Proc. Natl. Acad. Sci. USA
Vol. 79, pp. 4138–4142, July 1982
Genetics

sigma, a repetitive element found adjacent to tRNA genes of yeast

(repeated DNA sequences/5-base-pair direct repeats)

FRANCISCO J. DEL REY, THOMAS F. DONAHUE, AND GERALD R. FINK*

Section of Biochemistry, Molecular and Cell Biology, Cornell University, Ithaca, New York 14853

Contributed by Gerald R. Fink, March 31, 1982

ABSTRACT ***sigma*** **is a DNA element of about 340 base pairs (bp) that is repeated many times in the yeast genome. The element has 8-bp inverted repeats at its ends and is flanked by 5-bp direct repeats. The 5-bp repeats are different for each *sigma* and have no homology with the ends of the *sigma* sequence. *sigma* is located 16 or 18 bp from the 5′ end of several tRNA genes. Southern analysis of different yeast strains shows that the pattern of hybridization is different even for closely related strains.**

Recombinant DNA techniques have revealed the structure of dispersed repetitive DNA sequences in the genomes of a variety of eukaryotes. These sequences are often highly reiterated and can constitute a significant proportion of the total DNA in an organism. Several of these repeated DNAs have been cloned and their structures have been analyzed by hybridization and by DNA sequence analysis. Southern hybridization has shown that the position of these sequences in the genome can vary among individuals of the same species (1). Sequence analysis has shown that some dispersed repetitive sequences like the *copia* element of *Drosophila* (2) and the transposable yeast (Ty) element (3, 4) are capable of moving to a different chromosomal location by a nonhomologous recombination event. Transposition could explain the proliferation and variable chromosomal location of dispersed repetitive DNA elements.

In this report we describe *sigma*, a repetitive DNA element of yeast. The most unusual aspect of the element is its location within 16 or 18 base pairs (bp) of the 5′ end of several tRNA genes. The element is found in different locations in closely related yeast strains and its DNA sequence is highly conserved.

MATERIALS AND METHODS

Plasmid DNA. The clones, pYD2 and pYB12, were identified in a pMB9 clone bank of yeast tRNA genes from strain *Saccharomyces cerevisiae* +D4 (5) by hybridization to ^{32}P-labeled histidine tRNA (unpublished data). The lysine tRNA clone, pYL17, is from a pBR313 bank of *S. cerevisiae* S288C *Bam*HI restriction fragments (6) and cross-hybridizes with the yeast DNA insert in pYD2. The plasmids pFD2, pFD12, and pFD17 are the respective *Sal* I/*Hin*dIII (pYD2), *Eco*RI/*Eco*RI (pYB12), and *Hin*dIII/*Eco*RI (pYL17) subclones used for this study (Fig. 1).

The preparation of plasmid DNA and DNA fragments, gel electrophoresis, restriction mapping, and subcloning were performed as described (3, 7, 8).

DNA Sequence Analysis. DNA sequences were determined by the method of Maxam and Gilbert (9). The DNA fragments were labeled at their 3′ ends with deoxynucleoside [α-^{32}P]triphosphates and avian myeloblastosis virus reverse transcriptase (10). Fragments labeled at only one end were produced by a second restriction cleavage or by strand separation. The DNA was subjected to chemical degradation (11) and the reaction products were separated on 40- or 80-cm-long, 0.5-mm-thick gels of 10% or 6% 1:20 bisacrylamide/acrylamide. Autoradiography was with Kodak XR-5 x-ray film at −20°C, using an intensifying screen.

Hybridization Analysis. Total yeast DNA was digested with the restriction enzyme *Eco*RI. DNA fragments were separated on 0.6% agarose gels and transferred to nitrocellulose filter paper (Millipore, HAWP) according to the method of Southern (12). Filters were hybridized with DNA fragments end labeled with avian myeloblastosis virus reverse transcriptase and deoxynucleoside [α-^{32}P]triphosphate. The conditions of hybridization were 0.9 M NaCl/0.09 M trisodium citrate, 10× Denhardt's solution, 10 mM Tris·HCl (pH 7.4), and 0.1% sodium dodecyl sulfate at 50°C for 18 hr. Filters were washed extensively at 50°C in 0.3 M NaCl/0.03 M trisodium citrate/0.1% sodium dodecyl sulfate, dried, and exposed to Kodak XR-5 x-ray film at −70°C, using an intensifying screen. Colony hybridization was performed as described by Grunstein and Hogness (13), using the hybridization and washing conditions mentioned above.

