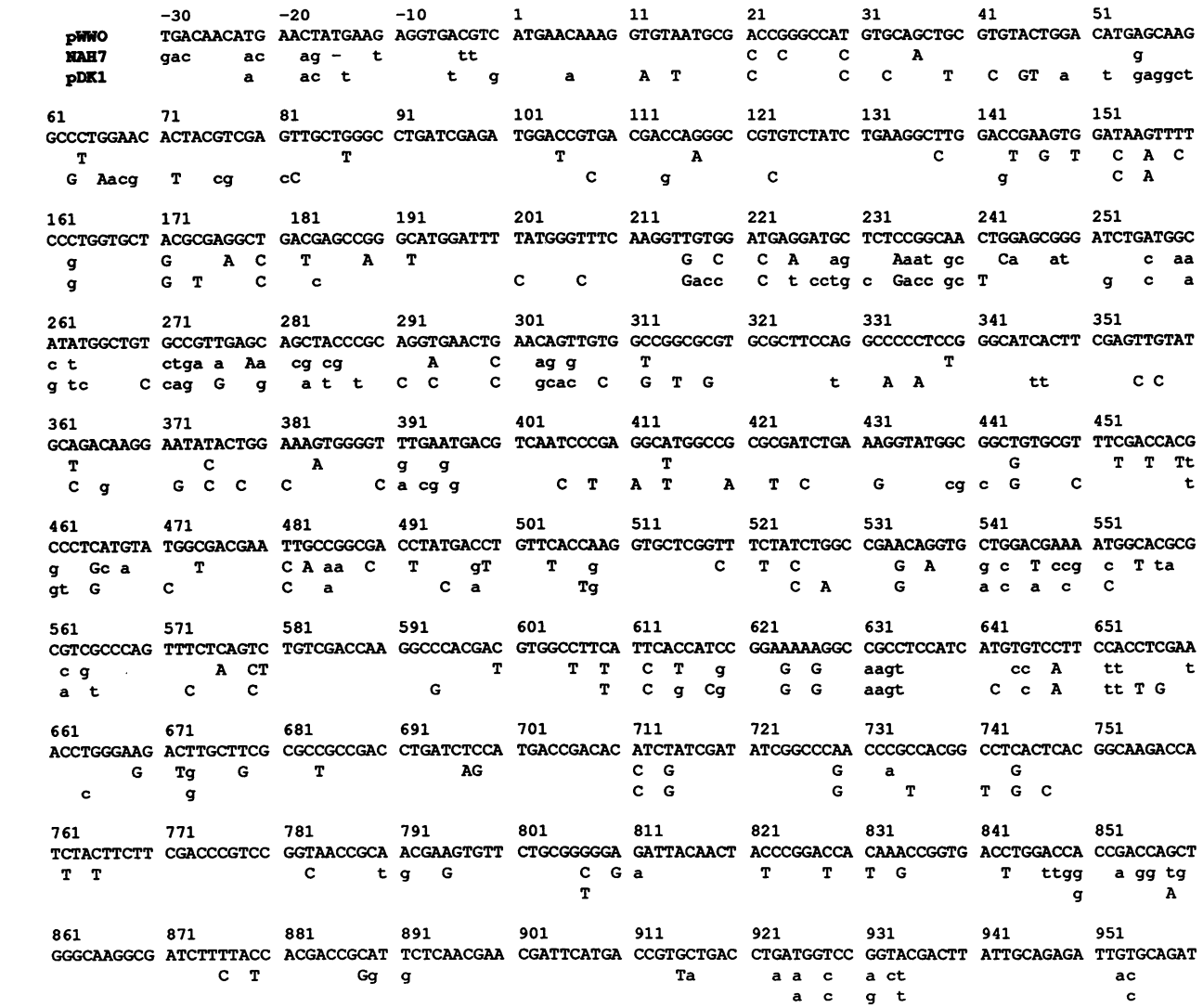

FIG. 3. Comparison of the nucleotide sequences of C23O genes encoded by TOL pWW0, NAH7, and TOL pDK1 plasmids. Positions in $xylE_{pDK1}$ and *nahH* which are different from those in $xylE_{pWW0}$ are indicated below the pWW0 sequence. Lowercase letters refer to nucleotide changes that alter the amino acid sequence. Nucleotide changes giving rise to degenerate codons with no accompanying amino acid changes are indicated in uppercase letters.

revealed that the 3′ 471 nucleotides encoding the carboxy half of the protein were somewhat more conserved than the 5′ 450 nucleotides comprising the amino terminus (76 and 68%, respectively). Furthermore, the highest degree of nucleotide conservation (82.2%) was found within the 264 nucleotides that comprise the C terminus (nucleotides 658 to 921 [region VI]). Importantly, this region also showed the greatest homology when nucleotides from any pair of genes were compared (average, 87.9%). These results find analogy with recent work suggesting that it is the carboxy terminus of C23O which plays a key role in substrate binding and catalysis (21).

