Proc. Natl. Acad. Sci. USA
Vol. 93, pp. 8485–8489, August 1996
Evolution

Translocation events in the evolution of aminoacyl-tRNA synthetases

(histidyl-tRNA synthetase/*Fugu rubripes*)

SYDNEY BRENNER* AND LUIS M. CORROCHANO†

Molecular Genetics, Department of Medicine, University of Cambridge School of Clinical Medicine, Hills Road, Cambridge CB2 2QQ, United Kingdom

Contributed by Syndey Brenner, March 19, 1996

ABSTRACT **We have characterized *hisS*, the gene encoding the histidyl-tRNA synthetase (HisRS) from the tetraodontoid fish *Fugu rubripes*. The *hisS* gene is about 3.5 kbp long and contains 13 exons and 12 introns of 172 bp, on average. The *Fugu hisS* gene encodes a putative protein of 519 amino acids with the three motifs identified as signatures of class 2 aminoacyl-tRNA synthetases. A model for the shifting of intron 8 between *Fugu* and hamster is proposed based on the successive appearance of a cryptic splicing site followed by an insertion mutation that created a new acceptor site. In addition, sequence comparisons suggest that the *hisS* gene has undergone a translocation through the first intron. As a result, the *Fugu* HisRS has an N-terminal sequence markedly different from that in the human and hamster enzymes. We propose that similar events have been responsible for variations at the N-terminal end of other aminoacyl-tRNA synthetases. Our analysis suggests that this involves exchanges through introns of two exons encoding an ancestral 32-amino acid motif.**

Aminoacyl-tRNA synthetases join an amino acid to its corresponding tRNA and are responsible for the specificity of the translation process (1, 2). Since aminoacyl-tRNA synthetases are key components of the contemporary protein-synthesis apparatus, it is possible that knowledge of the structure of genes encoding these enzymes might contain clues to the evolution of protein synthesis. It is known that many of the aminoacyl-tRNA synthetases have related sequences, and the 20 enzymes have been grouped into two classes based on exclusive sets of sequence motifs (3). Class 1 is characterized by the amino acid sequences "HIGH" and "KMSKS" and have a nucleotide binding fold similar to that found in other nucleotide-binding proteins; class 2 synthetases have three consecutive motifs and an active site based on a seven-stranded antiparallel β-sheet (2, 4–7). Aminoacyl-tRNA synthetases contain separable structural domains with different functions (8).

Sequence comparisons and phylogenetic analysis have been used to deduce the likely evolutionary history of aminoacyl-tRNA synthetases and suggests that the contemporary enzymes in the two classes might have arisen by gene duplication from single ancestors (9).

The genomic structure of the genes encoding these enzymes is likely to help in the elucidation of their evolutionary origins, and the position of introns might reveal relationships not easily found in sequence comparisons. Intron positions are only known for the human glutamyl-prolyl-tRNA synthetase (GluProRS), a single gene encoding the two activities in a fusion protein (10), human tryptophanyl-tRNA synthetase (TrpRS) (11), *Caenorhabditis elegans* histidyl-tRNA synthetase (HisRS) (12), and hamster HisRS (13).

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.

The genome of the tetraodontoid fish *Fugu rubripes*, with about 400 Mbp and small introns, is 7.5 times smaller than the human genome (14). The relative small size of this vertebrate genome allows the easy cloning and sequencing of genes from genomic libraries, and genomic structures can be rapidly found. To validate this approach for aminoacyl-tRNA synthetases, we selected the HisRS encoding gene whose intron structure is known in hamster and *C. elegans* (12, 13).

We report the genomic structure of the *Fugu* HisRS gene. A comparison of the intron positions in HisRS with other aminoacyl-tRNA synthetases provides evidence for relatively recent translocation events in the evolution of these enzymes.

