Proc. Natl. Acad. Sci. USA
Vol. 85, pp. 9101–9105, December 1988
Evolution

Tandemly repeated tRNA pseudogenes in photobacterium

(tRNA^{Pro}/gene cluster/evolution/*rho*-independent terminator)

SYLVIE GIROUX AND ROBERT CEDERGREN

Département de biochimie, Université de Montréal, Montréal, PQ H3C 3J7, Canada

Communicated by Boris Magasanik, June 23, 1988 (received for review March 30, 1988)

ABSTRACT **A region distal to three tRNA genes in *Photobacterium phosphoreum*, a Gram-negative eubacterium, unexpectedly contains a high number of repeated DNA segments that are closely related to the adjacent tRNA^{Pro} gene. The 5′ to 3′ order of this cluster is tRNA^{Pro}–tRNA^{His}–tRNA^{Pro} followed by eight tRNA^{Pro}-like structures interspersed by *rho*-independent terminators. The two tRNA^{Pro} genes, which are identical, and the tRNA^{His} gene have 86% and 87% positional identity, respectively, to their counterparts in the *argT* operon of *Escherichia coli*. The facts that these tRNA-like structures are not transcribed, in contrast to the tRNA retropseudogenes of eukaryotes, and that these structures are clustered near their progenitor suggest they are an unusual class of tRNA pseudogenes that arose by tandem duplication.**

The DNA-sequencing revolution has spawned a plethora of information on the structure of gene and gene products. This rapidly expanding data base has provided impetus for using gene sequences to probe evolutionary relationships among different organisms (1–3). However, upon this background of recently discovered molecular phylogenies, there is growing awareness that genomes themselves, far from being static repositories of genetic information, are mutable, fluid entities (4, 5). Therefore, whether, or to what extent, phylogenies based on a small portion of the genome faithfully portray the evolutionary relatedness among organisms becomes a valid question (6).

To evaluate the relative evolutionary importance of gene sequence (gene evolution) *versus* gene organization (genome evolution), a systematic study of gene arrangement is necessary; we have adopted tRNA gene clusters in eubacteria as a model. A cloned DNA fragment of one such cluster, the *argT* operon of *Escherichia coli* (7) containing the genes of tRNA^{Arg}, tRNA^{His}, tRNA^{Leu}, and tRNA^{Pro}, has been used as a genomic probe with DNA from a variety of bacteria and has permitted a preliminary evaluation of the similarity between each genomic DNA sample and the *E. coli* operon. These results directed the choice of the vibrio, *Photobacterium phosphoreum*, for further studies, because this organism was the most phylogenetically distant of those organisms that had DNA yielding strong hybridization signals with the *E. coli* operon. Vibrios are classified among the γ-3 subdivision of purple bacteria, close relatives to the enteric bacteria according to the bacterial taxonomy established by Woese *et al.* (8) from ribosomal RNA data. The 5S RNA sequences of the *E. coli* and *P. phosphoreum* have 84% positional identity (9). We report here the characterization and the unusual sequence* of a genomic DNA segment from *P. phosphoreum*, which hybridizes to the *E. coli argT* probe.

MATERIAL AND METHODS

Strain, Plasmid, Phage, and RNA. *P. phosphoreum* was obtained from the American Type Culture Collection (ATCC 11040). Plasmid pLC25-25 containing *argT* was provided by M. J. Fournier (University of Massachusetts, Amherst). The 1.9-kilobase (kb) fragment of DNA containing the *argT* operon was prepared by digestion of pLC25-25 with *Bam*HI and *Eco*RI and then ligated with pBR327 to give pBE1935. *E. coli* HB101 (10) was used as a recipient host for transformations with pBR327 (10). *E. coli* JM101 (11) was used to propagate the phage M13mp19. Pure tRNA$_3^{Pro}$ of *E. coli* was obtained from Subriden (Seattle).

DNA Preparation. *Photobacterium* was grown at 18°C in a medium containing 10 g of Bactotryptone, 20 g of NaCl, 5 g of yeast extract, 50 ml of 1 M Tris·HCl, pH 7.5, and 3 ml of glycerol per liter (using tap water); the pH was adjusted to 7.5 with HCl. A single colony was used to inoculate 20 ml of medium, which was then shaken overnight at 18°C. Genomic DNA was extracted as described by Birnboim and Doly (13).

