Volume 11 Number 24 1983 Nucleic Acids Research

Structure and evolution of a mouse tRNA gene cluster encoding $tRNA^{Asp}$, $tRNA^{Gly}$ and $tRNA^{Glu}$ and an unlinked, solitary gene encoding $tRNA^{Asp}$

James E.Looney* and John D.Harding+

922 Fairchild Center, Department of Biological Sciences, Columbia University, New York, NY 10027, USA

Received 17 October 1983; Revised and Accepted 14 November 1983

ABSTRACT

We have sequenced mouse tRNA genes from two recombinant λ phage. An 1800 bp sequence from one phage contains 3 tRNA genes, potentially encoding $tRNA^{Asp}$, $tRNA^{Gly}$, and $tRNA^{Glu}$, separated by spacer sequences of 587 bp and 436 bp, respectively. The mouse tRNA gene cluster is homologous to a rat sequence (Sekiya et al., 1981, Nucleic Acids Res. 9, 2239-2250). The mouse and rat $tRNA^{Asp}$ and $tRNA^{Gly}$ coding regions are identical. The $tRNA^{Glu}$ coding regions differ at two positions. The flanking sequences contain 3 non-homologous areas: a c. 100 bp insertion in the first mouse spacer, short tandemly repeated sequences in the second spacers and unrelated sequences at the 3' ends of the clusters. In contrast, most of the flanking regions are homologous, consisting of strings of consecutive, identical residues (5-17 bp) separated by single base differences and short insertions/deletions. The latter are often associated with short repeats. The homology of the flanking regions is c. 75%, similar to other murine genes.

The second λ clone contains a solitary mouse $tRNA^{Asp}$ gene. The coding region is identical to that of the clustered $tRNA^{Asp}$ gene. The 5' flanking regions of the two genes contain homologous areas (10-25 bp) separated by unrelated sequences. Overall, the flanking regions of the two mouse $tRNA^{Asp}$ genes are less homologous than those of the mouse and rat clusters.

INTRODUCTION

Analysis of DNA sequences indicates that mammalian tRNA genes are arranged in two different configurations in the genome. Some genes are present in clusters in which individual tRNA coding regions are separated by c. 100-500 base pairs of DNA (1-3). In contrast, several human $tRNA_i^{Met}$ genes (4) and a mouse $tRNA^{His}$ gene (5) are apparently not closely associated with other tRNA genes. These two types of tRNA gene organization are also found in other eucaryotes. Many Drosophila tRNA genes are clustered similarly to the mammalian examples (e.g. 6-10) and some are solitary (11). In yeast, most tRNA genes appear to be solitary (reviewed in Ref. 12).

The differences in clustering of mammalian tRNA genes as well as the existence of multiple copies of genes coding for the same tRNA raise many questions concerning tRNA gene structure, regulation and evolution which are

currently difficult to answer due to the relative paucity of sequence data. A fundamental issue is the extent to which the sequences of tRNA coding and flanking regions are conserved, both within a gene family in a particular mammalian species and among different species.

We have recently isolated several clones containing tRNA genes from a mouse recombinant DNA library in order to analyze in detail the sequence organization of mammalian tRNA gene families (5,13). In this paper we present DNA sequences from two different mouse genomic DNA clones. One sequence contains a cluster of genes coding for $tRNA^{Asp}$, $tRNA^{Gly}$, and $tRNA^{Glu}$, respectively, which is related to the rat clusters sequenced by Sekiya et al. (1) and Shibuya et al. (14). This makes possible the first comparison, to our knowledge, of cognate tRNA gene clusters from closely related mammals. The second clone contains a solitary $tRNA^{Asp}$ gene, the sequence of which we compare to that of the clustered $tRNA^{Asp}$ gene.

MATERIALS AND METHODS

Isolation of Mouse tRNA Genes

A DBA-2 mouse genomic DNA-λCharon 4A recombinant library was prepared as described in Ref. 5. Phage (c. 6,000-10,000 per plate) were spread on 100 mm square petri dishes (Lab Teck 4021) using 0.3 ml/plate of a concentrated culture of E. coli DP50 SupF and incubated overnight at 37°. The phage plaques were adsorbed to nitrocellulose filters saturated with E. coli DP50 SupF as described by Woo (15). The filters were treated as described by Woo, (except that the neutralization step with 1M Tris-HCl was omitted) and baked in vacuo at 80° for 2 hrs. The baked filters were hybridized with 2 x 10^7 cpm of iodinated mouse liver tRNA (2-5 x 10^7 cpm/ug), washed and autoradiographed essentially as described in Ref. 5. Autoradiographs of duplicate filters were compared to distinguish bona fide signals from background spots. Phage plaques were picked and successively replated until more than 90% of the plaques hybridized with the tRNA probe.

DNA Sequencing

A plaque purified λ clone or a restriction fragment derived from it was digested with a restriction endonuclease and the resulting fragments subcloned into phage M13mp9, as described in Ref. 5. Blunt end fragments were cloned into the SmaI site, NarI and HpaII fragments into the AccI site and Sau3A fragments into the BamHI site of M13mp9, respectively.