Computer Analysis. DNA sequence data were analyzed with an Apple II computer system. The programs were written by Forrest Fuller and modified by Edward Berry. The DNA homology search and its parameters have been reported (14).

RESULTS

DNA Sequences of *sigma* Elements. We determined the sequences of *sigma* elements from three different chromosomal locations (Fig. 1). A comparison of the DNA sequences of these *sigma*s shows that all three are nearly identical in sequence and size (Fig. 2*a*). The *sigma* in pFD2 differs from the *sigma*s in pFD12 and pFD17 in only 4 and 8 bp, respectively. One of these differences is the presence of an additional G·C bp (position 211) in pFD12 and pFD17 making them 341 bp, 1 bp longer than pFD2 (340 bp).

Although there are no extensive repeated sequences within the *sigma* element, there are a number of short direct and inverted repeats (Fig. 2*a*). The most conspicuous feature is an 8-bp inverted repeat that is present at the ends of the element. The DNA sequence is noteworthy only for its high A+T composition [65% for *sigma* as compared with 59% for total nuclear DNA (15)] and the absence of a large open reading frame (the longest would encode a peptide of 15 amino acids). There is no homology between *sigma* and the adjacent tRNA genes.

Each *sigma* element is flanked by a direct repeat of 5 bp (Fig 2*b*). These 5-bp repeats are different for each *sigma* and have no homology with the ends of the *sigma* sequence. *sigma* was compared with the *delta* portion of the Ty element, a transposable element of yeast known to make 5-bp repeats when it inserts into nonhomologous DNA. The *delta*s are direct repeats

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.

Abbreviation: bp, base pair(s).

* To whom reprint requests should be addressed.

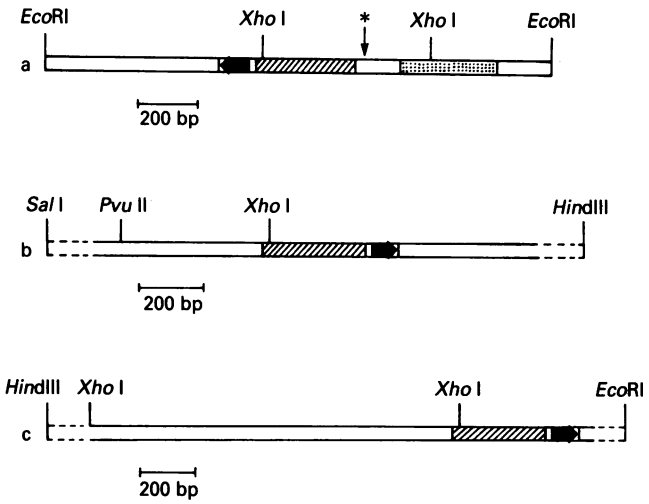

FIG. 1. Restriction map of the DNA fragments containing *sigma* elements and tRNA genes in plasmids pFD12 (*a*), pFD2 (*b*), and pFD17 (*c*). The *sigma* elements are represented by hatched boxes and the tRNA genes and the direction of transcription are shown with black arrows. The dotted box in plasmid pFD12 (*a*) represents the position of a "solo" *delta* sequence. The asterisk marks the *Rsa* I site used to cleave the 3'-labeled *Xho* I/*Xho* I fragment to generate the two fragments used as *sigma* and *delta* probes in hybridization analysis. A detailed restriction map and sequence analysis strategy will be presented elsewhere.

of about 330 bp at the ends of the Ty element. No region of extensive homology exists between *sigma* and *delta*.