Cloning tRNA Genes. A preparative *Sau3*A digest of 20 μg of genomic DNA was electrophoresed through a 1.0% agarose gel, and based on previous Southern blot information, the gel region containing DNA fragments of 2 to 3 kb was excised. The DNA was recovered from the gel slice by electroelution (unidirectional electroeluter, model UEA (International Biotechnologies, New Haven, CT). This size-enriched DNA was ligated to *Bam*HI-restricted pBR327, which had been previously dephosphorylated. The ligation mixture transformed *E. coli* HB101, and 1500 colonies were screened by colony hybridization (11) with a synthetic DNA probe complementary to the anticodon loop of tRNA^{His}, 5′-TGGAATCACAATCCA-3′. One of the positive clones was named pPPS70.

Subcloning and Nucleotide Sequencing. The *Sau*3A restriction fragment from pPPS70 was excised from an agarose gel and electroeluted as described above. The fragment was then ligated to dephosphorylated, *Bam*HI-restricted M13mp19. Both orientations were isolated and identified using synthetic DNA probes complementary to the coding strand or the anticoding strand of tRNA^{His}. Sequencing of both DNA strands was performed using the dideoxy chain-termination method of Sanger *et al.* (14) with different synthetic primers.

Northern (RNA) Blot and *in Vitro* Transcription Assay. A 2.0-kb *Hae*III fragment was ligated to *Sma* I-restricted pKK223-3 (15) to give the plasmid pKKH1 with the correct orientation for transcription. The 2.3-kb *Sau*3A fragment was filled-in and ligated with *Sma* I-restricted pKK223-3 to give pKKS4 with the correct orientation. *Bam*HI was used to liberate the DNA fragment containing the promoter and the insert from both plasmids. Fragments were purified as described above, and an *in vitro* transcription assay was performed as has been described (16). Crude transcription mixtures were analyzed by electrophoresis on 5% polyacrylamide gels containing 8.0 M urea.

Abbreviation: nt, nucleotide(s).

*The sequence reported in this paper is being deposited in the EMBL/GenBank data base (IntelliGenetics, Mountain View, CA, and Eur. Mol. Biol. Lab., Heidelberg) (accession no. X12975).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Cellular RNA was prepared by a phenol extraction of cells as described by Reed *et al.* (17). Samples of one OD_{260} unit of each extract was loaded on a 12% polyacrylamide gel containing 8.0 M urea. Northern blots were performed by electroblotting onto Hybond-N membranes (Amersham) in 0.25× TBE (1× TBE = 100 mM Tris/89 mM boric acid, pH 6.5/89 mM EDTA) for 2.5 hr at 30 V. The resulting membranes were probed with the 2.0-kb *Hae* III DNA fragment isolated from pPPS70, which had been 5′-labeled with ^{32}P by standard methods (11). Hybridizations were carried out as described in the protocol supplied by the manufacturer.

