Evolution of Isopenicillin N Synthase Genes May Have Involved Horizontal Gene Transfer[1]

Giddy Landan, Gerald Cohen,† Yair Aharonowitz,† Yuval Shuali,‡ Dan Graur,‡ and Dov Shiffman†*

*The Interdisciplinary Program for Fostering Excellence, Tel Aviv University; and Departments of †Microbiology and ‡Zoology, George S. Wise Faculty of Life Sciences, Tel Aviv University

The isopenicillin N synthase genes from three fungal species, three Gram-positive species, and one Gram-negative bacterial species share an unusually high sequence similarity. A phylogentic analysis was carried out to determine which type of evolutionary scenario best accounts for this similarity. The most plausible scenario is one in which a horizontal gene-transfer event, from the prokaryotes to the eukaryotes, occurred at a time close to the divergence between the Gram-positive and the Gram-negative bacteria.

Introduction

Horizontal gene transfer is defined as the transfer of genetic information from one genome to another, specifically between species. There are very few cases in which horizontal gene transfer has been convincingly demonstrated (Benveniste 1985), and in even fewer cases have the transferred genes retained their functionality (Gray and Fitch 1983; Hensel et al. 1989). Indeed, a horizontally transferred gene is not expected to remain functional in the host species, because such a gene has been probably reverse transcribed and/or replicated with error and most likely no longer contains the proper signals for transcription, mRNA maturation, and translation.

The isopenicillin N synthase (IPNS) gene has been found in a variety of microorganisms that produce penicillin and cephalosporin antibiotics, including unicellular eukaryotic species and Gram-positive ($Gram^+$) and Gram-negative ($Gram^-$) prokaryotes. At present seven IPNS genes have been cloned and sequenced: three from filamentous fungi, three from $Gram^+$ mycelial streptomycetes, and one from a $Gram^-$ bacterium. Comparison of the predicted amino acid sequences shows that the fungal and microbial proteins share >50% sequence identity. In contrast, comparisons between typical homologous proteins from prokaryotes and eukaryotes yield considerably lower sequence identities (Doolittle et al. 1986; Bardwell and Craig 1987; Hensel et al. 1989).

Weigel et al. (1988) advanced two possible explanations for the unusually high sequence similarity between the microbial and fungal IPNS genes. The similarity may have resulted from a slow but constant rate of evolutionary change of the gene, reflecting strict functional constraints on the IPNS protein, constraints that precluded most amino acid changes. Alternatively, the similarity may have resulted from a horizontal

1. Key words: horizontal gene transfer, isopenicillin N synthase.

Address for correspondence and reprints: Dan Graur, Department of Zoology, George S. Wise Faculty of Life Sciences, Tel Aviv University, Ramat Aviv 69978, Israel.

Mol. Biol. Evol. 7(5):399–406. 1990.

0737-4038/90/0705-0001$02.00

gene-transfer event from the bacterial lineage to the fungal lineage after the eukaryote-prokayote split. Weigel et al. prefer the latter explanation, since the rate of molecular evolution of the fungal IPNS gene is not very different from that of other fungal genes and because IPNS is a secondary metabolism gene that does not appear to be essential for the survival of the organism. Although the transfer hypothesis appears to be the more plausible of the two propositions, other evolutionary scenarios need to be considered. The purpose of the present note is (*a*) to enumerate the possible evolutionary routes that may account for the remarkable similarity between the IPNS genes of distantly related species and (*b*) to show which of these routes is compatible with the available sequence data.

Data

Seven DNA sequences of IPNS structural genes were analyzed: three from the Gram$^+$ streptomycetes *Streptomyces clavuligerus* (Leskiw et al. 1988), *S. lipmanii* (Shiffman et al. 1988; Weigel et al. 1988) and *S. jumonjinensis* (Shiffman et al. 1988); one from the Gram$^-$ bacterium *Flavobacterium sp.*, strain SC 12,154 (Shiffman et al., 1990); and three from the fungal species *Penicillium chrysogenum* (Carr et al. 1987), *Cephalosporium acremonium* (Samson et al. 1985), and *Aspergillus nidulans* (Ramon et al. 1987; Weigel et al. 1988). The coding sequences of the intronless genes vary in length from 978 bp in *Flavobacterium* to 1,014 bp in *C. acremonium.* Alignment of the sequences was achieved by the method of Wilbur and Lipman (1983). Improvements in the alignment were introduced by visual inspection. Only 945 nucleotide sites (314 sense codons + termination) for which an unambiguous alignment was obtained were used (fig. 1).

