Volume 12 Number 2 1984 Nucleic Acids Research

The 5S ribosomal genes in the *Drosophila melanogaster* species subgroup. Nucleotide sequence of a 5S unit from *Drosophila simulans* and *Drosophila teissieri*

Marie-Laure Samson and Maurice Wegnez

Centre de Génétique Moléculaire, Laboratoire Propre du CNRS, associé à l'Université Pierre et Marie Curie, Paris VI, F-91190, Gif-sur-Yvette, France

Received 18 October 1983; Revised and Accepted 23 November 1983

SUMMARY

The 5S genes of the eight species of the D. melanogaster subgroup have been mapped. The spacers, in contrast with coding regions, differ markedly between most species. One 5S gene unit has been sequenced for both D. simulans and D. teissieri. The mature 5S RNA region in these two species is identical to the corresponding region of D. melanogaster. Only 5 nucleotide variations occur between the D. melanogaster and D. simulans 5S gene spacers. The spacer in D. teissieri is very different. Only two segments, located one at each side of the coding region, are clearly homologous to corresponding sequences of D. melanogaster and D. simulans.

INTRODUCTION

The 5S ribosomal genes form a gene family present in all eucaryotic species. These small tandemly repeated genes are composed of two regions which evolve at drastically different rates. Evolution of the coding region (5S RNA, 120 nucleotides) is very slow as shown by the comparison of all available 5S RNA sequences (1). The spacer region, on the contrary, evolves much faster. Sequence homologies outside the 5S coding region in two Xenopus species, X. laevis and X. borealis, are few and short (2).

The high degree of homogeneity among members of gene families is attributed to a process known as concerted evolution (3,4). This term indicates that the units of a gene family tend to evolve in unison, and are thus kept similar to one another within a species, while they diverge in structure from the same gene family units in other species. Unequal crossing-over and gene conversion are the two mechanisms which have been proposed to account for this evolutionary process (5-11). The 5S genes of the melanogaster subgroup species seem to us a well suited system for an experimental approach to the problem of concerted evolution. In this paper, we report the mapping of the 5S genes from all eight sibling species of the melanogaster subgroup as well as the sequence of one 5S gene unit of D. simulans and D. teissieri.

MATERIALS AND METHODS

Drosophila stocks

The Drosophila species were kindly provided by L. Tsacas and J. David from the "Laboratoire de Biologie et Génétique Evolutives" in Gif-sur-Yvette, France. All of the melanogaster subgroup species originated from Africa : D. erecta, Ivory Coast, strain 220-5; D. mauritiana, Mauritius, str. 163-1; D. melanogaster, Ivory Coast, str. 192-1; D. orena, Cameroun, str. 188-1; D. sechellia, Seychelles, str. 228; D. simulans, Seychelles, str. 206-1; D. teissieri, Congo, str. 201-5; D. yakuba, Cameroun, str. 115.

Purification of Drosophila DNA

DNA was prepared from dechorionated 2 to 24 h old embryos. About 5 g of embryos were homogenized at 0°C in 10 ml of 10 mM NaCl, 10 mM EDTA, 10 mM Tris-HCl, pH. 8.0. The homogenate was filtered through nylon tissue (100 - 110 µm, then 30-40 µm mesh size), and centrifuged at 1,000 g for 5 min. The pellet (cells and nuclei) was resuspended in 7 ml of the homogenization buffer. After addition of SDS to 1 % and $NaClO_4$ to 1 M, the solution was extracted with chloroform. Ethanol precipitated DNA was dissolved in 1 mM EDTA, 10 mM Tris-HCl, pH 8.0 and incubated at 37°C with pronase (100 µg/ml) for 2 h. Following an additional chloroform extraction, DNA was purified by CsCl centrifugation.

Gel electrophoresis and Southern transfer

Horizontal 0.3 to 1.5 % agarose and vertical 6.2% polyacrylamide gels were used. To facilitate handling of dilute gels (0.3% agarose), a 3 mm thick 2% agarose layer was poured first. DNA transfer to Schleicher and Schüll BA 85 nitrocellulose filters was carried out in 1.5 M ammonium acetate, 20 mM NaOH (12). Retention of fragments smaller than 500 bp was much increased by long baking at 80°C (up to 10 h). Hybridization conditions were those of Wahl et al. (13).

Construction of plasmids with 5S gene insertions

The plasmid pUC8 was used (14). This vector carries a multiple cloning site inserted into a segment of the E. coli lac operon. Insertion of a DNA fragment into one of these sites leads to the inactivation of the gene that complements a defective β-galactosidase gene carried by the host (E. coli K12 strain JM 103). Cells transformed with recombinant plasmids produce white colonies on a background of blue colonies when plated on X-gal (15).