RESULTS

Characterization and Sequence of the *Photobacterium* tRNA Cluster. Probing of the *Sau*3A-restricted genomic DNA from *Photobacterium* permitted the identification of two 2.5-kb fragments that hybridized with the *argT* operon from *E. coli*. Because a variety of restriction enzymes gave two or more bands, *Photobacterium* must contain two different clusters of tRNA genes similar to *argT*. The positive clones obtained were found to contain the smaller (and less intense) *Sau*3A fragment. The possibility of rearrangements in these clones (identified as pPPS70) was precluded by showing that DNA fragments from pPPS70 or chromosomal DNA produced by a number of restriction enzymes were identical in size (data not shown). Also, the same fragment was found in three different isolates of *Photobacterium*. The hybridization of synthetic DNA probes complementary to each tRNA gene in the *argT* operon demonstrated the presence of a sequence highly similar to the anticodon loop of *E. coli* $tRNA^{Pro}$ (5′-CTGGTCCCAAACCAG-3′). Furthermore, the greater intensity of this hybridized band compared to that obtained with the probe for $tRNA^{His}$ suggested the presence of a repeated $tRNA^{Pro}$ gene.

```
   1 GATCTCCCAGTCAGCATGTTGGCTTTTTAAAGCCCGCTTGTTTTACAAGCACACAGAATT
  61 TAGTCAGTGATAACGTAGTTTTATAGCGTATCTCAGGCTTCTAGGATTAGCGGACCAGAT
 121 GGTTTGTTTTCTTAGTAACGATTTAGTCGGTGAATAGCGCAGTTTGGTAGCGCATCTGGT
 181 TTGGGACCAGAGGGTCGGGGGTTCGAATCCCTCTTCACCGACCACTTACAATGGTGGCTA
 241 TAGCTCAGTTGGTAGAGTCCCGGATTGTGATTCCGGTTGTCGCGAGTTCAAGCCTCGTTA
 301 GCCACCCCATTATTCAGGTTATCTTTTATTAGAGACCTTGGCCTTGTAAGCCTAAACATT
 361 GTCGGTGAATAGCGCAGTTTGGTAGCGCATCTGGTTTGGGACCAGAGGGTCGGGGGTTCG
 421 AATCCCTCTTCACCGACCACCATTAAGAAAACCGTATCGAAAGATACGGTTTTTTTTCGT
 481 CTGTAGAAAATGGTGAAGCGCGTAGTTTGGGGGCGCATCTCTGATGTTAGGAATAGCGGA
 541 CCAGAGGGTCGGGGGTTCCCTGATATTAGAACCTTGCTCTTCACCGACCACCATTAAGAA
 601 ACCTGAATCGAAAGATTGAGGTTTTTTTTTCGTCTGTATAAAAGTAAATAGGGGTCGTTGG
 661 TTTAAAGCCAGAGGGTCGGGGGTTCACTGATATTAGAATCTTGCTCTTCACCGACCACCA
 721 TTAAGAAACCTGAATTGAAAGATTGAGGTTTTTTTTTCGTCTGTAGAAAATGGTGAAGCGC
 781 GTAGTTTGGGGGCGCATCTCTGATGTTAGGAATAGCGGACCAGAGGGTCGGGGGTTCCCT
 841 GATATTAGAACCTTGCTCTTCACCGACCACCATTAAGAAACCTGAATCGAAAGATTCAGG
 901 TTTTTTTTTCATCTGTAGAAAAGTAAATCTGTATCATTGGTTTAAAGCTAGAGGGTCGGGG
 961 GTTCACCGATATTAGAATCTTGGCTCTTCACCGACCGCCATTAAGAAAACCGTATCGAAA
1021 GATACGGTTTTTTTTTCATCTATAGAAAATGGTGAAGCGCGTAGTTTGGGGGCGCATCTCT
1081 GATGTTAGGAATAGCGGACCAGAGGGTCGGGGGTTCCCTGATATTAGAACCTTGGCTCTT
1141 CACCGACCGCCATTAAGAAACTTGAATCGAAAGATACGGTTTTTTTTCGTCTGTAGAAAA
1201 TGGTGAAGCGCGTAGTTTGGGGGCGCATCTCTGATGTTAGGAATAGCGGACGAGAGGGTC
1261 GGGGGTTCCCTGATATTAGAACCTTGCTCTTCACTACCGCCATTAAGAAACCTGAATCGA
1321 AAGATTCAGGTTTTTTTTTCATCTGTAGAAAAGTAAATCTGTATCATTGGTTTAAAGCTAG
1381 AGGGTCGGGGGTTCACCGATATTAGAACCTTGGCTCTTCACCGACCACCATTAAGAAACC
1441 TGAATCGAAAGATTGAGGTTTTTTTTTCGTCTATAGAAAATGGTGAAGTGCGTAGTTTGGT
1501 AGTGCATCTCTGATGTTAGTAATAGCGGACCAGAGGGGCGGGGGTTCCCTGATATTAGAA
1561 TCTTGCTCTTCACCTATCACCATTAAGAAAACCGTATCGAAAGATTGAGGTTTTTTTTCG
1621 TCTGTCTTAAATGATGAGAAAACAGTTCGAATATCTGCGCAGATGTGTGCTGTAAATTTC
1681 ACCATTGCTGGTTGTAATAGCTATATCCACTGATGTTTATGTTTTTAGTAGTTATTGTCG
1741 CTATTTATTTTAATCTAAACGATATTTTGTTTAATTAATGTTAATTTAATTGGCGGTGGT
1801 TAGTGATTTTTTATTCATGATGGTGTTTTTAAGTGTTTATAGGTGTTTGGTTTATTGTAC
1861 GGTTTTTATTTTGCTTTTGTTGTTATTCAAATGGTTGTTGGTTTTGTTTAGTTTAGTTTA
1921 GTGTACTGTTGTTTTATTATGCGTGTATATGAAATAAATATGCGCAATATATAAATAATG
1981 ATAGTTTTTTACTTTGAGTTGTTGCTTTTTTGTAATATTTTTGACAGATTGAATGGAGAC
2041 ATGCAATCAGACAATAGTATTGGCATGGAAAGGATTCAATTTTTCAGAATGGCAGGATGC
2101 TGCCAGTAGTGGAGTCTGATAATGGCATTATTAGAAGTAAAAAATCTTCGTATAGAATAC
2161 CCCTCACGACATGGCGTACATGCAGCGGTGAAATCACTGTCATTCACCATTGAGCGTGGT
2221 GAAATTGTGGGTGTTGTTGGTGAGTCTGGAGCGGGTAAATCGACGGTGGGTAATGCTGTT
2281 ATTGACCTATTAAGTCCCCCTGGACAGATC
```