We have examined a number of other short, repeated, eukaryotic DNA sequences for homology with *sigma*. No extensive

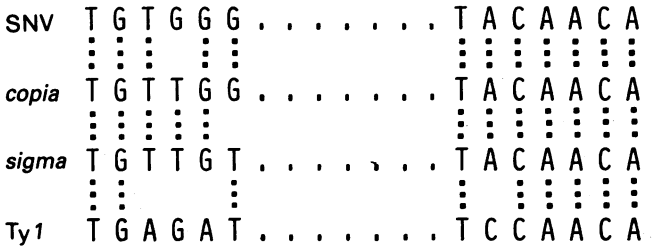

FIG. 3. Sequence homologies between the ends of *copia* (17), Ty (4), spleen necrosis virus (SNV) (19), and *sigma*. The first 6 and last 7 nucleotides of each element are given. The dots between nucleotides signify identity at that position.

homology is observed with the "*Alu*" sequence (16), the direct repeats flanking *copia* (17), or the 5S RNA from *Xenopus* and yeast (18). There is a similarity between the ends of *sigma*, *copia*, chicken spleen necrosis virus (19), and Ty (Fig. 3). Two regions in *sigma* share partial sequence homology with the *Alu* sequences. One of these regions is interesting because it includes part of the upstream half of the RNA polymerase III split promoter of the *Alu* sequence (16). That is, the *Alu* A36 sequence, 5'-G-G-C-T-C-A-C-G-C-C-T-G-T-A-A-T-C-C-C-A-G-A-3', is in the same relative position (+15 to +36) as the *sigma* sequence, 5'-G-G-C-T-C-G-A----G-T-A-A-T-A-C-C-G-G-A-3' (+13 to +30).

***sigma* Elements Are Adjacent to tRNA Genes.** Each of the three *sigma* sequences is located 16 or 18 bp from the 5' end of a tRNA gene (Fig. 2*b*). The *sigma* elements of clones pFD2 and pFD12 are located next to histidine tRNA genes and the *sigma* element of clone pFD17 is next to a lysine tRNA gene. The histidine tRNA genes in pFD2 and pFD12 have identical sequences (unpublished data), but the *sigma* elements are in