MATERIALS AND METHODS

Cloning and Sequencing the *Fugu hisS* Gene. PCR (15) was performed with two sets of nested primers designed to anneal to conserved areas of *hisS*, the gene encoding HisRS. The primers used were His1F, 5′-GAYACNCCNGTNTTYGA-3′, corresponding to the amino acid sequence, DTPVFE, in the human HisRS (13, 16); His1R, 5′-CCNGTRTARTARTC-3′, GTYYD; His2F, 5′-TAYCAYATHGCNAARGT, YHIAKV; and His2R, 5′-TTYTTNACYKCYTCCCA-3′, K/NKVE/AEW. *Taq* polymerase (Perkin–Elmer/Cetus) was used with 1.5 mM Mg^{2+} to amplify and clone a segment of the *Fugu hisS* gene, which was verified by sequencing. A λ2001 *Sau*3AI *Fugu* genomic library (14) was screened with the *hisS* fragment using the method of Church and Gilbert (17). A phage designated λHis was obtained that contained the *Fugu hisS* gene and was confirmed by PCR. The 15 kbp of *Fugu* DNA from λHis was subcloned and sequenced by the chain-termination method (18) using synthetic oligonucleotides as primers. The *hisS* coding sequences were identified by similarities with other HisRS by using the program BLAST (19) accessed from the Human Genome Mapping Program Resource Centre (Harrow, U.K.).

Cloning cDNAs Encoding *Fugu* HisRS by PCR. Total *Fugu* RNA isolated from muscle was used for cDNA synthesis as described (20). PCR was performed using 1–2 μl of the cDNA mix and several set of primers: His7W5 (5′-CATAGCGTTTCCTGTTC-3′) and His2W2 (5′-ATAAAACTCTCTATAACGTC-3′), His7W8 (5′-GCTGAACCTTCATCCTC-3′) and His2W2, and HisAF (5′-TGACAACCCAGCCATGACCC-3′) and His6W12 (5′-TGCGAGTCTGAATGGCG-3′). The first two sets of primers were used to amplify the 5′ end of the gene whereas the last set was used to amplify the 3′ end. The amplified DNA was cloned and sequenced by the chain-

Abbreviations: HisRS, histidyl-tRNA synthetase; GluProRS, glutamyl-prolyl-tRNA synthetase; TrpRS, tryptophanyl-tRNA synthetase; GlyRS, glycyl-tRNA synthetase.

Data deposition: The sequence reported in this paper has been deposited in the GenBank data base (accession no. Z54243)

*To whom reprint requests should be sent at the present address: King's College, Cambridge CB2 1ST, United Kingdom.
†Present address: Departamento de Genética, Universidad de Sevilla, Apartado 1095, E-41080 Sevilla, Spain.

termination method (18) by using synthetic oligonucleotides as primers.

RESULTS

The similarity between human and *Fugu* HisRS allowed the identification of coding sequences and introns in the *hisS* gene sequence. The position of introns was further confirmed by sequencing three overlapping clones containing a *Fugu hisS* cDNA from 407 bp upstream of the putative initiator methionine to 65 bp downstream of the stop codon. The cDNA clones contained several in-frame stop codons upstream of the first methionine, and the stop codon in the last exon, confirming the coding sequence predicted by similarities with human and hamster HisRS (13, 16).

The *hisS* gene comprises 13 exons and 12 introns in about 3.5 kbp (Table 1). The coding sequence has a G+C content of 48.9%, and the intronic DNA has a G+C of 36.5%. Introns in the *Fugu hisS* gene follow the GT/AG rule. All introns are in position 0 (between codons) except intron 8, which is in position 1, after the first nucleotide of the codon. The average intron size is 172 bp, but most introns are about 100 bp in length (Table 1).