FIG. 1. DNA sequence of the 2310-bp *Sau*3A fragment isolated from *P. phosphoreum*. The tRNA genes are underlined with the anticodon boxed. The putative terminator sequences are indicated with arrows. Numbers 1 to 9 above the sequence indicate the first nucleotide of the nine repeated segments. H indicates the *Hae* III restriction site used to construct pKKH1.

```
1 CGGTGAATAGCGCAGTTTGGTAGCGCATCTGGTTTG-----------GGACCAGAGGGTCGGGGGTTC--------
2 TGGTGAAGCGCGTAGTTTGGGGGCGCATCTCTGATGTTAGGAATAGCGGACCAGAGGGTCGGGGGTTCCCTGATAT
3 -----------GTAAATAGGGGTCG---TTGGTTTA-----------AAGCCAGAGGGTCGGGGGTTCACTGATAT
4 TGGTGAAGCGCGTAGTTTGGGGGCGCATCTCTGATGTTAGGAATAGCGGACCAGAGGGTCGGGGGTTCCCTGATAT
5 -----------GTAAATCTGTATCA---TTGGTTTA-----------AAGCTAGAGGGTCGGGGGTTCACCGATAT
6 TGGTGAAGCGCGTAGTTTGGGGGCGCATCTCTGATGTTAGGAATAGCGGACCAGAGGGTCGGGGGTTCCCTGATAT
7 TGGTGAAGCGCGTAGTTTGGGGGCGCATCTCTGATGTTAGGAATAGCGGACGAGAGGGTCGGGGGTTCCCTGATAT
8 -----------GTAAATCTGTATCA---TTGGTTTA-----------AAGCTAGAGGGTCGGGGGTTCACCGATAT
9 TGGTGAAGTGCGTAGTTTGGTAGTGCATCTCTGATGTTAGTAATAGCGGACCAGAGGGGCGGGGGTTCCCTGATAT

1 --GAATCC---CTCTTCACCGACCACCATTAAGAAAACCGTATCGAAAGATAC-GGTTTTTTTTCGTCTGTAGAAAA
2 TAGAACCTT-GCTCTTCACCGACCACCATTAAGAAACCTGAATCGAAAGATTGAGGTTTTTTTTCGTCTGTATAAAA
3 TAGAATCTT-GCTCTTCACCGACCACCATTAAGAAACCTGAATTGAAAGATTGAGGTTTTTTTTCGTCTGTAGAAAA
4 TAGAACCTT-GCTCTTCACCGACCACCATTAAGAAACCTGAATCGAAAGATTCAGGTTTTTTTTCATCTGTAGAAAA
5 TAGAATCTTGGCTCTTCACCGACCGCCATTAAGAAAACCGTATCGAAAGATAC-GGTTTTTTTTCATCTATAGAAAA
6 TAGAACCTTGGCTCTTCACCGACCGCCATTAAGAAACTTGAATCGAAAGATAC-GGTTTTTTTTCGTCTGTAGAAAA
7 TAGAACCTT-GCTCTTCAC-TACCGCCATTAAGAAACCTGAATCGAAAGATTCAGGTTTTTTTTCATCTGTAGAAAA
8 TAGAACCTTGGCTCTTCACCGACCACCATTAAGAAACCTGAATCGAAAGATTGAGGTTTTTTTTCGTCTATAGAAAA
9 TAGAATCTT-GCTCTTCACCTATCACCATTAAGAAAACCGTATCGAAAGATTGAGGTTTTTTTTCGTCTGTCTTAAA
```

FIG. 2. Alignment of a tRNA^{Pro} gene and its adjacent terminator with the following eight pseudogenes. The sequence number refers to the numbering in Fig. 1.

The nucleotide sequence of the 2310-base-pair (bp) fragment was determined and is shown in Fig. 1. tRNA genes in the fragment were inferred from the ability of a region to fold into the canonical tRNA cloverleaf structure and the amino acid-charging specificity from the sequence of the putative anticodon. Three tRNA sequences in the order 5′-tRNA^{Pro}–tRNA^{His}–tRNA^{Pro}-3′, underlined in Fig. 1, were detected. The assignment of these genes is also strongly supported by the presence of characteristic nucleotides of the tRNA^{Pro} and tRNA^{His} families, respectively (18).