In some cases, (see Results) M13 subclones were screened for the presence of tRNA coding regions with iodinated tRNA probe as described above for λ

phage. Use of the Woo procedure (15) allows M13 phage particles containing either the coding (tRNA-like) or non-coding DNA strands to be detected with a tRNA probe. The two DNA orientations can be distinguished by the intensity of the hybridization signal, as described in detail elsewhere (J. Looney, J. Han and J. Harding, submitted).

DNA inserts in M13 subclones were sequenced by the dideoxy chain termination procedure (16).

RESULTS

Isolation and Sequence Analysis of Mouse tRNA Genes

We screened c. 100,000 plaques from a DBA-2 mouse DNA-λ Charon 4A recombinant library with iodinated tRNA. The first screening step yielded more than 50 areas on the phage plates that hybridized with the tRNA probe, from 25 of which we plaque-purified individual recombinant phage. DNA was isolated from 21 cultures of individual plaque-purified phage and initially characterized by EcoRI digestion followed by Southern blotting (17) with iodinated tRNA. Two of the clones which gave strong hybridization signals, termed λMt2 and λMt4, were characterized by DNA sequence analysis.

λMt2 contains 3 EcoRI mouse DNA fragments c. 7, 4.1 and 0.5 kb in size. Iodinated tRNA hybridized to the 4.1 kb EcoRI fragment only. DNA sequencing (see below) identifies 3 tRNA genes in this fragment. Other Southern blotting experiments (not shown) suggest that the 4.1 kb fragment also contains at least one other tRNA gene.

The sequence of an 1800 bp region of the 4.1 kb EcoRI fragment of λMt2 was obtained from 32 M13 subclones as shown in Fig. 1A. The sequence of both DNA strands was determined, except for c. 200 bp at each end, which were determined on one strand. The latter sequences were obtained at least twice from independently isolated M13 subclones.

λMt4 contains a single 19 kb EcoRI mouse DNA fragment. A 639 bp sequence was determined from 5 M13 subclones, as shown in Fig. 1B. 50% of the sequence was determined on both DNA strands. The remainder of the sequence was determined on one strand from several independent sequencing reactions. The 639 bp sequence contains a single tRNA gene (see below) which is present on a 306 bp AluI fragment and a 350 bp Sau3A fragment. Southern blot analysis of λMt4 DNA indicates that single AluI and Sau3A fragments of these expected sizes hybridize with iodinated tRNA. We therefore conclude that λMt4 probably contains a single tRNA gene.

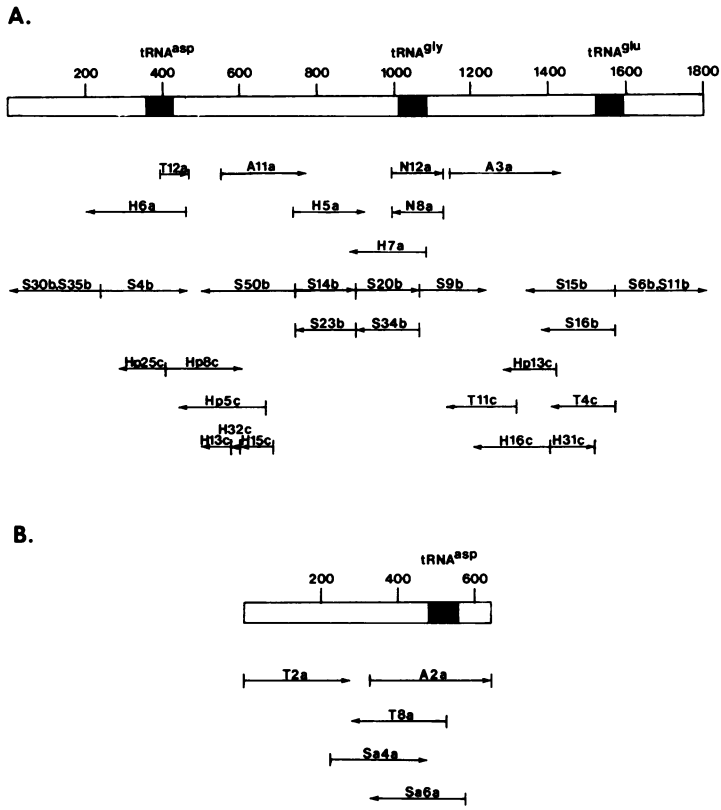

Figure 1. Sequencing Strategies - λMt2 and λMt4.

Part A. The bar represents an 1800 bp sequence present in phage λMt2 which contains 3 tRNA genes, indicated by the filled in areas. The 5' end of the sequence, relative to the non-coding (tRNA-like) DNA strand is to the left.