The *Fugu hisS* gene encodes a putative protein of 519 amino acids with the three motifs identified as signatures of class 2 aminoacyl-tRNA synthetases (result not shown). As in the hamster gene, the position of introns in the *Fugu hisS* gene are located close to known functional and structural domains. The *Fugu* HisRS is very similar to its human counterpart: 83.5% of the residues are similar and 72.3% are identical. The similarities extend throughout the whole protein, but the amino end is markedly different between the two enzymes. Interestingly, an intron is present between the segment of low similarity and the rest of the gene, suggesting that the first exon and the rest of the gene might have different evolutionary origins (Fig. 1). We could not find any significant similarity between the amino acid sequence of the first exon of *Fugu* HisRS and any other protein in the Protein Identification Resource and SwissProt data bases.

DISCUSSION

If we define the length of a gene from the initiator codon to the stop codon, then the *Fugu hisS* gene is 3.5 kbp long as compared with the hamster gene which is 18 kbp in length (13). Both genes contain 12 introns and 13 exons. As shown in Table 1, the lengths of the coding sequence are similar, about 1.5 kbp; thus there is an eight-fold expansion of intron lengths in the hamster gene. Most of the introns in the *Fugu* gene are small; 10 are 150 bp or less. Other *Fugu* genes contain similarly small introns (14, 21–23). The difference in the length of introns can account for a major part of the overall reduction of genome size of *Fugu*.

Table 1. Size of exons and introns in the HisRS encoding gene of *Fugu* and hamster

No.	Exon size, bp*		Intron size, bp	
	Fugu	Hamster	*Fugu*	Hamster†
1	105	90	120	240
2	90	90	150	7,500
3	120	120	105	1,800
4	96	96	665	1,100
5	126	125	102	1,000
6	108	109	72	120
7	99	99	128	350
8	94	99	120	2,900
9	128	122	80	100
10	273	240	129	180
11	117	117	86	300
12	147	145	306	900
13	57	72		
Total	1560	1524	2063	16,500

Hamster data have been taken from Tsui and Siminovitch (13).
*The limits of the first and the last exons correspond to the predicted coding sequences.
†The size of the last intron in hamster has been estimated from the total length of the HisRS encoding gene (13).

All the introns in the *Fugu hisS* gene are in the same position as in the published hamster sequence (13), except for intron 8 which is shifted 5 bp relative to the hamster intron. The sequence around the 5′-splicing site of intron 8 in *Fugu* is the same as the 3′ end of hamster exon 7 (Fig. 2). We suggest that the shifting of intron 8 can be explained by the following plausible sequence of events. The *Fugu* sequence is taken to be the ancestral form. A cryptic mutation in the intron creating an alternative donor (GT) splicing site could be then followed by an insertion mutation in the exon creating a new active "AG" acceptor site. It has been customary, in the comparison of intron positions, to assume that introns in different homologous genes that occupy positions close to each other, regardless of phase, represent the same original intron, one representative having moved by an undefined process of intron sliding. This has been rightly criticized, and these cases have been used as evidence for independent insertions (24–26). We are fortunate here to be able to provide a plausible basis for an intron shift which has occurred relatively recently.