Results

The nucleotide sequences of IPNS genes exhibit a very high degree of sequence identity (table 1). Within the streptomycetes the sequence identity is 91%–92%, and within fungi it is 81%–85%. Sequence identity between *Flavobacterium* and streptomycetes is 82%, between *Flavobacterium* and fungi it is 73%–80%, and between streptomycetes and fungi it is 75%–84%.

Using (*a*) Fitch's (1977) maximum-parsimony method and (*b*) the neighbor-joining method (Saitou and Nei 1987) with number of nonsynonymous substitutions per site (Nei and Gojobori 1986), we obtained the same unrooted tree for the seven IPNS genes. This tree requires a total of 494 amino acid replacements. The unrooted tree obtained by the neighbor-joining method is shown in figure 2. The three taxonomic classes are clearly distinguishable.

Positioning the root on a phylogenetic unrooted tree is possible either by reference to an outgroup or by assuming a constant evolutionary rate over the entire tree. Since we do not have an outgroup sequence, we assumed that the molecular clock applies for the tree, and we consequently position the root at a point preceding the Gram$^+$/Gram$^-$ divergence (see arrow in fig. 2).

Discussion

To reconstruct the evolutionary history of the IPNS gene two main issues must be clarified: (*a*) the rate of evolutionary change of the gene and (*b*) the type of horizontal gene-transfer event that may have taken place. We can therefore envision four possible evolutionary scenarios: (*a*) a constant rate of evolution with no horizontal gene transfer, (*b*) a constant rate of evolution with a gene-transfer event, (*c*) varying rates of evolution