We cloned one 5S gene unit of D. simulans from totally Mbo I digested DNA. The digest was fractionated on a 5% polyacrylamide gel. Slices were cut from the gel around the 380 bp region, and the DNA eluted as described by

Maxam and Gilbert (16). The 5S gene enriched DNA was ligated to Bam HI linearized pUC8 DNA. Transformation of the host JM 103 was achieved according to the procedure of Mandel and Higa (17). The method of Grunstein and Hogness (18) was used for screening. The screening probe was a purified Alu I fragment from plasmid pBR7A (kindly provided by V. Pirrotta and C. Tschudi) that contains 10 copies of the 5S gene of D. melanogaster and 650 bp of pBR322 DNA.

For the cloning of a D. teissieri 5S gene unit, we first screened our genomic library, constructed in vector λ 1059 (19,20). One isolated recombinant, λ Te5S, that has a 20 kb insert containing several 5S genes, was then used for subcloning. DNA of λ Te5S, totally digested with Mbo I, was ligated to Bam HI linearized pUC8 DNA. Subsequent steps were as for D. simulans 5S gene cloning.

DNA sequencing

Plasmid DNA was purified by the method of Birnboim and Doly (21). The 5S gene restriction fragments were dephosphorylated, fractionated on 5% polyacrylamide gels and labelled at their 5'-ends using T4 polynucleotide kinase (16). Fragments with a unique ^{32}P-labelled end were obtained either by strand separation or restriction. Sequencing reactions were according to the Maxam and Gilbert procedure (16).

RESULTS

Restriction mapping of the 5S RNA genes in the melanogaster species subgroup

Restriction mapping of the 5S RNA genes in the eight species of the melanogaster subgroup was carried out according to standard procedures. In some cases, partial digests were performed. It was possible to use the D. melanogaster 5S gene as probe since RNase T_1 and RNase A fingerprints of the 5S RNA from the melanogaster subgroup species are undistinguishable (M.Wegnez, unpublished results). In the conditions of high stringency we used (50% formamide, 42°C), there was no crosshybridization between the spacer of the D. melanogaster 5S gene and the spacers of the 5S genes of the less closely related species, i.e. D. teissieri, D. yakuba, D. erecta and D. orena.

An example of the patterns produced by restriction analysis of the 5S genes is given in Fig.1. The small size of the restriction fragments in most species shows that at least one Taq I site occurs within the 5S gene unit. The reaction was incomplete with D. melanogaster DNA (lanes 3 and 4), leading

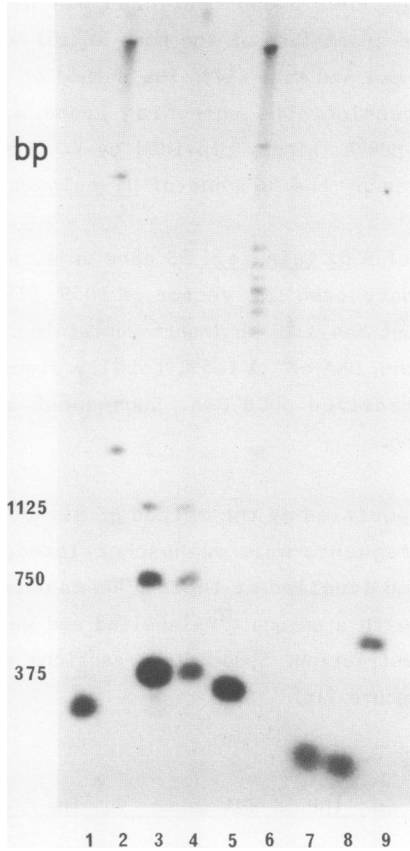

Figure 1 - Autoradiogram of embryonic Drosophila DNA digested with Taq I, fractionated on a 1.2 % agarose gel and hybridized to a ^{32}P-labelled 5S gene probe. DNA lanes : (1) D. erecta, (2) D. mauritiana, (3 and 4) D. melanogaster, (5) D. orena, (6) D. simulans, (7) D. teissieri, (8) D. yakuba, (9) D. nasuta.

to the appearance of 750 and 1125 bp fragments, i.e. dimeric and trimeric forms of the 5S gene unit. The patterns of D. mauritiana (lane 2), D. simulans (lane 6) and D. sechellia (not shown), showing high molecular weight fragments, are not due to partial digestion. This result was repeatedly obtained with increased concentrations of enzyme and longer times of incubation. Thus there is no Taq I site in most of the 5S gene units of these species. Five other restriction enzymes were used for the mapping. The restriction maps of the 5S genes of the eight melanogaster subgroup species are shown in Fig.2.