The first two exons of the human HisRS gene encode a 32-amino acid repeat sequence that is also present at the N-terminal ends of eukaryotic HisRS, TrpRS, and glycyl-

```
                                  ---------------repeat---------------
                                  *   .. **.. ** .*. ...*    .. . *  *..**. **. .. ....  ...      .
ProRS-Human          672-CTTSEDSLVLYNRVAVQGDV.VRELKAKKAPK ED VDAA VKQLLSLKAEYKEKTGQEYKPGNPPA EIGQNISSNSSASIL
                             ESKSLYDEVAAQGEV.VRKLKAEKSPK=AK INEA VECLLSLKAQYKEKTGKEYIPGQPPL SQSSDSSPTRNSEPAGLETP
                             EAKVLFDKVASQGEV.VRKLKTEKAPK=DQ VDIA VQELLQLKAQYKSLIGVEYKPVSATG AEDKDKKKKEKENKSEKQNK
ProRS-Drosophila        1118-AKDELTQEINAQGEK.VRAAKGNKAAK EV IDAE VAKLLALKAKYKEVTGTDF.PVAGRG GGGGGGSAKKAPKEAQPKPA
TrpRS-Human           1-MPNSEPASLLELFNSIATQGEL.VRSLKAGNASK=DE IDSA VKMLVSLKMSYKAAACEDYKADCPPG NPAPTSNHGPDATEAEEDFV

GlyRS-Human           1-MDGAGAEEVLAPLRLAVRQQGDL.VRKLKEDKAPQ VD VDKA VAELKARKRVLEAKELALQPKDDIVD RAKMEDTLKRRFFYDQAFAI
GlyRS-Bombyx          1-MADPKIEEILAPLRANVKEQGDL.VRKLKEEKAPE ID IKKA VAELKTRKKILEDKELSLAPAEDLFD RAKMEDLIKRRFFYDQSFSM

HisRS-Human                 1-MAERAALEELVKLQGER.VRGLKQQKASA EL IEEE VAKLLKLKAQLGPDESKQKFVLKTPK GTRDYSPRQMAVREKVFDVI
HisRS-Hamster               1-MASPA.LEELVLNSRHRLVRGLKQQKASA DQ=IEEE VAKLLKLKAQLGHDESKQKFVLKTPK=GTRDYSPRQMAVREKVFDVI

HisRS-Fugu             1-MLAMHCARVCSVLMGCRTTTRALSIRSFPGVTL AQ=IDEE VAKLLELKAHLGGDDGKHQFVLKTAK=GTRDYNPKQMAIREKVFNTI
HO3-Human              1-MPLLGLLPRRAWASLLSQLLRPPCASCTGAVRCQ SQ=VAEA V.....LTSQLKAHQEKPNFIIKTPK GTRDLSPQHMVVREKILDLV
                                  ---------------repeat---------------
```

FIG. 1. A repeat in several aminoacyl-tRNA synthetases. The symbol "*" indicates a conserved amino acid, and the symbol "·" indicates that most of the amino acids in that position are conserved among ProRS and TrpRS. Introns are marked by the symbol "=" and a gap in the amino acid sequence is shown with a dot. The position of the intron in the hamster HisRS gene comes from Tsui and Siminovitch (13); the intron positions in the human genes encoding TrpRS, and GluProRS from Frolova *et al.* (11) and Kaiser *et al.* (10) respectively, and position of the HO3 intron was obtained from GenBank (accession no. U18936). Only the last repeat from the GluProRS of *Drosophila* is shown.

A

DNA sequence around intron 8 of the histidyl-tRNA synthetase gene

```
HisRS-Fugu      TATGTTGGTATGCAAGgtgagattt---tctgtagGTGGAATGGATTTGGCTGAACGT
HisRS-Hamster   TATGTCCAGCAGCACGGTGAGgtaaa-----gctccccagGTGTGTCTGGTAGAGCAG
HisRS-Human     TATGTCCAGCAACATGGTGGG                    GTATCCCTGGTGGAACAG
```

B

Model for the shift of intron 8

Taking the *Fugu* intron as the ancestral, an A to G change creates a cryptic splicing site

```
                    ↓
 Y  V  G  M  Q                      G  G  M  D  L  A  E  R
TATGTTGGTATGCAAGgtgagattt---tctgtagGTGGAATGGATTTGGCTGAACGT
TATGTTGGTATGCAAGgtgaggttt---tctgtagGTGGAATGGATTTGGCTGAACGT
```

An insertion of a G creates a frameshift (top) which would be lethal unless the cryptic splicing site becomes functional (bottom).

```
                                    ↓
 Y  V  G  M  Q                      G  G  D  G  F  G STOP
TATGTTGGTATGCAAGgtgaggttt---tctgtagGTGGAGATGGATTTGGCTGAACGT
TATGTTGGTATGCAAGGTGAGgttt---tctgtaggtggagATGGATTTGGCTGAACGT
 Y  V  G  M  Q  G  E                   M  D  L  A  E  R
```