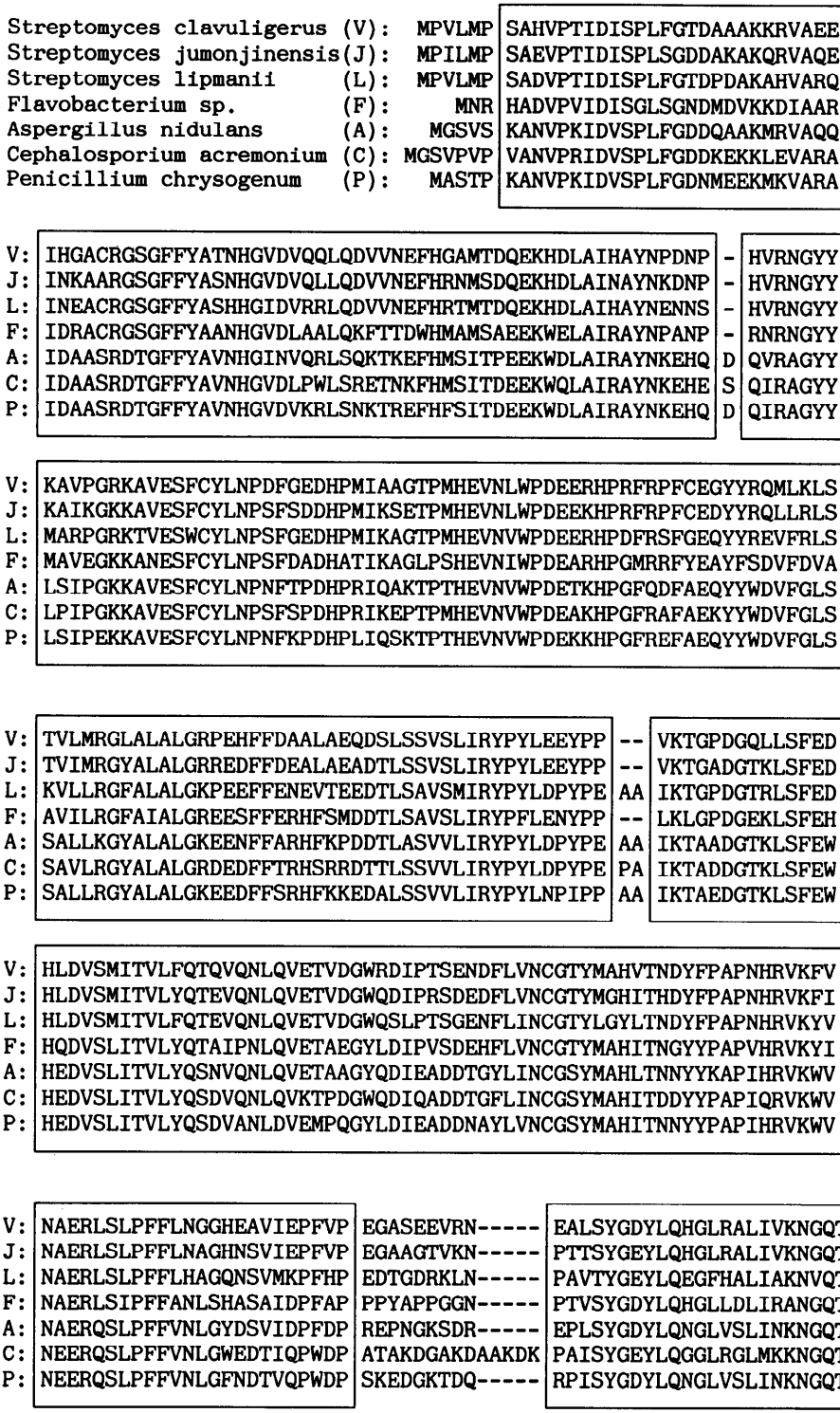

FIG. 1.—Predicted amino acid sequences of seven IPNS genes: V = *Streptomyces clavuligerus* (Leskiw et al. 1988); J = *S. jumonjinensis* (Shiffman et al. 1988); L = *S. lipmanii* (Shiffman et al. 1988; Weigel et al. 1988); F = *Flavobacterium* species (Shiffman et al. 1990); A = *Aspergillus nidulans* (Ramon et al. 1987; Weigel et al. 1988); C = *Cephalosporium acremonium* (Samson et al. 1985); and P = *Penicillium chrysogenum* (Carr et al. 1987). Boxed amino acid residues represent the 314 amino acids for which an unambiguous alignment was obtained. Numbers denote amino acid position.

Table 1
Percent Sequence Similarity between Seven IPNS Genes

Species	Species						
	Streptomyces clavuligerus	*S. jumonjinensis*	*S. lipmanii*	*Flavobacterium* species	*Aspergillus nidulans*	*Cephalosporium acremonium*	*Penicillium chrysogenum*
Streptomyces clavuligerus		82.2	73.3	61.6	60.3	59.4	58.1
S. jumonjinensis	92.0		72.4	62.9	61.6	63.2	59.7
S. lipmanii	91.9	91.3		58.4	60.0	58.7	56.2
Flavobacterium species	82.4	82.2	81.8		58.4	58.1	57.1
A. nidulans	75.6	76.0	77.1	73.2		76.8	82.5
C. acremonium	81.8	83.7	83.5	81.0	81.4		78.1
P. chrysogenum	79.4	78.1	79.1	78.3	82.9	85.2	

NOTE.—Similarities at the amino acid level are shown above the diagonal; similarities at the nucleotide level are shown below the diagonal.

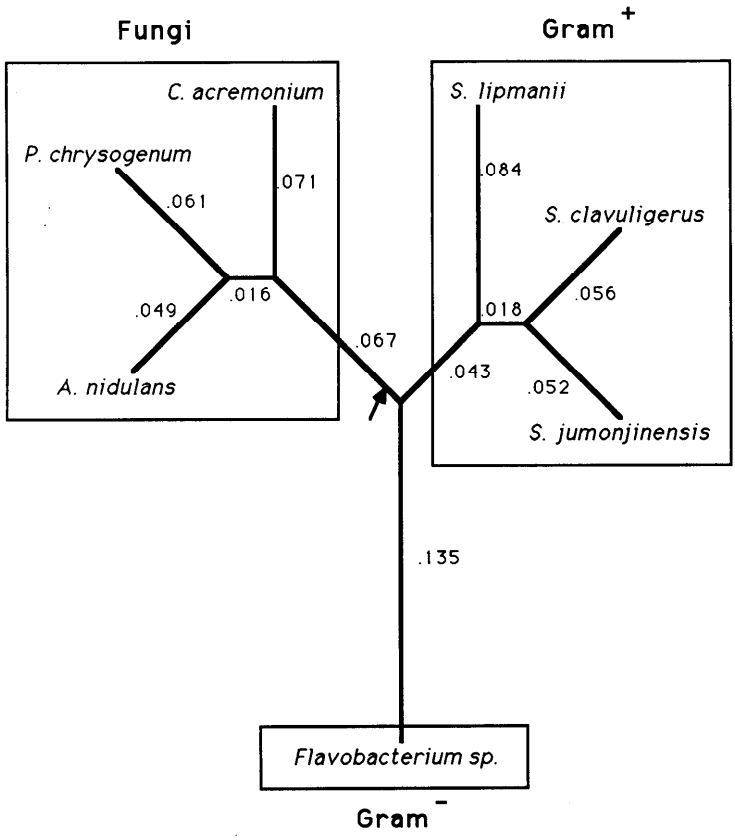

FIG. 2.—Unrooted phylogenetic tree derived from analysis of IPNS nucleotide sequences. The tree was constructed according to the neighbor-joining method (Saitou and Nei 1987). The branch length is proportional to the number of nonsynonymous substitutions per site (Nei and Gojobori 1986). The arrow indicates the position of the root, obtained by dividing into two equal parts the longest pathway in the tree.