The M13 subclones used to determine the 1800 bp sequence are shown below the bar. The beginning of the mouse DNA insert relative to the M13 sequencing primer is indicated by a vertical line; the arrowhead indicates the end of the sequence read from the gel. An arrowhead followed by a vertical line indicates that the entire mouse DNA insert of the M13 subclone was sequenced. Arrows pointing to the right indicate that DNA corresponding to the coding strand was sequenced; arrows pointing to the left indicate that the non-coding strand was sequenced.

Part B. The bar represents a 639 bp sequence in phage λMt4 which contains a $tRNA^{Asp}$ gene, indicated by the filled in area. M13 subclones used to determine the sequence are indicated below the bar.

The first letter of each M13 subclone designation in Parts A and B indicates the restriction digest of the lambda phage from which it was isolated. A, AluI; H, HaeIII; Hp, HpaII; N, NarI; S, SmaI; Sa, SauIIIA; T, ThaI. The last letter of each designation indicates the method by which the subclone was isolated. The suffix a denotes clones isolated by plaque screening with labelled tRNA; b denotes clones derived from a SmaI digest of λMt2, picked at random and sequenced; c denotes clones derived from two purified SmaI fragments of λMt2, picked at random and sequenced.

Nucleotide Sequence of the 1800 bp Fragment of λMt2

The nucleotide sequence of the 1800 bp fragment of λMt2 is shown in Fig. 2 (upper case letters). We searched the sequence for GTTC and GATC tetranucleotides, which are diagnostic for tRNA coding regions, and constructed tRNA-like cloverleaf structures from the surrounding nucleotides. As indicated in Fig. 1A and Fig. 2, the 1800 bp sequence contains 3 genes which can encode $tRNA^{Asp}$ (positions 356-427), $tRNA^{Gly}$ (positions 1015-1086) and $tRNA^{Glu}$ (positions 1523-1594), respectively. The 3 genes are present in the same polarity and are separated by spacer regions of 587 and 436 bp, respectively.

The 3 tDNA cloverleaf structures encoded in the 1800 bp sequence are shown in Fig. 3. Each tDNA exhibits a typical tRNA-like stem-loop configuration. All of the residues present at potentially invariant or semi-variant positions are characteristic of sequenced eucaryotic cytoplasmic tRNAs (18). As is typical of eucaryotic tRNA genes, the 3' terminal CCA residues found in mature tRNAs are not encoded in the DNA.

The tDNA structures are not precisely homologous to any sequenced tRNA. The $tDNA^{Asp}$ structure (Fig. 3A) differs at two positions from sequenced rat liver $tRNA^{Asp}$ (19). At position 32 in the anticodon loop the tDNA contains a C residue and the sequenced tRNA, a U residue. At position 33, the tDNA contains a T residue and the tRNA, a C residue. Sequenced bovine and rabbit liver $tRNA^{Asp}$ contain the same residues at positions 32 and 33 as the mouse $tDNA^{Asp}$ sequence, although they differ at some other positions (20, 21). To our knowledge, mammalian $tRNA^{Gly}$ and $tRNA^{Glu}$ species having the same anticodons as the mouse tDNAs have not been sequenced. The $tDNA^{Gly}$ sequence is 93% homologous to B. mori $tRNA^{Gly}$ (22).

As shown in Fig. 2, the 3' flanking regions of the $tRNA^{Asp}$ and $tRNA^{Gly}$ genes contain likely transcription termination sites (comprised of 4 or more consecutive T residues on the non-coding DNA strand (23) 12 bp from the end of the coding region. The $tRNA^{Glu}$ gene contains 4 consecutive T residues 63 bp from the end of the coding region.

Sequences outside the tRNA coding regions are ratherly uniformly G-C rich (average of 68% GC for the entire sequence). They contain numerous short tandem and non-tandem repeats. Computerized dot-matrix analysis of the sequence does not, however, reveal any predominant repeated sequence. The only long segment of tandemly repeated sequence occurs between residues 1321 and 1362, in which the heptanucleotide AACGCCC occurs 6 times, with a one base-pair mismatch in the last repeat.