The result is a 5 nucleotide shift in the position of the intron but only a G to E mutation in the amino acid sequence

```
old   Y  V  G  M  Q  G  G  M  D  L  A  E  R
new   Y  V  G  M  Q  G  E  M  D  L  A  E  R
```

FIG. 2. Intron shift in the *hisS* gene. (*A*) Comparison of the sequence around intron 8 in *Fugu*, human, and hamster HisRS encoding genes is shown. Lowercase letters denote intron sequence, and uppercase, exon sequence. The segment of the *Fugu* intron 8 identical to hamster exon sequence is in boldface type. The human and hamster sequences were obtained from Tsui and Siminovitch (13). (*B*) A model for the shift of intron 8. A cryptic splicing site is created by an A to G mutation, and an insertion of a G creates a frameshift that would be lethal unless the cryptic splicing sites becomes functional, resulting in an intron shift. Nucleotide changes are in boldface type and marked by arrows.

tRNA synthetase (GlyRS), but absent in their bacterial counterparts (refs. 16, 27, and 28; Fig. 1). Several copies of the same repeat are also present at the N-terminal end of the ProRS coding region of the GluProRS of human and *Drosophila* (29, 30). The presence of a conserved sequence in different aminoacyl-tRNA synthetases suggests that they had a common origin (8, 28, 31). The DNA sequence encoding the repeat is split by an intron in the genes encoding hamster HisRS and human TrpRS, and in two of the three repeats of the human GluProRS (Fig. 1). Although the position of this intron differs by two codons in the ProRS and HisRS genes, we will treat it as the same intron. It is tempting to speculate that this intron might also be present in a similar location in the vertebrate genes encoding GlyRS. The function of the repeat is not known, but the first two exons of the gene encoding human HisRS, including the repeat, are necessary for enzymatic activity (28). However, in the case of the human TrpRS gene, alternative splicing eliminates the exon encoding the first half of the repeat without loss of function (32).

The presence of the repeat as well as the intron in the gene encoding TrpRS, a class 1 enzyme, suggests that this sequence shares a common origin with the same two exons of the genes encoding human ProRS, and HisRS class 2 enzymes. Fig. 1 also reveals that the similarities between human ProRS and TrpRS extend beyond the repeat and end at a position where another intron is present in the human HisRS gene. We can best explain these results by assuming that the two-exon repeat was present in the ancestor gene that gave rise to at least some, if not all, of the class 2 enzymes, and that the TrpRS gene acquired this sequence by translocation into the second intron of a duplicated class 2 enzyme gene (Fig. 3). We have to explain the absence of this intron from the TrpRS gene by its subsequent loss, which is more plausible than an accurate fusion between two coding sequences. We think these events took place a long time ago, at least before the divergence of insects and mammals as deduced from the similarity of the structures of ProRS and GlyRS.

The amino end of the *Fugu* HisRS is very different from the human HisRS, and it only has the second half of the 32-amino acid repeat (Fig. 1). We note that the first intron clearly separates the region of low similarity from the rest of the protein, suggesting that the first exon in the human and *Fugu hisS* genes had different evolutionary origins. We propose that the *Fugu* gene has been translocated through the first intron from its original position into a new location, capturing an exon and acquiring a new sequence at its amino end (Fig. 3). This is supported by the fact that we could not detect any similarity between the 5′ upstream regions of the human and *Fugu hisS* genes. Such translocations may be important in evolution, allowing coding sequences to acquire novel regulatory properties, but they could also be gratuitous. In any event, we can predict that the genes would occupy different positions in the human and fish genomes.

An analogous translocation seems to have occurred in a sequence homologous to the human HisRS gene, designated HO3 in Fig. 1, which is found just upstream of the gene but in the opposite orientation (33). This homolog has an amino end very different from the other human and hamster HisRS, with an intron separating regions of high and low similarity. Like *Fugu* HisRS, HO3 has most likely lost the amino end of the repeat by translocation of an inverted duplication through the first intron (Fig. 3). Since such a sequence was not found in *Fugu*, this is also a comparatively recent event.