Further intimations were revealed when comparisons of additional regions were made. For example, a comparison of both DNA and amino acid sequence homologies within the amino-terminal portion of the gene gave the unexpected result that $xylE_{pWW0}$ and *nahH* were most similar (84.7% similarity for nucleotides and 86.7% similarity for amino acids). This level of homology existed despite the fact that regions I and III of the amino terminus are separated by a segment (II) of 99 nucleotides that exhibit the least conservation for the entire gene (48.5% at the nucleotide level). Analysis of variable positions within region II further revealed no consistent pattern of homology, with the most common observation being that all three genes were different. We interpret these observations as implying that evolutionary constraints on both nucleotide and amino acid sequences within this region are reduced and therefore the usefulness of the homology comparisons is decreased. In contrast to the N terminus, within the carboxy half of the gene, $xylE_{pDK1}$ and $xylE_{pWW0}$ were found to be most similar (88.5% for nucleotides, 89.2% for amino acids); it is noteworthy that the final 264 nucleotides comprising the 3′ end of the gene (VI) showed a striking 95.5% homology. Contrary to this, in region V, *nahH* and $xylE_{pDK1}$ exhibited the highest homology; however, the significance of this comparison is

TABLE 1. Comparison of nucleotide and predicted amino acid sequences of C23O genes

Region	Nucleotides		Amino acids	
	No.	% Homology	No.	% Homology
Complete gene (regions I to VI)	1 to 921		1 to 307	
Nonidentical positions	258/921	72.0	68/307	77.9
$xylE_{pWW0}/xylE_{pDK1}$ match	88/258	81.5	15/68	82.7
$xylE_{pDK1}/nahH$ match	55/258	78.0	16/68	83.1
$xylE_{pWW0}/nahH$ match	90/258	81.8	21/68	84.7
All 3 genes different	25/258		16/68	
N-terminal portion (regions I to III)	1 to 450		1 to 150	
Nonidentical positions	143/450	68.2	38/150	74.7
$xylE_{pWW0}/xylE_{pDK1}$ match	27/143	74.2	2/38	76.0
$xylE_{pDK1}/nahH$ match	24/143	73.6	6/38	78.7
$xylE_{pWW0}/nahH$ match	74/143	84.7	18/38	86.7
All 3 genes different	18/143		12/38	
C-terminal portion (regions IV to VI)	451 to 921		151 to 307	
Nonidentical positions	115/471	75.6	30/157	80.9
$xylE_{pWW0}/xylE_{pDK1}$ match	61/115	88.5	13/30	89.2
$xylE_{pDK1}/nahH$ match	31/115	82.2	10/30	87.3
$xylE_{pWW0}/nahH$ match	16/115	79.0	3/30	82.8
All 3 genes different	7/115		4/30	
Region I	1 to 213		1 to 71	
Nonidentical positions	56/213	73.7	12/71	83.1
$xylE_{pWW0}/xylE_{pDK1}$ match	15/56	80.8	0/12	83.1
$xylE_{pDK1}/nahH$ match	8/56	77.5	1/12	84.5
$xylE_{pWW0}/nahH$ match	33/56	89.2	10/12	97.2
All 3 genes different	0/56		1/12	
Region II	214 to 312		72 to 104	
Nonidentical positions	51/99	48.5	19/33	42.4
$xylE_{pWW0}/xylE_{pDK1}$ match	9/51	57.6	2/19	48.5
$xylE_{pDK1}/nahH$ match	13/51	61.6	4/19	54.5
$xylE_{pWW0}/nahH$ match	13/51	61.6	4/19	54.5
All 3 genes different	16/51		9/19	
Region III	313 to 450		105 to 150	
Nonidentical positions	36/138	73.9	7/46	84.8
$xylE_{pWW0}/xylE_{pDK1}$ match	3/36	76.1	0/7	84.8
$xylE_{pDK1}/nahH$ match	3/36	76.1	1/7	87.0
$xylE_{pWW0}/nahH$ match	28/36	94.2	4/7	93.5
All 3 genes different	2/36		2/7	
Region IV	451 to 606		151 to 202	
Nonidentical positions	49/156	68.6	12/52	76.9
$xylE_{pWW0}/xylE_{pDK1}$ match	25/49	84.6	3/12	82.7
$xylE_{pDK1}/nahH$ match	12/49	76.3	4/12	84.6
$xylE_{pWW0}/nahH$ match	8/49	73.7	2/12	80.8
All 3 genes different	4/49		3/12	
Region V	607 to 657		203 to 219	
Nonidentical positions	19/51	62.7	6/17	64.7
$xylE_{pWW0}/xylE_{pDK1}$ match	1/19	64.7	1/6	70.6
$xylE_{pDK1}/nahH$ match	13/19	88.2	4/6	88.2
$xylE_{pWW0}/nahH$ match	4/19	70.6	1/6	70.6
All 3 genes different	1/19		0/6	
Region VI	658 to 921		220 to 307	
Nonidentical positions	47/264	82.2	12/88	86.4
$xylE_{pWW0}/xylE_{pDK1}$match	35/47	95.5	9/12	96.6
$xylE_{pDK1}/nahH$ match	6/47	84.5	2/12	88.6
$xylE_{pWW0}/nahH$ match	4/47	83.7	0/12	86.4
All 3 genes different	2/47		1/12	