tRNA asp
tRNA gly

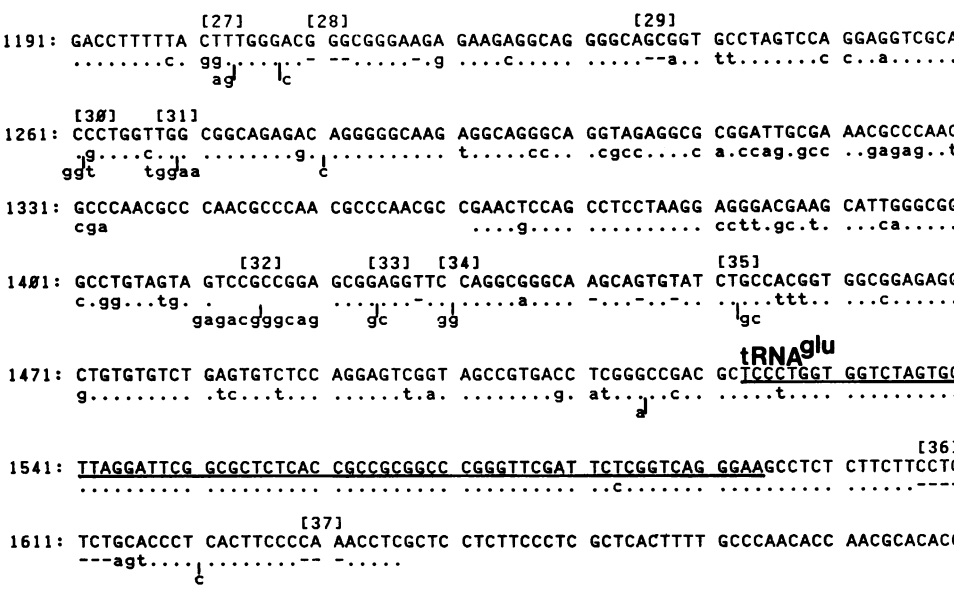

Figure 2. Sequence of the tRNA Gene Cluster in λMt2.

The non-coding strand of the 1800 bp segment of mouse DNA is given in upper case letters with the 5' end to the left. The three tRNA coding regions are underlined. Each tRNA gene would be transcribed from left to right. Homologies with the cognate rat sequence determined by Sekiya et al. (1) are indicated beneath the mouse sequence. Dots correspond to identical nucleotides in the rat and mouse sequences. Dashes represent deletions in the rat sequence and lower case letters represent mismatched nucleotides or insertions in the rat sequence. The rat sequence of Sekiya et al. (1) begins at the residue homologous to residue 13 of the mouse sequence and extends beyond the last homologous residue shown (position 1636). The rat sequence shown in the figure is contiguous to that of Sekiya et al. (1) except for 76 bp (between the residues homologous to mouse positions 571 and 745, respectively) which are not homologous to mouse sequences and are not included.

The numbers in brackets mark insertions or deletions of 2 or more bp in the mouse and rat sequence comparison. These are categorized in respect to their association with short repeats in Table I.

Comparison of the 1800 bp Sequence of λMt2 With a Cognate Rat Sequence

Sekiya et al. (1) have published the sequence of a rat tRNA gene cluster which is similar to the mouse sequence reported here and Shibuya et al. (14) have published the sequences surrounding the tRNA genes in 4 other related rat gene clusters. Both the mouse and rat clusters contain the $tRNA^{Asp}$, $tRNA^{Gly}$ and $tRNA^{Glu}$ genes in the same orientation. They differ on a gross level primarily in the length of the spacer sequence between the $tRNA^{Asp}$ and $tRNA^{Gly}$

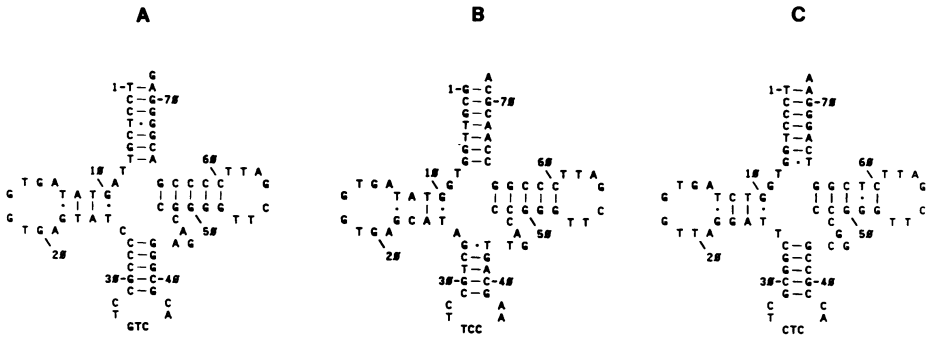

Figure 3. The Cloverleaf Structures of the tDNAs Encoded by the 1800 bp Sequence of λMt2.

PartA. The $tDNA^{Asp}$ encoded by residues 356-427 in Fig. 2 and residues 485-557 in Fig. 4.

Part B. The $tDNA^{Gly}$ encoded by residues 1015-1086 in Fig. 2.

Part C. The $tDNA^{Glu}$ encoded by residues 1523-1594 in Fig. 2.

genes, 587 bp in the mouse and 464 bp in the rat. The spacers between the $tRNA^{Gly}$ and $tRNA^{Glu}$ genes are about the same length (436 bp in the mouse, 429 bp in the rat).