Exon shuffling has been proposed as a major mode for the evolution of protein diversity. If this happened at a very early stage of protein evolution, then there needs to be some correspondence between primitive exons and secondary estructure elements or else most of the combinations would be

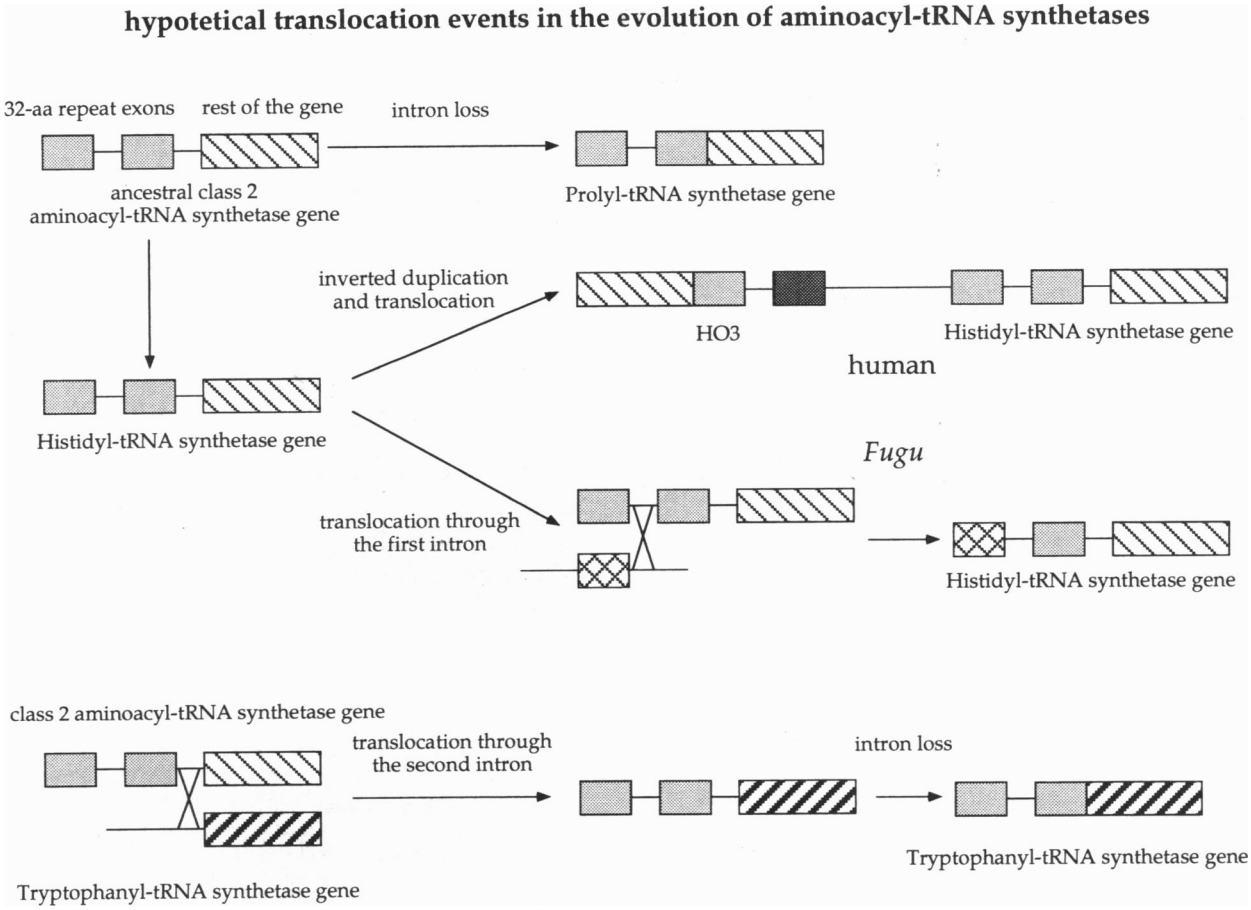

FIG. 3. Hypothetical translocation events in the evolution of aminoacyl-tRNA synthetases. The two exons encoding the 32-amino acid repeat and the rest of the gene are shown by grey and hatched boxes, respectively. An ancestral class 2 aminoacyl-tRNA synthetase gene containing the two exons encoding the 32-amino acid repeat gave rise to the genes encoding at least ProRS and HisRS. Present-day ProRS in animals contain several copies of the repeat and is fused to the gene encoding GluRS (not shown). The gene encoding HisRS suffered an inverted duplication and a translocation in the human lineage resulting in two genes in opposite orientation. The translocation might have occurred through the first intron resulting in the capturing of a new exon in the HO3 gene (HisRS homolog). In the fish lineage, a translocation through the first intron allowed the capturing of a new exon in the *Fugu* HisRS gene. The human TrpRS gene (a class 1 enzyme) has captured the two exons repeat by translocation into a class 2 gene. A further intron loss resulted in the present-day structure of the human TrpRS gene.

useless. Whether a residue of this is still present in contemporary genes is still a matter of debate (25, 26, 34–37). However, fusion of larger protein domains would also be facilitated by exchanges through introns, which is less demanding than the recombination of coding sequences. There are many examples of this in the structure of complex genes, and in many instances an intron can be found to mark the boundary between two different large domains (38). We prefer to call these events translocations, rather than using the looser term of exon shuffling, since we want to emphasize that this is a recombination process which must lead to a change in the position of a gene in the genome.