less clear given that this region comprises only 5% of the gene.

Although the overall structure of $xylE_{pDK1}$ appears to be very similar to that of $xylE_{pWW0}$ and *nahH*, it was surprising to find that the relatedness between the two *xylE* genes was no greater than that between either *xylE* gene and *nahH*. A priori, one may have hypothesized that the two *xylE* genes would be more similar given that they occur in isofunctional operons. The finding that, at the N terminus, the resemblance is greater for $xylE_{pWW0}$ and *nahH* while, at the C terminus, $xylE_{pWW0}$ and $xylE_{pDK1}$ are more related could be interpreted to mean that $xylE_{pWW0}$ represents a genetic hybrid of the other two C23Os, i.e., having borrowed the N-terminal half from *nahH* and the C-terminal half from $xylE_{pDK1}$. This hypothesis is certainly consistent with the well-recognized propensity of bacterial plasmid genes to undergo genetic recombination. This further finds analogy with the notion that proteins may evolve piecemeal via the interchange of DNA encoding separate domains (7, 15). Alternatively, it is possible that $xylE_{pWW0}$ and $xylE_{pDK1}$ diverged after the separation of *xylE* and *nahH* ancestors, but, because of selective pressure exerted on the hosts of $xylE_{pWW0}$ and *nahH*, the nucleotide sequences encoding the N-terminal half of these two genes were better conserved than those of $xylE_{pDK1}$. In any event, the present work suggests that the evolution of C23Os may have occurred independently of the plasmid-encoded pathway (TOL, NAH) in which they are found. At the same time, evolutionary constraints appear to have been greater on the structural conservation of the C terminus than on other regions of the gene, a conclusion consistent with recent work indicating that it is this part of C23O which is most important catalytically (21).

This research was supported by grant 003594-030 from the Texas State Higher Education Coordinating Board.