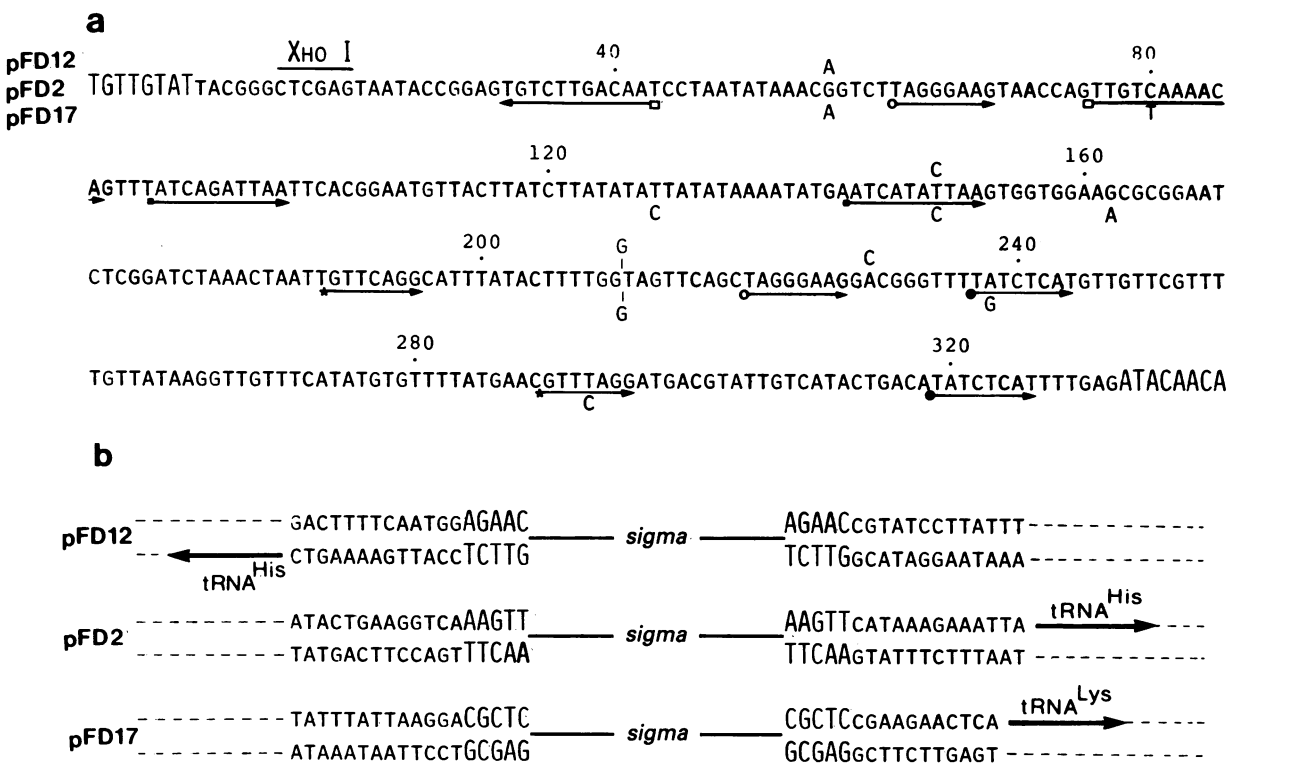

FIG. 2. Nucleotide sequence of the *sigma* element and flanking regions. The *sigma* sequence is written only once (*a*), as it exists in plasmid pFD2. Only one strand of the sequence is shown. Where the nucleotide sequences of the three *sigmas* differ, the base changes are placed below (pFD17) or above (pFD12) the pFD2 sequence. The 8-bp inverted repeat at the ends of *sigma* is written in larger letters. Repeats within the element are designated by arrows. (*b*) Flanking regions in the three different locations in which *sigma* was found. The larger type indicates the 5-bp duplication in the adjoining chromosomal DNA. The location of the tRNA genes found close to the *sigma* sequences is shown with black arrows.

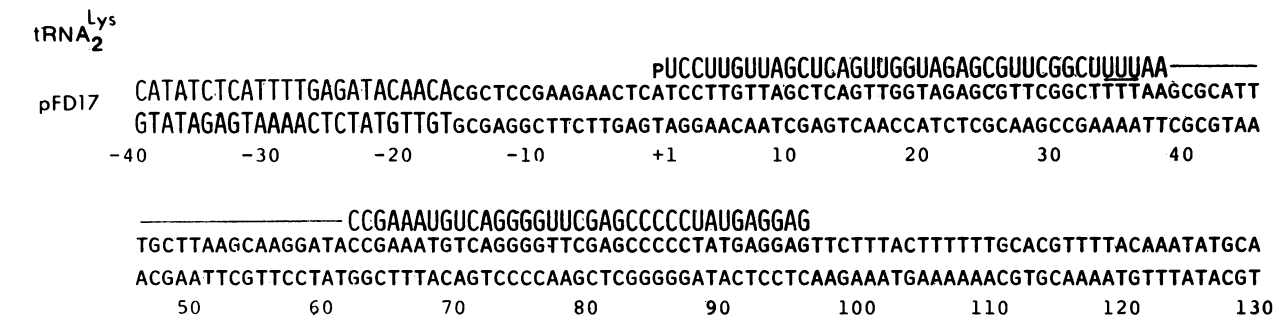

FIG. 4. DNA sequence of the yeast tRNA$_2^{Lys}$ gene (pFD17) and the previously described RNA sequence of tRNA$_2^{Lys}$ (20) (top row above the DNA sequence with the anticodon underlined). The terminal C-C-A of the tRNA is not encoded by the DNA. A 23-bp intervening sequence between positions 39 and 61 is marked by a line. The larger type in the DNA sequence shows the end of the *sigma* element found in the vicinity of the tRNA$_2^{Lys}$ gene.

inverted orientation with respect to the histidine tRNA gene. The DNA sequence of the lysine tRNA gene in clone pFD17 is complementary to the tRNA$_2^{Lys}$ sequence (20), except for a 23-bp region that is present only in the DNA sequence (+39 to +61, see Fig. 4). The 23-bp intervening sequence occurs 3′ to the anticodon of the lysine tRNA at a position similar to that found in other tRNA genes known to contain intervening sequences (21, 22).