We can find no credible promoter in the 140-nucleotide (nt) segment 5′ to the first tRNA^{Pro} gene, which means that this DNA segment is probably linked to a promoter further upstream, because the tRNA sequences are so well conserved. We believe that the promoter sequences in *Photobacterium* would be similar to those in *E. coli*, since even *Bacillus subtilis*, a distantly related Gram-positive bacterium, has a promoter sequence (in its vegetative phase) that is almost identical to that of *E. coli* (19, 20). Downstream from the three tRNA genes, a region showing dyad symmetry (indicated with arrows in Fig. 1) followed by a run of eight thymines strongly resembles a *rho*-independent terminator (21). This same symmetrical feature is subsequently repeated eight times. Interspersed among the terminators are five 118-nt and three 93-nt segments that show a very high similarity among themselves and, surprisingly, with the preceding tRNA (see Fig. 2). All eight segments contain a tRNA CCA terminus or a close relative thereof (CCG); this is also found in the tRNA^{Pro} and tRNA^{His} genes, but not in all tRNA genes (20).

The 118-nt sequence can be folded into a cloverleaf structure (Fig. 3) with an insertion in the "tRNA" anticodon loop and another in the T loop. The folding pattern of the 93-nt element is less tRNA-like due to a deletion of some of the 5′ region of the tRNA^{Pro} gene. Nevertheless, the sequence is clearly related to the tRNA and shares the T-loop insertion of the 118-nt segment (Fig. 3). Although there is no insertion in the anticodon loop of the shorter element, the nucleotides in the loop have been mutated from TTTGGGA to TTTAAAG. The existence of these insertions and mutations (transitions) would most likely render inactive the transcription product of these elements, in line with the hypothesis that they are bacterial tRNA pseudogenes.

The anticodon-loop insertion CTGATGTTAGGAATAGC and the T-loop insertion CTGATATTAGAACCTTG are clearly related to each other and possess the consensus sequence of TGATKW, where K represents thymine or guanine and W represents thymine or adenine, found in the inverted terminal repeats of bacterial transposons (22). This latter similarity suggests a possible transposon role in the origin of the repetitive sequences, although subsequent deletion of the transposon does not normally leave a portion of the inverted repeat in the genome. Similarity of the above also extends to the consensus TGATG stem of the REP or P.U. (repetitive extragenic palindromes) found in *E. coli* and *Salmonella* (23, 24).

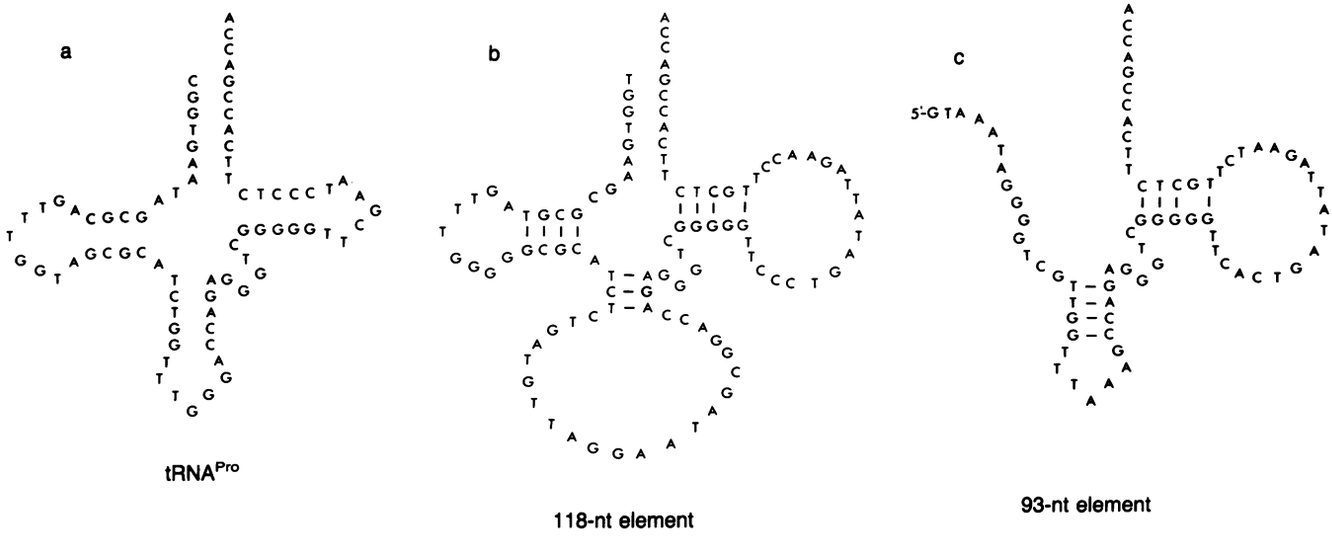

FIG. 3. Hypothetical secondary structure of tRNA^{Pro} (*a*), the 118-nt pseudoelement (*b*), and the 93-nt pseudoelement (*c*).