A direct comparison of the mouse sequence reported here and the rat sequence determined by Sekiya et al. (1) is shown in Fig. 2. The $tRNA^{Asp}$ and $tRNA^{Gly}$ coding regions are identical in the mouse and rat clusters. The $tRNA^{Glu}$ coding regions differ at two positions. The mouse gene contains a C at position 1526 (position 4 in the aminoacyl stem of the $tDNA^{Glu}$ structure in Fig. 3C) whereas the rat gene contains a T at this position. As in the mouse sequence, the four $tRNA^{Glu}$ genes sequenced by Shibuya et al. (14) contain a C at this position. The other coding region change occurs at position 1583 in Fig. 2 (position 61 in the T-stem of the $tDNA^{Glu}$ structure) where the mouse DNA contains a T residue and the rat DNA a C residue in all 5 published sequences (1,14). We note that neither of these differences would markedly affect the base pairing of the stems of the mouse or rat tDNAs. The mouse aminoacyl stem contains a C-G base pair whereas the rat stem contains a T-G base pair (U-G in the RNA), which is fairly common in tRNAs (18). The opposite occurs in the case of the T-stems (T-G base pair in the mouse, C-G in the rat).

The sequence comparison shown in Fig. 2 also indicates that the regions outside the tRNA coding regions of the mouse and rat gene clusters are homologous. The sequences are comprised of strings of consecutive, homologous

nucleotides interrupted by single base differences and insertions/deletions of single base pairs or short strings of nucleotides. The mouse and rat sequences contain 59 strings of 5 or more consecutive homologous base pairs. If each insertion/deletion is scored as being equivalent to a single point mutation, we calculate that the mouse sequence is 75% homologous to the rat sequence. This calculation involves 1145 bp of the mouse sequence, not including the tRNA coding regions (which are almost 100% homologous, see above) and three regions which cannot be aligned with the rat sequence (see below).

When making the sequence comparison in Fig. 2, we considered a large number of possibilities by inspection with some help from computer analysis. Whenever possible, we minimized the number of postulated insertions and deletions. In some cases, minor variations of the alignment are possible, for instance in the precise positioning of small deletions/insertions or in scoring two or more consecutive point mutations as an insertion in both DNAs. These potential variations in the alignment do not significantly affect the calculated sequence homology.

Postulated insertions/deletions of 2 or more base pairs are marked in Fig. 2 by a number in brackets above the mouse sequence. The postulated insertions/deletions are often associated with short repeats, as has also been found in other eucaryotic genes (see Ref. 24 and refs. within) and can be divided into 7 categories as summarized in Table I. Most of the insertion/deletions are associated with either tandem repeats or closely associated non-tandem repeats (Categories 1 and 2, Table I). In some cases, as in example [4] (see Table I and Fig. 2) the repeats are identical. In other cases, as example [12], one of the repeats contains a point mutation. It is also necessary to postulate in some cases that the conserved repeats in mouse and ratDNA contain point mutations. For instance, in example [4], the mouse DNA contains the sequence CAGTCG which we postulate is homologous to the rat sequence, cactcg. Another case is seen in example [1], in which the second TCCC repeat in the mouse DNA (the first repeat is absent in the rat DNA) has also been mutated in the rat DNA, perhaps by inversion of TCC to cct. A further point about Table I is that a given Example can sometimes be placed in two different categories if a point mutation has been assumed to have occurred. For instance, Example [9], which we place in category 3 in a non-mutated form can also be placed in category 1 if it is assumed that CGGT and CGGG are two tandem repeats, one of which has been mutated in the fourth nucleotide.

Categories 1-3 and 7 in Table I have also been documented in other eu-

TABLE I: INSERTIONS AND DELETIONS IN THE MOUSE AND RAT tRNA GENE CLUSTERS

	Category	Cases[a]	Examples[b]
1.	Tandem Repeat Deleted/Inserted	4,7,12*,16*, 17,19,20*,21, 23,25,30*,35	[4]: CAGTCG------TCA cactcgcactcggca [12]: TTTGGAAAGACA tttg----gaca
2.	Non-Tandem Repeat Deleted/Inserted	1,18,22*,27, 28,29,33,34,36	[1]: CCCCTCCCGGTCCCG cccc----ggcctgc
3.	Non-Tandem Repeat and Intervening Residues Deleted/Inserted	5*,9	[9]: AGACGGTCGGG aga----cggg
4.	Residues Between Non-Tandem Repeats Deleted/Inserted	15,26,31*	[15]: TGCGCTAAGCGCAGGCCAGCGC ag-gc----------gcgc
5.	Insertions in Both Sequences, Between Non-Tandem Repeats	3,11,32	[3]: CCC GCAG CCCAGCAG cccctgtctgtcccca-cag
6.	Insertions in Both Sequences	2,8,13,14	[8]: GGT CCCCCAGGG CTG ggtggtctggcggaactg
7.	Uncertain	6,10,24,37	[6]: GGGTGGCAACGGTC gg------acgggc

[a] Cases are marked by bracketed numbers in Fig. 2. * indicates that one of the repeats contains a point mutation.

[b] The mouse sequence from Fig. 2 is in upper case, the rat sequence, in lower case. Repeats are underlined.

caryotic genes (24). In addition, we find three other categories of insertions/deletions, numbered 4-6 in Table I. These regions also often contain small repeats. For instance, in Example [15], the sequence GCGC is found within the mouse sequence which is not present in the rat DNA and is also repeated in the common flanking residues.