We also believe that such a translocation will prove to be the major source of evolutionary diversity in the evolution of complex organisms. Through it, novel paths of gene regulation can be explored. Since many genes contain introns in their 5′ untranslated sequences, translocation through these introns would be the best way to achieve such translocations, because the exchange process need not depend on the phases of the recombining introns. Since the position of a gene in a genome changes as a consequence of translocation, the comparative study of genome structures might contain important clues for the evolution of complex functions, even though many of these translocation events may not be productive but gratuitous.

Aminoacyl-tRNA synthetases are old proteins and are composed of several structural and functional motifs (8). We have found that gene translocation appears as the likely mechanism responsible for the divergence in similarity at the amino end between *Fugu* and human HisRS. The relevance of translocations in the evolution of aminoacyl-tRNA synthetases can only be assessed as more results are obtained on the genomic structure of the genes encoding these proteins. We have shown that the *Fugu* genome, with small introns, is an attractive alternative for the task.

We thank Drs. G. Elgar and B. Venkatesh for the *Fugu* DNA and genomic library, and the *Fugu* RNA, respectively. L.M.C. was supported by a Wellcome Trust Postdoctoral Fellowship. This work was done in the Medical Research Council Molecular Genetics Unit, which was closed in 1992.

1. Schimmel, P. (1987) *Annu. Rev. Biochem.* **56,** 125–158.
2. Carter, C. W. (1993) *Annu. Rev. Biochem.* **62,** 715–748.
3. Eriani, G., Delarue, M., Poch, O., Gangloff, J. & Moras, D. (1990) *Nature (London)* **347,** 203–206.
4. Burbaum, J. J. & Schimmel, P. (1991) *J. Biol. Chem.* **266,** 16965–16968.
5. Moras, D. (1992) *Trends Biochem. Sci.* **17,** 159–164.
6. Eriani, G., Cavarelli, J., Martin, F., Ador, L., Rees, B., Thierry, J.-C., Gangloff, J. & Moras, D. (1995) *J. Mol. Evol.* **40,** 499–508.
7. Härtlein, M. & Cusack, S. (1995) *J. Mol. Evol.* **40,** 519–530.
8. Delarue, M. & Moras, D. (1993) *Bioessays* **15,** 675–687.