among the lineages with no gene transfer, and (*d*) varying rates of evolution with a gene-transfer event. From a parsimonious point of view, scheme (*d*) is not a satisfactory explanation, since it involves two independent assumptions. Consequently, this scheme will not be considered further. In the following we attempt to determine which of the other three schemes—(*a*), (*b*), or (*c*)—is the most consistent with the data.

Constant Rate of Evolution

If it is assumed that the rate-constancy hypothesis holds, the root and the relative lengths of the branches can be inferred (fig. 3a). Figure 3b shows a phylogenetic tree for Gram$^+$, Gram$^-$, and fungi that is based on 5S rRNA sequences (Hori and Osawa 1987). In Hori and Osawa's tree, the ratio a/b is 0.42, while in the IPNS tree a/b = 0.94. In other words, by using the rate-constancy hypothesis, Hori and Osawa concluded that eukaryotes and prokaryotes diverged ~2.4 billion years ago, while the Gram$^+$ and Gram$^-$ bacteria diverged ~1 billion years ago. In comparison, if we assume a constant rate of substitution in the case of the IPNS genes, and if we use Hori and Osawa's estimate for the eukaryote-prokaryote divergence, then the Gram$^+$ and Gram$^-$ bacteria would have diverged from each other 2.3 billion years ago. Alternatively, if the divergence between Gram$^+$ and Gram$^-$ is assumed to have occurred

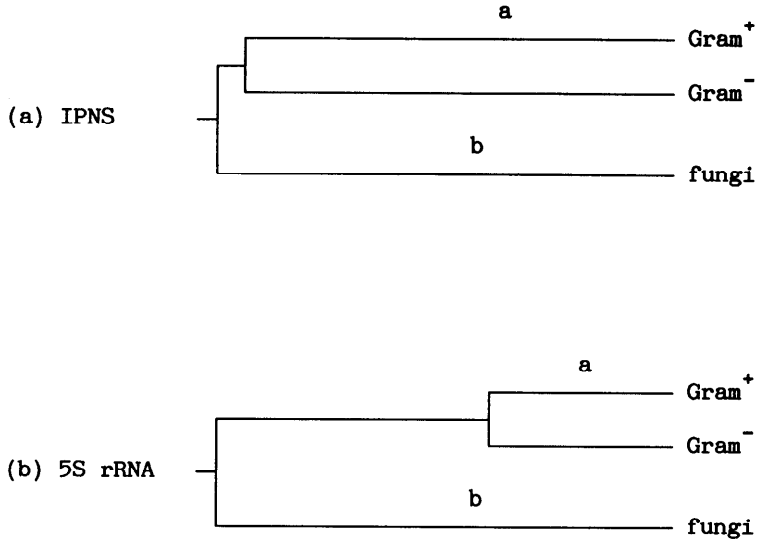

FIG. 3.—Schematic comparison of phylogenetic trees of [panel (a)] IPNS and [panel (b)] 5S rRNA genes, for three taxonomic classes: Gram$^+$, Gram$^-$, and fungi. The rooted phylogenetic tree in panel (a) was obtained by the neighbor-joining method (Saitou and Nei 1987) and by assuming a constant substitution rate. The 5S rRNA phylogenetic tree [in panel (b)] was adapted from Hori and Osawa (1987).

1 billion years ago, then the prokaryote-eukaryote split should have occurred at about the same time, obviously a gross underestimate. Moreover, the 5S rRNA studies of Chen et al. (1984) show that the intrafungal and intramicrobial similarities are much like those found in the IPNS genes, whereas the similarities between the fungal and bacterial 5S rRNA sequences are considerably less than those in the IPNS genes. The constant-rate hypothesis, (*a*), is therefore incompatible with the data.