In contrast to the homologous regions of the mouse and rat sequences, 3 regions are so different that no convincing alignment is possible. These are: 1). The region between positons 571 and 744 of the mouse sequence. As noted above, the spacer sequence which contains these residues is longer in the mouse cluster than in the rat, which explains most of the gap in homology.

However, 76 bp of the rat sequence cannot be aligned in this region (and are not shown in Fig. 2). 2). The region between positions 1310 and 1361 of the mouse sequence. These residues mainly comprise the tandemly repeated sequence AACGCCC, as noted above. In the rat DNA, this region contains the sequence, GCCAGG, which is repeated 5 times. 3). The region between positions 1636 and 1800 of the mouse sequence. This region of the rat DNA (not shown in Fig. 2) clearly shares little homology with the mouse sequence. It contains 26 tandemly repeated CT dinucleotides which are not present in the mouse DNA.

Nucleotide Sequence of the $tRNA^{Asp}$ Gene Present in λMt4

The 639 bp sequence derived from λMt4 is shown in Fig. 4. It contains a single $tRNA^{Asp}$ gene, which, as discussed above, is probably the only tRNA gene in λMt4. The $tRNA^{Asp}$ coding region is identical to that present in clone λMt2. A probable transcription termination site, consisting of 6 consecutive T residues, is located 13 bp from the 3' end of the coding region.

Comparison of the sequences flanking the $tRNA^{Asp}$ coding regions in λMt4 and λMt2 indicates the presence of 6 homologous regions as indicated in Fig. 4.

```
  1: CGCGTATTGA CCAGGTGTTC AACGGCAAAT TTCTGCCCTT CTGATGTCAG AAAGGTAAAG TGATTTTCTT

 71: TCTGGTATTC AGTTGCTGTG TGTCTGGTTT CAGCAAAACC AAGCTCGAAG GCATTAACGC CTCTGCCCGT
                                                    (1) ..t.g..... .aga....c.

141: TACCCGAAAA ATTGGGTGAC CACCGGCGTA CCGGCGAGCC CGAGAGCGGA GGGAAGCTCA GAGGGTAGAT
     --.... (24)

211: GGCGATCAGG CTAGCGCTAG ACAGCATGGG ATTGGGGGGA GGTGCGGGGA GGAACCGAGA GTGTCTGGGG
                                          (143) .. ..--...g-- ..c.. (155)

281: CAGGGGTAGG CTGGGTTACG GGAAAACGCT GAGTGGAGTG GGGACACAGC TAATGGAACG CTTTCATTTT
                        (173) ....g.c..c.c. (187)
                                 c   c

351: AGTTCTTTTC CGTTGCGATA GGAGCCATGA TGGCAGCACG AAAAGCAAAA ACTATTGGCG TAAGTTTGTT
   (242) .....g g......g.- .... (26Ø)            (266) ..g. .gagg....a.. .tg.g. (286)

                                                                          tRNA^asp
421: GAGGGGCGGT GCGGTCTTGA ATATCTATTC AAGAGCCCTA TCAGAAAGGC CACCACTGTG CAATTCCTCG
       (299) .. .-....g.. (3Ø8)

491: TTAGTATAGT GGTGAGTATC CCCGCCTGTC ACGCGGGAGA CCGGGGTTCG ATTCCCCGAC GGGGAGGCAC

561: AATGATCGTT TTTTGATGGA CCTGGATACC CGTCGTGGCT CTAATTCCGA CCTGGATACC GGCTACTTTA

631: ATTCCAGCT
```

Figure 4. Sequence of the 639 bp Region of λMt4 which Contains a $tRNA^{Asp}$ Gene.
The non-coding strand of the 639 bp sequence is given in upper case letters with the 5' end to the left. The $tRNA^{Asp}$ coding region is underlined; it would be transcribed from left to right. The tDNA cloverleaf structure encoded is shown in Fig. 3A.
Homologous regious in the 5' flanking sequences of the $tRNA^{Asp}$ gene in λMt2 are shown below the upper case letters. Dots mark identical nucleotides and lower case letters, mismatches, respectively. The numbers indicate the coordinates of the homologous regions taken from Fig. 2.

Taken together, the homologous sequences include 107 of the 478 bp (24%) of the 5' flanking region of the $tRNA^{Asp}$ gene of λMt4. The total homology is 69% (76 homologous positions out of a total of 110 possible positions). The homology of the individual regions varies from 62% to 82%.