9. Nagel, G. M. & Doolittle, R. F. (1995) *J. Mol. Evol.* **40,** 487–498.
10. Kaiser, E., Hu, B., Becher, S., Eberhard, D., Schray, B., Baack, M., Hameister, H. & Knippers, R. (1994) *Genomics* **19,** 280–290.
11. Frolova, L. Y., Grigorieva, A. Y., Sudomoina, M. A. & Kisselev, L. L. (1993) *Gene* **128,** 237–245.
12. Amaar, Y. G. & Baillie, D. L. (1993) *Nucleic Acids Res.* **21,** 4344–4347.
13. Tsui, F. W. L. & Siminovitch, L. (1987) *Nucleic Acids Res.* **15,** 3349–3367.
14. Brenner, S., Elgar, G., Sandford, R., Macrae, A., Venkatesh, B. & Aparicio, S. (1993) *Nature (London)* **366,** 265–268.
15. Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. & Erlich, H. A. (1988) *Science* **239,** 487–491.
16. Raben, N., Borriello, F., Amin, J., Horwitz, R., Fraser, D. & Plotz, P. (1992) *Nucleic Acids Res.* **20,** 1075–1081.
17. Church, G. M. & Gilbert, W. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 1991–1995.
18. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
19. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) *J. Mol. Biol.* **215,** 403–410.
20. Corrochano, L. M. (1991) *Genet. Anal. Tech Appl.* **8,** 134–135.
21. Elgar, G., Rattray, F., Greystrong, J. & Brenner, S. (1995) *Genomics* **27,** 442–446.
22. Mason, P. J., Stevens, D. J., Luzzatto, L., Brenner, S. & Aparicio, S. (1995) *Genomics* **26,** 587–591.
23. Baxendale, S., Abdulla, S., Elgar, G., Buck, D., Berks, M., Micklem, G., Durbin, R., Bates, G., Brenner, S., Beck, S. & Lehrach, H. (1995) *Nat. Genet.* **10,** 67–76.
24. Dibb, N. J. & Newman, A. J. (1989) *EMBO J.* **8,** 2015–2021.
25. Rogers, J. H. (1990) *FEBS Lett.* **268,** 339–343.
26. Palmer, J. D. & Logsdon, J. M. (1991) *Curr. Opin. Genet. Dev.* **1,** 470–477.
27. Nada, S., Chang, P. K. & Dignam, J. D. (1993) *J. Biol. Chem.* **268,** 7660–7667.
28. Raben, N., Nichols, R., Dohlman, J., McPhie, P., Sridhar, V., Hyde, C., Leff, R. & Plotz, P. (1994) *J. Biol. Chem.* **269,** 24277–24283.
29. Fett, R. & Knippers, R. (1991) *J. Biol. Chem.* **266,** 1448–1455.
30. Cerini, C., Kerjan, P., Astier, M., Gratecos, D., Mirande, M. & Sémériva, M. (1991) *EMBO J.* **10,** 4267–4277.
31. Arnez, J. G., Harris, D. C., Mitschler, A., Rees, B., Francklyn, C. S. & Moras, D. (1995) *EMBO J.* **14,** 4143–4155.
32. Tolstrup, A. B., Bejder, A., Fleckner, J. & Justesen, J. (1995) *J. Biol. Chem.* **270,** 397–403.
33. O'Hanlon, T. P., Raben, N. & Miller, F. W. (1995) *Biochem. Biophys. Res. Commun.* **210,** 556–566.
34. Dorit, R. L., Schoenbach, L. & Gilbert, W. (1990) *Science* **250,** 1377–1382.
35. Long, M., Rosenberg, C. & Gilbert, W. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 12495–12499.
36. Mattick, J. S. (1994) *Curr. Opin. Genet. Dev.* **4,** 823–831.
37. Stolzfus, A., Spencer, D. F., Zuker, M., Logsdon, J. M. & Doolittle, W. F. (1994) *Science* **265,** 202–207.
38. Doolittle, R. F. (1995) *Annu. Rev. Biochem.* **64,** 287–314.