The *sigma* element is not adjacent to all the genes coding for tRNA^{His} in yeast. We have identified five other independent clones containing histidine tRNA genes on unique restriction fragments (unpublished data). When plasmids containing each of these are probed in Southern hybridization with a DNA fragment containing *sigma*, none gives a positive signal. The DNA sequence of one of these clones contains a histidine tRNA gene identical in sequence to that found in pFD2 and pFD12. However, no *sigma* sequences are present in the 5′ or 3′ regions flanking this histidine tRNA gene.

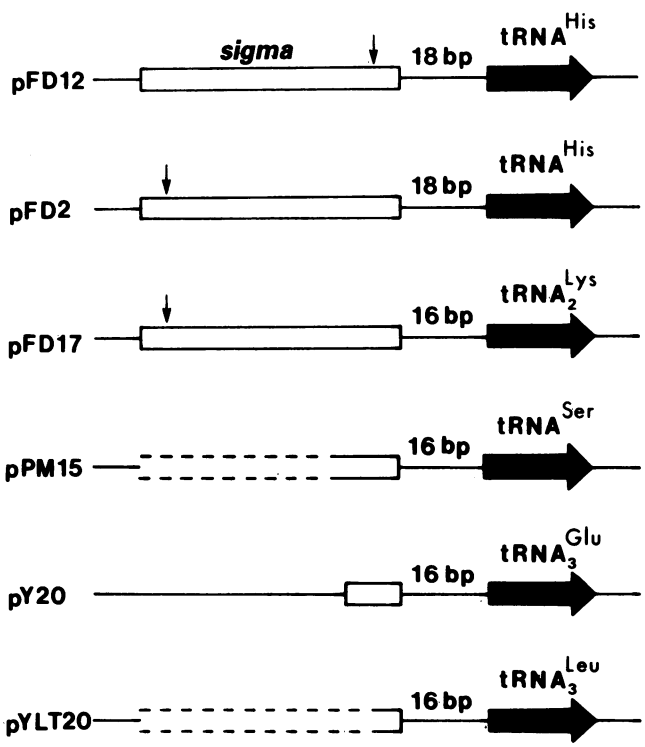

FIG. 5. Schematic representation of the position and orientation of the *sigma* element with respect to the adjacent tRNA genes. Data concerning pPM15, pY20, and pYLT20 are from refs. 23, 24, and 25, respectively. The published sequence data for pPM15 and pYLT20 include only 24 and 84 bases of *sigma* 5′ to the tRNA gene. The *sigma* sequence in pY20 is truncated and contains only 72 bp. The box represents the *sigma* element and the dark arrow represents the tRNA gene. The small vertical arrow shows the position of the *Xho* I site within the element.

A search for *sigma* in the published DNA sequences of other yeast tRNA genes reveals that *sigma* is adjacent to a number of these genes (Fig. 5). *sigma* is located 16 bp from the 5′ end of the genes encoding serine (23), glutamate (24), and leucine tRNA (25). The *sigma* adjacent to the glutamate tRNA gene is truncated and contains many base pair differences from the intact *sigmas* (13 changes in 72 bp). The published sequence data for the serine and leucine genes do not extend far enough in the region 5′ to the tRNA to include an entire *sigma* element. *sigma* is not found adjacent to certain genes for proline (26), glutamate (27), or serine (23) tRNAs.

***sigma* Is a Dispersed Repetitive Element.** Southern analysis of six different yeast strains suggests that *sigma* sequences are present in a number of different chromosomal locations. Total genomic DNA from each of these strains was hybridized to a ^{32}P-labeled *sigma* probe (the *Xho* I/*Rsa* I fragment in Fig. 1). All six yeast strains contain multiple fragments that hybridize to *sigma* (Fig. 6, lanes a–f). Although this hybridization suggests that *sigma* is located at different positions in these strains, we can not rule out restriction site polymorphism as an explanation of the differences in hybridization patterns. The pattern of hybridization with *sigma* is clearly different from that obtained when the *delta* element is used as a probe (Fig. 6, lanes g–l). This hybridization analysis suggests that laboratory strains have at least 20–25 copies of *sigma* per haploid genome. To estimate the proportion of tRNA genes associated with *sigmas*, we screened a clone bank of yeast tRNAs (5) with the *Xho* I/*Rsa* I fragment, which contains *sigma* sequences (Fig. 1). Of the 225 clones examined, 15 contain *sigma*. This tRNA clone bank is not complete. Southern analysis of yeast genomic DNA, using purified ^{32}P-labeled tRNA^{His} as a probe, identifies seven yeast fragments containing histidine tRNA genes, but only four of these fragments are represented in the tRNA bank (unpublished data). So there could be additional tRNA genes with an associated *sigma*. Moreover, we have no information on whether there are *sigmas* without associated tRNA genes.