REFERENCES

1. **Burlage, R. S., S. W. Hooper, and G. S. Sayler.** 1989. The TOL (pWW0) catabolic plasmid. Appl. Environ. Microbiol. **55**:1323–1328.
2. **Chatfield, L. K., and P. A. Williams.** 1986. Naturally occurring TOL plasmids in *Pseudomonas* strains carry either two homologous or two nonhomologous catechol-2,3-oxygenase genes. J. Bacteriol. **168**:878–885.
3. **Crawford, I. P., B. P. Nichols, and C. Yanofsky.** 1980. Nucleotide sequence of the *trpB* gene in *Escherichia coli* and *Salmonella typhimurium*. J. Mol. Biol. **142**:489–502.
4. **Ghosal, D., I.-S. You, and I. C. Gunsalus.** 1987. Nucleotide sequence and expression of gene *nahH* of plasmid NAH7 and homology with gene *xylE* of TOL (pWW0). Gene **55**:19–28.
5. **Harayama, S., M. Rekik, A. Wasserfallen, and A. Bairoch.** 1987. Evolutionary relationships between catabolic pathways for aromatics: conservation of gene order and nucleotide sequences of catechol oxidation genes of pWW0 and NAH7 plasmids. Mol. Gen. Genet. **210**:241–247.
6. **Harayama, S., and K. N. Timmis.** 1989. Catabolism of aromatic hydrocarbons by *Pseudomonas*, p. 151–174. *In* D. A. Hopwood and K. I. Chater (ed.), Genetics of bacterial diversity. Academic Press, Inc. (London), Ltd., London.
7. **Hartl, E. L.** 1989. Anecdotal, historical and critical commentaries on genetics. Genetics **122**:1–6.
8. **Johnston, J. B., and I. C. Gunsalus.** 1977. Isolation of metabolic plasmid DNA from *Pseudomonas putida*. Biochem. Biophys. Res. Commun. **75**:13–19.
9. **Kaplan, J. B., W. K. Merkel, and B. P. Nichols.** 1985. Evolution of glutamine amidotransferase genes. Nucleotide sequences of the pabA genes from *Salmonella typhimurium*, *Klebsiella aerogenes* and *Serratia marcescens*. J. Mol. Biol. **183**:327–340.
10. **Keil, H., M. R. Lebens, and P. A. Williams.** 1985. TOL plasmid pWW15 contains two nonhomologous independently regulated catechol-2,3-oxygenase genes. J. Bacteriol. **163**:248–255.
11. **Kunz, D. A., and P. J. Chapman.** 1981. Isolation and characterization of spontaneously occurring TOL plasmid mutants of *Pseudomonas putida* HS1. J. Bacteriol. **146**:952–964.
12. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labelled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
13. **Nakai, C., H. Kagamiyama, M. Nozaki, T. Nakazawa, S. Inouye, Y. Ebina, and A. Nakazawa.** 1983. Complete nucleotide sequence of the metapyrocatechase gene on the TOL plasmid of *Pseudomonas putida* mt-2. J. Biol. Chem. **258**:2923–2928.
14. **Osborne, D. J., R. W. Pickup, and P. A. Williams.** 1988. The presence of two complete homologous *meta* pathway operons on TOL plasmid pWW53. J. Gen. Microbiol. **134**:2965–2975.
15. **Phillips, D. C., M. J. E. Sternberg, and B. J. Sutton.** 1983. Intimations of evolution from the three-dimensional structures of proteins, p. 145–173. *In* D. S. Bendall (ed.), Evolution from molecules to men. Cambridge University Press, Cambridge, England.
16. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
17. **Schrier, P. H., and R. Cortese.** 1979. A fast and simple method for sequencing DNA cloned in the single-stranded bacteriophage M13. J. Mol. Biol. **129**:169–172.
18. **Shaw, L. E., and P. A. Williams.** 1988. Physical and functional mapping of two cointegrate plasmids derived from RP4 and TOL plasmid pDK1. J. Gen. Microbiol. **134**:2463–2474.
19. **Tanaka, T., and B. Weisblum.** 1975. Construction of a colicin E1-R factor composite plasmid in vitro: means for amplification of deoxyribonucleic acid. J. Bacteriol. **121**:354–362.
20. **Voss, J., D. Kunz, and R. Benjamin.** 1989. Abstr. Annu. Meet. Am. Soc. Microbiol. 1989, K 182, p. 275.
21. **Williams, P. A., S. J. Assinder, and L. E. Shaw.** 1990. Construction of hybrid *xylE* genes between the two homologous genes from TOL plasmid pWW53: comparison of the kinetic properties of the gene products. J. Gen. Microbiol. **136**:1583–1589.
22. **Williams, P. A., L. E. Gibb, H. Keil, and D. J. Osborne.** 1988. Organization of and relationships between catabolic genes of TOL plasmids, p. 339–358. *In* S. R. Hagedorn, R. S. Hanson, and D. A. Kunz (ed.), Microbial metabolism and the carbon cycle. Harwood Academic Publishers, New York.
23. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of M13mp18 and pUC19 vectors. Gene **33**:103–119.