DISCUSSION

Comparison of the Mouse and Rat tRNA Gene Clusters

The mouse tRNA gene cluster that we have sequenced is similar to the rat clusters analyzed by Sekiya et al. (1) and Shibuya et al. (14). The sequences of the mouse $tRNA^{Asp}$ and $tRNA^{Gly}$ coding regions are identical to those present in the rat cluster of Sekiya et al. (1). The $tRNA^{Glu}$ coding regions of the two clusters differ in two positions which would not appreciably affect the secondary structures of the tRNA transcripts (see Results). The mouse $tRNA^{Gly}$ and $tRNA^{Glu}$ coding regions differ more substantially from some of those sequenced by Shibuya et al. (14) which contain deletions and are presumably inactive pseudogenes. It is not known whether any of the tRNA genes present in the mouse or rat clusters are expressed in vivo. The fact that the tDNA structures shown in Fig. 3 and those in the rat cluster of Sekiya et al. (1) have all of the structural features of cytoplasmic tRNAs suggests that the genes can encode biologically active molecules. In addition, we have recently shown that the three mouse genes are transcribed by mammalian RNA polymerase III in vitro (R. Rooney and J. Harding, unpublished) and similar results have been cited for the rat genes by Shibuya et al. (14).

With the exception of unrelated regions noted in Results, 5' and 3' flanking regions of the mouse and rat tRNA gene clusters are about 75% homologous. Two other rat-mouse gene pairs exhibit similar degrees of homology. The transcription initiation regions of rRNA genes are 69% homologous by our calculation (25) and the 5' flanking regions of immunoglobulin kappa chain J2 sequences are 82% homologous (26). Like the tRNA gene clusters, these sequences differ by single base differences and small deletions/insertions. Thus the tRNA gene clusters appear to have evolved at about the same rate and exhibit the same types of mutations as other murine sequences.

The fact that the rat tRNA gene clusters are reiterated c. 10 times in the genome (1) complicates the interpretation of the significance of the homology values somewhat. (We have no data for the mouse cluster, but it is reasonable to expect that it is reiterated also.) Analysis of more mouse and rat clusters may reveal different degrees of homology. However, there is evidence that the rat clusters are significantly more homologous to each other than to

the mouse sequence. From the data of Shibuya et al. (14) we calculate the proximal 5' and 3' flanking regions (c. 100-200 bp from the ends of the tRNA coding sequences) of 5 of the rat clusters are 89%-98% homologous.

The sequence comparison suggests the following observations concerning the evolution of the murine tRNA gene clusters. 1) The gene clusters probably evolved from a sequence (or closely related sequences) present in the common ancestor of mice and rats. The alternative possibility, that the clusters arose independently after divergence of the two rodents, is unlikely because of the sequence homologies in the spacer regions. 2) As cited above, the regions immediately flanking the tRNA coding regions of 5 rat clusters are more homologous to each other than they are to the mouse sequence. This suggests that the mouse and rat clusters diverged from an ancestral sequence by mutation and then duplicated after the rat-mouse species divergence. Alternatively, two different (although closely related) sequences in the ancestor of mice and rats could have given rise to the present day sequences. 3) The 3 tRNA genes have remained associated at least since the divergence of mice and rats (probably c. 10 million years ago (27)). This is the first indication to our knowledge that tRNA gene clusters are relatively stable structures, at least over rather short evolutionary time spans. The functional significance, if any, of tRNA gene clustering is not known. It will be of interest to determine whether cognate sequences are present in non-murine mammals. Maintenance of the cluster over longer evolutionary times would suggest a functional role for clustering.

Comparison of Two Mouse $tRNA^{Asp}$ Gene Sequences

The $tRNA^{Asp}$ coding regions present in λMt2 and λMt4 are identical. In contrast, the 5' flanking regions share a more limited "patchwork" pattern of homology, as shown in Fig. 4. We are less certain of the significance of this comparison than of that of the mouse and rat tRNA gene clusters because the sequences surrounding the two mouse $tRNA^{Asp}$ genes contain fewer and shorter strings of consecutive, homologous base pairs than the cognate rat and mouse sequences. However, the proposed homologies involve few insertions or deletions, are linear in the 5' to 3' direction and are probably greater than would be expected from random sequences.

A similar "patchwork" pattern of 5' flanking region homology is also seen in two human $tRNA_i^{Met}$ genes (4). Furthermore, we have recently sequenced a mouse $tRNA_i^{Met}$ gene which is homologous to one of the human genes. Thus the "patchwork" homology pattern can be conserved over rather long intervals of evolutionary time.

The presence of dispersed DNA sequences containing identical tRNA coding

regions raises questions about the origins of tRNA gene families and the processes influencing sequence homology. Different DNA sequences containing identical tRNA coding regions could arise by the duplication of a tRNA structural gene and some flanking sequences followed by mutation of the latter. This is likely to be the case for the two mouse $tRNA^{Asp}$ genes. Furthermore, the fact that the two mouse sequences share less homology than the mouse and rat sequences suggests that the solitary and clustered mouse $tRNA^{Asp}$ genes diverged before the mouse-rat species divergence. A complicating factor is the possibility of gene conversion which can homogenize related, unlinked sequences and has been shown to occur in the tRNA genes of the fission yeast, S. pombe (28). If gene conversion has occurred between the two mouse $tRNA^{Asp}$ genes, it has been limited to the coding regions. Different members of a tRNA family could also originate by the introduction of tRNA coding regions into unrelated DNA sequences via RNA intermediates, as has been suggested for mammalian Alu family sequences (29), small nuclear RNA pseudogenes (30) and a mouse tRNA pseudogene (31). This may not have occurred in the case of the solitary mouse $tRNA^{Asp}$ gene, which lacks the flanking direct repeats that are the hallmark of most Alu family and pseudogene sequences.