DISCUSSION

sigma has a number of features in common with transposable elements of both prokaryotes and eukaryotes. The transposable elements studied so far have repeated sequences at their ends and generate direct repeats of target sequences in the process of transposition (2, 3, 4, 17, 28). *sigma* has 8-bp inverted repeats at its ends and is flanked by 5-bp direct repeats. In addition,

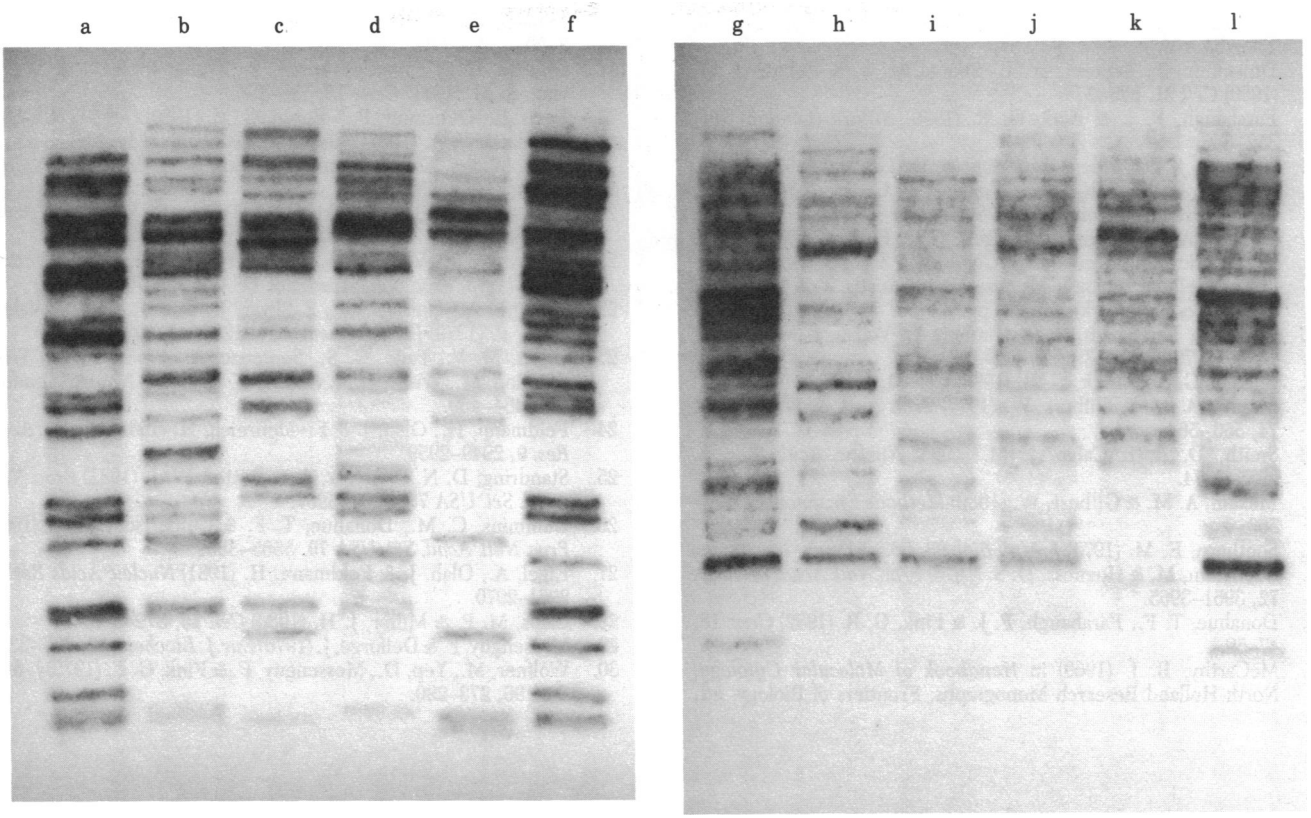

FIG. 6. Southern hybridization analysis of six different yeast strains, using *sigma* (lanes a–f) or *delta* (lanes g–l) as a probe. Chromosomal DNA, prepared from six interbreeding yeast strains as described by Roeder and Fink (8), was cleaved in each case with *Eco*RI. The *sigma* and *delta* sequence probes used were the *Xho* I/*Rsa* I fragments of plasmid pFD12 (see Fig. 1) labeled at the 3′ end of the *Xho* I sites with [α-^{32}P]dCTP and reverse transcriptase. Fragments were separated in a 6% polyacrylamide gel, eluted, and used as a probe. Shown are the patterns of *S. cerevisiae* S288C (a, g), *S. norbensis* (b, h), *S. cerevisiae* E1278b (c, i) (J. M. Wiame), *S. cerevisiae* F114 (d, j) (Fleischmann's), *S. carlbergensis* JHill (e, k), and *S. cerevisiae* D585-11C (f, l).

closely related strains appear to have *sigma*s located at different positions in the genome. Despite these similarities to transposable elements there is, as yet, no direct evidence for transposition of *sigma*.

The *sigma* could represent a portion of a larger transposable element that is present at one of the three sites whose sequences we have determined. However, because there is no homology in the 5′ and 3′ regions flanking these *sigma*s, no element larger than *sigma* can be repeated at all three sites. A model consistent with our data is a large transposable element flanked by *sigma* elements transposed into each of these regions, making a 5-bp repeat at the junction between each *sigma* and the target DNA. Recombination between the *sigma*s excised the main portion of the element and left a solo *sigma* flanked by 5-bp repeats. This recombination event would be similar to *delta-delta* recombination, which has been observed to occur for several Ty elements (3, 4).