Varying Rates of Evolution

Another possible explanation for the high similarity observed between the fungal and bacterial IPNS sequences is that the rate of substitutions may have varied among the lineages under study. For example, the IPNS genes could have evolved, on average, more than six times more slowly prior to the divergence among the fungi and the bacteria than it did after the divergence.

If there are differences in the rates of amino acid replacement among the lineages, and if these differences are caused by changes in the intensity of purifying selection, we expect to find different patterns of amino acid replacements in the different lineages. To measure the intensity of purifying selection, we used the mean chemical distance between proteins (Graur 1985), where higher mean chemical distances reflect weaker selective pressures. The mean chemical distance between neighboring nodes of the species tree were compared. In terms of its mean chemical distance, no tree segment was found to be significantly different from any of the others. This indicates that, if there are differences in the rate of replacement among the different lineages, these differences are not caused by varying intensities of purifying selection. It is, however, impossible to rule out differences in mutation rates. The mean chemical distance averaged ~58, a value that implies very intense purifying selective pressures in all the lineages.

Gene Transfer

If it is assumed that a gene-transfer event is the cause for the high sequence similarity between fungal and bacterial IPNS genes, the question of the time and direction of the transfer still needs to be addressed. The unrooted tree reconstructed by the maximum-parsimony method clearly shows that the three Gram$^+$ genes form a natural clade and that the three fungal genes form another. A transfer event therefore could not have happened after the fungal or bacterial speciation. Thus, the topology of the tree turns out not to be informative as to the direction of the horizontal gene transfer. Other data suggest that, if a transfer event did occur, its direction was probably from the bacteria to the fungi. In bacteria, the IPNS gene is part of an antibiotic gene cluster, whereas in some fungi the genes encoding the enzymes of the pathway are dispersed over several chromosomes (Kovacevic et al. 1989). In addition, the bacteria possess a more elaborate biosythetic capacity to produce β-lactam antibiotics than do the fungi. Also, all the IPNS genes lack introns, which favors the idea of a transfer from the bacteria to the fungi, given that fungal genes sometimes possess introns. Thus, the similarity between eukaryotic and prokaryotic IPNS genes represents a dramatic instance of functional xenology, i.e., homologous genes that have retained their functionality long after a horizontal gene-transfer event between distinct species (Gray and Fitch 1983).

The IPNS gene is not the only penicillin synthase gene to exhibit high conservation between fungi and bacteria. Deacetoxycephalosporin C synthase, which is responsible for the expansion of the thiazolidine ring of penicillin N, is also highly conserved (Kovacevic et al. 1989). Thus, if a transfer event occurred, it probably involved not only the IPNS gene but also some other genes belonging to the penicillin biosynthetic pathway.

The question of whether a horizontal gene-transfer event actually occurred will be answered unequivocally only when an outgroup sequence becomes available. Such an outgroup could be used to locate the root of the phylogenetic tree obtained by the maximum-parsimony method, thereby proving or refuting the horizontal gene-transfer hypothesis. A possible outgroup sequence might be an IPNS gene from archaebacteria, if indeed archaebacteria possess such a gene. Alternatively, a paralogous gene that diverged from IPNS prior to the eukaryote-prokaryote split can be used. The conclusion would, then, be a topological one, i.e., independent both of rates and of assumptions of rate constancy. The time scales under study are very long, and the IPNS gene is obviously a very conservative gene. Consequently, conclusions based on rate considerations alone are likely to remain tentative at best, even if more IPNS sequences from other related species become available.

Acknowledgments

G.C., Y.A., and D.S. were supported by a joint grant from the National Council for Research and Development (Israel) and by the G.S.F. (Munich). D.G. and G.L. were supported by a grant from the U.S.-Israel Binational Science Foundation. Part of the study was performed when D.G. was a visiting lecturer at the Institute of Genetics and Ecology, Aarhus University (Denmark).

LITERATURE CITED

BARDWELL, J. C. A., and E. A. CRAIG. 1987. Eukaryotic M_r 83,000 heat shock protein has a homologue in *Escherichia coli*. Proc. Natl. Acad. Sci. USA **84**:5177–5181.

BENVENISTE, R. E. 1985. The contribution of retroviruses to the study of mammalian evolution. Pp. 359–417 *in* R. J. MACINTYRE, ed. Molecular evolutionary genetics. Plenum, New York.

CARR, L. G., P. L. SKATRUD, M. E. SCHEETZ II, S. W. QUEENER, and T. D. INGOLIA. 1987. Cloning and expression of the isopenicillin N synthetase gene from *Penicillium chrysogenum.* Gene **48**:257–266.