ACKNOWLEDGEMENTS

This research was supported by National Foundation, March of Dimes Grant 1-802 and NIH Grant GM26884 to J.D.H. J.E.L. was supported by NIH training grant 5732 GM07216. We wish to thank Barbara Ross for preparing figures.

* Current address: Department of Biochemistry, University of Virginia School of Medicine, Charlottesville, VA, 22908.

+ To whom to address correspondence.

REFERENCES

1. Sekiya, T., Kuchino, Y., and Nishimura, S. (1981) Nucleic Acids Res. 9, 2239-2250.
2. Sekiya, T., Nishizawa, R., Matsuda, K., Taya, Y., and Nishimura, S. (1982) Nucleic Acids Res. 10, 6411-6419.
3. Roy, K.L., Cooke, H., and Buckland, R. (1982) Nucleic Acids Res. 10, 7313-7322.
4. Santos, T., and Zasloff, M. (1981) Cell 23, 699-709.
5. Han, J.H., and Harding, J.D. (1982) Nucleic Acids Res. 10, 4891-4900.
6. Hovemann, B., Sharp, S., Yamada, H., and Soll, D. (1980) Cell 19, 889-895.
7. Hosbach, H.A., Silberklang, M., and McCarthy, B.J. (1980) Cell 21, 169-178.
8. Robinson, R.R., and Davidson, N. (1981) Cell 23, 251-259.
9. Addison, W.R., Astell, C.R., Delaney, A.D., Gillam, I.C., Hayashi, S.,

Miller, R.C., Rajput, B., Smith, M., Taylor, D.M., and Tener, G.M. (1982) J. Biol. Chem. 257, 670-673.
10. Indik, Z.K., and Tartoff, K.D. (1982) Nucleic Acids Res. 10, 4159-4172.
11. Sharp, S., DeFranco, D., Silberklang, M., Hosbach, H.A., Schmidt, T., Kubli, E., Gergen, J.P., Wensink, P.C., and Soll, D. (1981) Nucleic Acids Res. 9, 5867-5882.
12. Guthrie, C., and Abelson, J. (1982) in The Molecular Biology of the Yeast Saccharomyces: Metabolism and Gene Expression (Strathern, J.N., Jones, E.W., and Broach, J.R. eds.) pp. 487-528. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
13. Han, J.H., and Harding, J.D. (1983) Nucleic Acids Res. 11, 2053-2064.
14. Shibuya, K., Noguchi, S., Nishimura, S., and Sekiya, T. (1982) Nucleic Acids Res. 10, 4441-4448.
15. Woo, S.L.C. (1979) Methods Enzymol. 68, 389-395.
16. Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H., and Roe, B.A. (1980) J. Mol. Biol. 143, 161-178.
17. Southern, E.M. (1975) J. Mol. Biol. 98, 503-517.
18. Gauss, D.H., and Sprinzl, M. (1983) Nucleic Acids Res. 11, rl-r53.
19. Kuchino, Y., Shindo-Okada, N., Ando, N., Watanabe, S., and Nishimura, S. (1981) J. Biol. Chem. 256, 9059-9062.
20. Vakharia, V.N. (1981) Fed. Proc. 40, 1753 Abstr. 1234.
21. Vakharia, V.N., and Singhal, R.P. (1982) Biochem. Biophys. Res. Commun. 105, 1072-1081.
22. Kawakami, M., Nishio, K., and Takemura, S. (1978) FEBS Lett. 87, 288-290.
23. Bogenhagen, D.F., and Brown, D.D. (1981) Cell 24, 261-270.
24. Jones, C.W., and Kafatos, C.F. (1982) J. Mol. Evol. 19, 87-103.
25. Financsek, I., Mizumoto, K., and Muramatsu, M. (1982) Gene 18, 115-122.
26. Burstein, Y., Breiner, A.V., Brandt, C.R., Milcarek, C., Sweet, R.W., Warszawski, D., Ziv, E., and Schechter, I. (1982) Proc. Natl. Acad. Sci. U.S.A. 79, 5993-5997.
27. Romer, A.S. (1966) Vertebrate Paleontology, 3rd Ed., University of Chicago Press, Chicago.
28. Munz, P., Amstutz, H., Kohli, J., and Leupold, U. (1982) Nature 300, 225-231.
29. Jagadeeswaran, P., Forget, B.G., Weissman, S.M. (1981) Cell 26, 141-142.
30. Van Arsdell, S.W., Denison, R.A., Bernstein, L.B., Weiner, A.M., Manser, T., Gesteland, R.F. (1981) Cell 26, 11-17.
31. Reilly, J.G., Ogden, R., and Rossi, J.J. (1982) Nature 300, 287-289.