The most striking feature of *sigma* is its location 16 to 18 bp upstream from the 5′ end of tRNA genes. The close association between *sigma* and the tRNA gene could be explained by several models for the mechanism of *sigma* transposition. The presence of *sigma* adjacent to tRNA genes may reflect its preference in the process of transposition for some structural feature of the tRNA gene. The available data suggest that *sigma* has no preference for particular tRNAs (identical histidine tRNA genes may or may not have an adjacent *sigma*) or for tRNA genes with intervening sequences (the leucine and lysine tRNA genes have intervening sequences but the histidine, serine, and glutamate tRNA genes do not). Alternatively, the tRNA gene could be part of the transposition unit. Transposition of tRNA genes could account for the presence of multiple copies of identical tRNA genes in yeast. One argument against this model is that identical histidine tRNA genes have different DNA sequences in the 18-bp region between the *sigma* and the tRNA gene.

The proximity of *sigma* to the site of initiation of tRNA transcription suggests the obvious possibility that *sigma* confers some special transcriptional properties on the adjacent tRNA gene. However, *sigma* cannot be required for the transcription of all tRNA genes because the majority of tRNA genes do not contain *sigma* sequences.

sigma could be part of a common regulatory control involving tRNA genes. tRNAs have been implicated in the general control of amino acid biosynthesis in yeast (29). The manifestation of general control is derepression of several amino acid biosynthetic pathways in response to starvation for any one of a number of amino acids (30). *sigma* could provide a common regulatory region in front of a subset of tRNA genes involved in the general control. Further experiments on the *in vivo* and *in vitro* behavior of cloned *sigma* sequences should help to elucidate any genetic and regulatory phenomena associated with these elements.

The authors thank F. Chumley, K. Durbin, A. Hinnebusch, and F. Winston for their critical review of this manuscript and F. Ormsbee for its preparation. We also thank Maynard Olson for helpful discussions. This work was supported by National Institutes of Health Grants CA23441 and GM15408 to G.R.F., who is an American Cancer Society Professor of Biochemistry. F.D.R. is the recipient of a fellowship from the Comite Conjunto Hispano-Norteamericano para la Cooperacion Cientifica y Tecnologica, and T.F.D. was supported by the Jane Coffin Childs Memorial Fund for Medical Research.