CHEN, M.-W., J. ANNE, G. VOLCKAERT, E. HUYSMANS, A. VANDENBERGHE, and R. DE WACHTER. 1984. The nucleotide sequences of the 5S rRNAs of seven molds and a yeast and their use in studying ascomycete phylogeny. Nucleic Acids Res. **12**:4881–4892.

DOOLITTLE, R. F., D. F. FENG, M. S. JOHNSON, and M. A. MCCLURE. 1986. Relationship of human protein sequences to those of other organisms. Cold Spring Harbor Symp. Quant. Biol. **51**:447–455.

FITCH, W. M. 1977. On the problem of discovering the most parsimonious tree. Am. Nat. **111**: 223–257.

GRAUR, D. 1985. Amino acid composition and the evolutionary rates of protein coding genes. J. Mol. Evol. **22**:53–62.

GRAY, G. S., and W. M. FITCH. 1983. Evolution of antibiotic resistance genes: the DNA sequence of a kanamycin resistance gene from *Staphylococcus aureus.* Mol. Biol. Evol. **1**:57–66.

HENSEL, R., P. ZWICKL, S. FABRY, J. LANG, and P. PALM. 1989. Sequence comparison of glyceraldehyde-3-phosphate dehydrogenases from three urkingdoms: evolutionary implications. Can. J. Microbiol. **35**:81–85.

HORI, H., and S. OSAWA. 1987. Origin and evolution of organisms as deduced from 5S ribosomal RNA sequences. Mol. Biol. Evol. **4**:445–472.

KOVACEVIC, S., B. J. WEIGEL, M. B. TOBIN, T. D. INGOLIA, and J. R. MILLER. 1989. Cloning, characterization and expression in *Escherichia coli* of the *Streptomyces clavuligerus* gene encoding deacetoxycephalosporin C synthetase. J. Bacteriol. **171**:754–760.

LESKIW, B. K., M. AHARONOWITZ, M. MEVARECH, S. WOLFE, L. C. VINING, D. W. S. WESTLAKE, and S. JENSEN. 1988. Cloning and nucleotide sequence determination of the isopenicillin N synthetase gene from *Streptomyces clavuligerus.* Gene **62**:187–196.

NEI, M., and T. GOJOBORI. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. **3**:418–426.

RAMON, D., L. CARRAMOLINO, C. PATINO, F. SANCHEZ, and M. A. PENALVA. 1987. Cloning and characterization of the isopenicillin N synthetase gene mediating the formation of the β-lactam ring in *Aspergillus nidulans.* Gene **57**:171–181.

SAITOU, N., and M. NEI. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4**:406–425.

SAMSON, S. M., R. BELAGAJE, D. T. BLANKENSHIP, J. L. CHAPMAN, D. PERRY, P. L. SKATRUD, R. M. VANFRANK, E. P. ABRAHAM, J. E. BALDWIN, S. W. QUEENER, and T. D. INGOLIA. 1985. Isolation, sequence determination, and expression in *E. coli* of the isopenicillin N synthetase gene from *Cephalosporium acremonium.* Nature **318**:191–194.

SHIFFMAN, D., G. COHEN, Y. AHARONOWITZ, H. PALISSA, H. VON DOHREN, H. KLEINKAUF, and M. MEVARECH. 1990. Nucleotide sequence of the isopenicillin N synthase gene (*pcbC*) of the unicellular Gram negative *Flavobacterium sp.* SC12,154. Nucleic Acids Res. **18**:660.

SHIFFMAN, D., M. MEVARECH, S. E. JENSEN, G. COHEN, and Y. AHARONOWITZ. 1988. Cloning and comparative sequence analysis of the gene coding for isopenicillin N synthase in *Streptomyces.* Mol. Gen. Genet. **214**:562–569.

WEIGEL, B. J., S. G. BURGETT, J. V. CHEN, P. L. SKATRUD, C. A. FROLIK, S. W. QUEENER, and T. D. INGOLIA. 1988. Cloning and expression in *Escherichia coli* of isopenicillin N synthetase from *Streptomyces lipmanii* and *Aspergillus nidulans.* J. Bacteriol. **170**:3817–3826.

WILBUR, W. J., and D. J. LIPMAN. 1983. Rapid similarity searches of nucleic acid and protein data banks. Proc. Natl. Acad. Sci. USA **80**:726–730.

MASATOSHI NEI, reviewing editor

Received July 27, 1989; revision received March 20, 1990

Accepted May 8, 1990