Expression of the tRNA Cluster. The lack of a promoter in the sequenced fragment was confirmed by an *in vitro* transcription assay. Referring to Fig. 4, lane 1 shows the production of the expected major transcript of about 480 nt from transcription of the *argT* operon of *E. coli*. Lane 4, on the other hand, shows that no major transcript is made from the *Photobacterium* fragment of pPPS70. Major transcripts of about 180 and 520 nt were seen by use of a template composed of either the *Hae* III fragment (pKKH1, lane 2) or the *Sau*3A fragment (pKKS4, lane 3) under control of the *tac* promoter. The doublet around 180 nt indicates that the 3′ terminus of these transcripts is heterogeneous (21). Transcription terminates at the first putative *rho*-independent terminator in the segment, and this suggests that the $tRNA^{Pro}$ pseudogenes are not expressed *in vivo*, whether or not the tRNA genes themselves are expressed.

Subsequently, a Northern (RNA) blot was used to confirm the lack of pseudogene expression *in vivo*. Hybridization of total RNA extracted from both *Photobacterium* and *E. coli* transformed with pPPS70, pKKH1, and pKKS4 revealed a single band comigrating with a purified *E. coli* $tRNA_3^{Pro}$ (Fig. 5, lanes 2, 5, 6, and 7). RNA from *E. coli* transformed with pBR327 or pBE1935 as well as the $tRNA_3^{Pro}$ standard did not cross-hybridize with the *Photobacterium* probe (Fig. 5, lanes 1, 3, and 4). Although pPPS70 has no promoter, *E. coli* transformed with this plasmid could produce $tRNA^{Pro}$, albeit to a lesser extent by the use of the upstream tet^r gene promoter present in pBR327.

Phylogenetic Analysis of the Cluster. The tandem repeats reported here could have resulted from one or a variation of two possible mechanisms—*i.e.*, a one-event, nine-fold repetition of the original element or several duplication events of a single element. Therefore, to relate the individual copies of the repeated DNA segments, the nine aligned sequences of Fig. 2 were submitted to phylogenetic analysis using an algorithm based on maximum parsimony (25). The resulting tree is shown in Fig. 6, and from this topology, a chronology

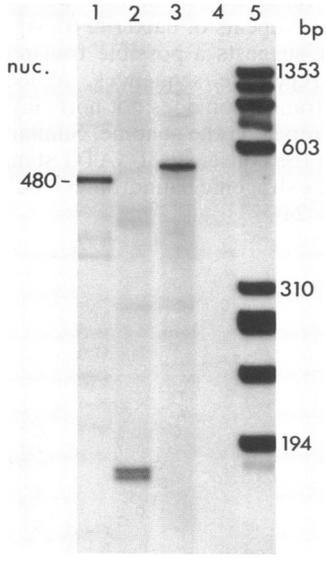

FIG. 4. *In vitro* transcription assay using purified restriction fragments as template. Lanes: 1, the transcription of the *E. coli argT* restriction fragment; 2, products from the transcription of the *Bam*HI restriction fragment from pKKH1 containing the last $tRNA^{Pro}$ followed by the nine repetitions fused to the *tac* promoter; 3, the transcription of the *Bam*HI restriction fragment from pKKS4 containing the entire *Sau*3A fragment fused to the *tac* promoter; 4, transcription of the *Sau*3A restriction fragment from pPPS70; 5, ϕx-174 restriction fragment DNA digested with *Hae* III and end-labeled with [γ-^{32}P]ATP and T4 kinase as marker.