1. Cameron, J. R., Loh, E. Y. & Davis, R. W. (1979) *Cell* **16**, 739–751.
2. Dunsmuir, P., Brorein, W. J., Simon, M. A. & Rubin, G. M. (1980) *Cell* **21**, 575–579.
3. Farabaugh, P. J. & Fink, G. R. (1980) *Nature (London)* **286**, 352–356.
4. Gafner, J. & Philippsen, P. (1980) *Nature (London)* **286**, 414–418.
5. Petes, T. D., Broach, J. R., Wensink, P. C., Hereford, L. M., Fink, G. R. & Botstein, D. (1978) *Gene* **4**, 37–49.
6. Hinnen, A., Farabaugh, P., Ilgen, C. & Fink, G. R. (1979) in *Eucaryotic Gene Regulation*, ICN–UCLA Symposia on Molecular and Cellular Biology, eds. Axel, R., Maniatis, T. & Fox, C. F. (Academic, New York) Vol. 14, pp. 43–50.
7. Ilgen, C., Farabaugh, P. J., Hinnen, A., Walsh, J. M. & Fink, G. R. (1979) *Genet. Eng.* **1**, 117–132.
8. Roeder, G. S. & Fink, G. R. (1980) *Cell* **21**, 239–249.
9. Maxam, A. M. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 560–564.
10. Smith, D. R. & Calvo, J. M. (1980) *Nucleic Acids Res.* **8**, 2255–2274.
11. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
12. Southern, E. M. (1975) *J. Mol. Biol.* **98**, 503–517.
13. Grunstein, M. & Hogness, D. S. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 3961–3965.
14. Donahue, T. F., Farabaugh, P. J. & Fink, G. R. (1982) *Gene* **18**, 47–59.
15. McCarthy, B. J. (1969) in *Handbook of Molecular Cytology*, North-Holland Research Monographs: Frontiers of Biology, ed. Lima-de-Faria, A. (North-Holland, Amsterdam), Vol. 15, pp. 3–20.
16. Duncan, C. H., Jagadeeswaran, P., Wang, R. R. C. & Weissman, S. M. (1981) *Gene* **13**, 185–196.
17. Levis, R., Dunsmuir, P. & Rubin, G. M. (1980) *Cell* **21**, 581–588.
18. Erdmann, V. A. (1981) *Nucleic Acids Res.* **9**, r25–r42.
19. Shimotohono, K., Mizutani, S. & Temin, H. M. (1980) *Nature (London)* **285**, 550–554.
20. Madison, J. T., Boguslawski, S. J. & Teetor, G. H. (1972) *Science* **176**, 687–689.
21. Valenzuela, P., Venegas, A., Weinberg, F., Bishop, R. & Rutter, W. J. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 190–194.
22. Venegas, A., Quiroga, M., Zaldivar, J., Rutter, W. J. & Valenzuela, P. (1979) *J. Biol. Chem.* **254**, 12306–12309.
23. Olson, M. V., Page, G. S., Sentenac, A., Piper, P. W., Worthington, M., Weiss, R. B. & Hall, B. D. (1981) *Nature (London)* **291**, 464–469.
24. Feldmann, H., Olah, J. & Freidenreich, H. (1981) *Nucleic Acids Res.* **9**, 2949–2959.
25. Standring, D. N., Venegas, A. & Rutter, W. J. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 5963–5967.
26. Cummins, C. M., Donahue, T. F. & Culbertson, M. J. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 3565–3569.
27. Eigel, A., Olah, J. & Feldmann, H. (1981) *Nucleic Acids Res.* **9**, 2961–2970.
28. Calos, M. P. & Miller, J. H. (1980) *Cell* **20**, 579–595.
29. Messenguy, F. & Delforge, J. (1976) *Eur. J. Biochem.* **67**, 335–339.
30. Wolfner, M., Yep, D., Messenguy, F. & Fink, G. F. (1975) *J. Mol. Biol.* **96**, 273–290.