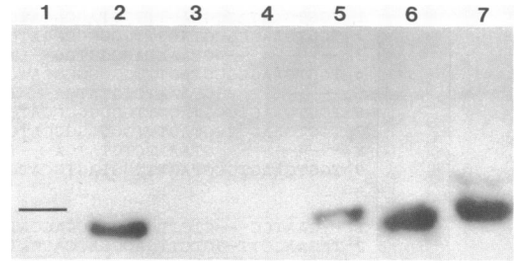

FIG. 5. Northern blot using RNA species isolated from different sources hybridized with the ^{32}P-labeled 2.0-kb repetitive segment from *Hae* III-restricted pPPS70. Line indicates position of band stained by ethidium bromide. Lanes: 1, 3 μg of purified $tRNA_3^{Pro}$ from *E. coli*; 2, one OD_{260} unit of RNA extracted from *P. phosphoreum*; 3, one OD_{260} unit of RNA extracted from *E. coli* transformed with pBR327; 4, one OD_{260} unit of RNA extracted from *E. coli* transformed with pBE1935 (*argT*-containing plasmid); 5, one OD_{260} unit of RNA extracted from *E. coli* transformed with pPPS70; 6, one OD_{260} unit of RNA extracted from *E. coli* transformed with pKKH1; 7, one OD_{260} unit of RNA extracted from *E. coli* transformed with pKKS4.

of events leading to the emergence of the repeated segments can be inferred. As a first step, the $tRNA^{Pro}$ gene would have been duplicated along with its adjacent terminator sequence. Next, the 17-nt segment would have been inserted in the T loop of the downstream gene copy; this lengthened gene must be the progenitor of all subsequent repeated segments. Next, duplication of the progenitor segment followed by an insertion in the anticodon loop of one copy and the deletion of part of the 5′ region of the other would have occurred. The resulting elements were then the progenitors for the two classes of pseudogenes. Although the sequence of the subsequent events cannot be determined with precision, it can be concluded that most, if not all, events leading to the tandem repeats involved the duplication of single elements.

DISCUSSION

Although simple duplications of the type reported here arise at relatively high rates (10^{-4}) in eubacterial genomes, the reversion is equally high (26). Occasionally a higher gene dose can give a selective advantage to a cell resulting in the maintenance of repetitive DNA. Such may be the case for the few known tandem repeats in bacteria: (*i*) the 178-nt distal region of the *E. coli* $tRNA^{Tyr}$ gene is repeated 3.14 times (27); (*ii*) the 3′-flanking region of *E. coli* and *Salmonella* M1 RNA component of RNase P is repeated 3.5 times (16); (*iii*) the operon *leuV* from *E. coli* contains three tandem leucine tRNA genes with exactly the same sequence (28) and (*iv*) uncharacterized, reiterated DNA is reported in purple bacteria of the α subdivision (29). Maintenance of the repetitive DNA in the first two examples may be related to their potential open reading frame (16), whereas the repeats in the third case may be a direct result of the abundant use of codons translated by the encoded $tRNA^{Leu}$. However, unlike these tandem dupli-

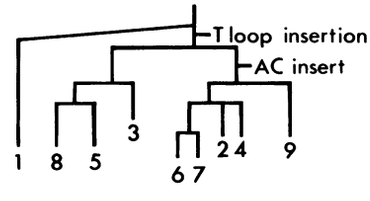

FIG. 6. Phylogenetic tree relating the nine sequences of Fig. 2. The tree was constructed from a maximum-parsimony algorithm. Sequence 1 is the $tRNA^{Pro}$ gene sequence with its adjacent terminator.

cations, the DNA fragment studied here has no neighboring protein coding region; nor can it produce functional $tRNA^{Pro}$ from the downstream "mutant" genes.

Pseudogenes are thought to arise in the chromosome by either the accumulation of deleterious mutations in a gene copy or the insertion of a retrotranscribed copy of a cytoplasmic RNA molecule. Highly repeated retropseudogenes related to tRNAs are prominent in many eukaryotic, particularly animal, genomes, where they are dispersed and are characterized by transcriptional activity (30). Pseudogenes in bacteria have few precedents: a single-copy pseudo $tRNA^{Ile}$ gene has been found in the 16S–23S ribosomal RNA spacer region of *Euglena gracilis* chloroplast DNA (31). However, the endosymbiotic chloroplastic genome is hardly a standard for eubacteria, even though a common origin is well documented (25). Thus, the repeated tRNA pseudogenes that we have observed are unusual and represent another class of eubacterial pseudogenes.

Finally, the phylogenetic distribution of the repeated element, though absent in *E. coli*, as determined by hybridization (data not shown), has not yet been determined. The number of inferred mutations (mainly deletions and insertions) among different copies, however, would support at least a moderate distribution among related organisms, implying that selection must somehow favor repetitive sequences. This, in turn, emphasizes the problem of rationalizing its apparent evolutionary conservation, because bacterial genomes are thought not to contain nonfunctional DNA (32). Whether these sequences are involved in maintaining a particular, intrinsic DNA structure or in binding cellular proteins is unknown at the present time.

We thank W. F. Doolittle for suggestions on the manuscript, M. J. Fournier for donation of pLC25-25 containing the *argT* cluster, and Emmanuelle Lemyre for technical help. This work was supported by the National Science and Engineering Research Council of Canada. S.G. holds a Medical Research Council Doctoral Studentship and R.C. is a Fellow of the Canadian Institute of Advanced Research.

1. Woese, C. R. (1987) *Microb. Rev.* **51,** 221–271.
2. Field, K. G., Olsen, G. J., Lane, D. J., Giovannoni, S. J., Ghiselin, M. T., Raff, E. C., Pace, N. R. & Raff, R. A. (1988) *Science* **239,** 748–753.
3. Hori, H. & Osawa, S. (1987) *Mol. Biol. Evol.* **4,** 445–472.
4. Borst, P. & Greaves, D. R. (1987) *Science* **235,** 658–667.
5. Shapiro, J. A. (1985) *Int. Rev. Cytol.* **93,** 25–56.
6. Cedergren, R. J. & Lang, B. F. (1985) *Biosystems* **18,** 263–267.
7. Hsu, L. M., Klee, H. J., Zagorski, J. & Fournier, M. J. (1984) *J. Bacteriol.* **158,** 934–942.
8. Woese, C. R., Weisburg, W. G., Hahn, C. M., Paster, B. J., Zablen, L. B., Lewis, B. J., Macke, T. J., Ludwig, W. & Stackebrandt, E. (1985) *Syst. Appl. Microbiol.* **6,** 25–33.
9. Macdonell, M. T., Swartz, D. G., Ortiz-Conde, B. A., Last, G. A. & Colwell, R. R. (1986) *Microbiol. Sci.* **3,** 172–178.
10. Bolivar, F., Rodriguez, R. L., Green, P. J., Betlach, M. C., Heyneker, H. L., Boyer, H. W., Crosa, J. H. & Falkow, S. (1977) *Gene* **2,** 95–113.
11. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
12. Silhavy, T. J., Berman, M. L. & Enquist, L. W. (1984) *Experiments in Gene Fusions* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
13. Birnboim, H. C. & Doly, J. (1979) *Nucleic Acids Res.* **7,** 1513–1523.
14. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
15. de Boer, H. A., Comstock, L. J. & Vasser, M. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 21–25.
16. Reed, R. E. & Altman, S. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 5359–5363.
17. Reed, R. E., Baer, M., Guerrier-Takada, C., Donnis-Keller, H. & Altman, S. (1982) *Cell* **30,** 627–636.
18. Nicoghosian, K., Bigras, M., Sankoff, D. & Cedergren, R. J. (1987) *J. Mol. Evol.* **26,** 341–346.
19. Johnson, W. C., Morau, C. P. & Losick, R. (1983) *Nature (London)* **302,** 800–804.
20. Vold, B. S. (1985) *Microb. Rev.* **49,** 71–80.
21. Platt, T. (1986) *Annu. Rev. Biochem.* **55,** 339–372.
22. Kleckner, N. (1981) *Annu. Rev. Genet.* **15,** 341–404.
23. Stern, M. J., Ferro-Luzzi Ames, G., Smith, N. H., Robinson, N. C. & Higgins, C. F. (1984) *Cell* **37,** 1015–1026.
24. Gilson, E., Clément, J. M., Britlag, D. & Hofnung, M. (1984) *EMBO J.* **3,** 1417–1421.
25. Cedergren, R., Gray, M. W., Abel, Y. & Sankoff, D. (1988) *J. Mol. Evol.*, in press.
26. Anderson, R. P. & Roth, J. R. (1977) *Annu. Rev. Microbiol.* **31,** 473–505.
27. Egan, J. & Landy, A. (1978) *J. Biol. Chem.* **253,** 3607–3622.
28. Duester, G., Campen, R. K. & Holmes, W. M. (1981) *Nucleic Acids Res.* **9,** 2121–2139.
29. Flores, M., Gonzales, V., Brom, S., Martinez, E., Pinero, D., Romero, D., Davila, G. & Palacios, R. (1987) *J. Bacteriol.* **169,** 5782–5788.
30. Weiner, A. M., Deininger, P. L. & Efstratiadis, A. (1986) *Annu. Rev. Biochem.* **55,** 631–661.
31. Orozco, E. M., Rushlow, K. E., Dodd, J. R. & Hallick, R. B. (1980) *J. Biol. Chem.* **255,** 10997–11003.
32. Doolittle, W. F. & Sapienza, C. (1980) *Nature (London)* **284,** 601–603.