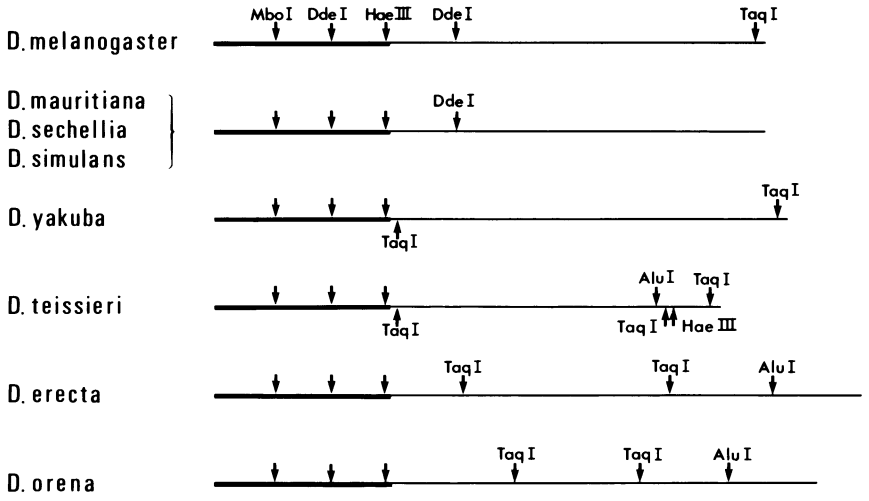

Figure 2 - Restriction maps of the 5S gene unit in the eight D. melanogaster subgroup species. The sequence coding for the mature 5S RNA (120 base pairs) is indicated by a solid box. The direction of transcription is from left to right. The D. melanogaster 5S RNA precursor has 15 additional nucleotides at its 3'-end (22). The precursor specific sequence has been drawn as a solid line in continuity with the spacer sequence. The D. simulans and D. teissieri restriction maps were checked with the corresponding sequences (Fig. 4 and 5). The Taq I site located between the Alu I and Hae III sites in the D. teissieri spacer, not detected before, was added to the map. The lengths, in base pairs, of the 5S gene units are respectively : D. melanogaster, D. mauritiana, D. sechellia and D. simulans : 375; D. yakuba : 390; D. teissieri : 345; D. erecta : 440; D. orena : 410.

Heterogeneity of the 5S genes

Some length heterogeneity between members of the D. melanogaster 5S gene family has been demonstrated (23). This heterogeneity is due to variable copies of a repetitive heptamer (23). To examine 5S gene length heterogeneity in the other species of the subgroup, we used polyacrylamide rather than agarose gels. As shown in Fig.3, significant length heterogeneities were detected. In the case of D. teissieri for example, there are 3 size classes of 5S gene units. We found the same pattern when analysing a cloned array of approximately thirty 5S gene units from D. teissieri inserted into plasmid pUC8 (data not shown).

Sequencing of a 5S gene unit of D. simulans

As shown in Fig.2, all the melanogaster subgroup species have only one Mbo I site in the 5S gene. We thus cloned one 5S gene unit of D. simulans in plasmid pUC8, starting from Mbo I digested DNA. The clone used for

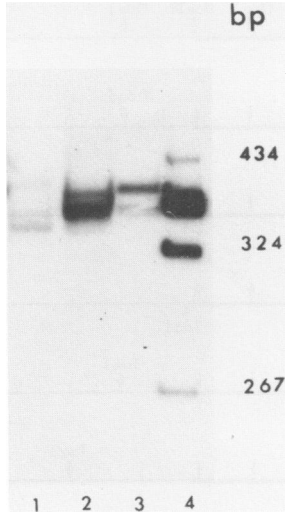

Figure 3 - Autoradiogram of embryonic Drosophila DNA digested with Mbo I, fractionated on a 6.2% polyacrylamide gel and hybridized to a ^{32}P-labelled 5S gene probe. DNA lanes: (1) D. teissieri, (2) D. melanogaster, (3) D. simulans, (4) Hae III digest of pBR7A.

sequencing, pSi5S, was the only 5S clone obtained out of 1000 tested colonies. The 5S gene insert migrates in polyacrylamide gels as the most abundant class of the D. simulans 5S gene family, i.e. the slowest migrating one of Fig.3.

The sequence of the 5S gene, shown in Fig.4, is almost identical to the D. melanogaster 5S gene sequence (23). The heptameric block (GCTGCCT) located at the 3'-end of the coding region, is present in 4 copies. An other heptamer (TTGGCTA), present once in D. melanogaster , is duplicated in the D. simulans sequence (nucleotides 187-193 and 194-200, Fig.4). There are only 5 nucleotide variations between the D. simulans and D. melanogaster spacers (Fig.4).

Sequencing of a 5S gene unit of D. teissieri

We subcloned one 5S gene unit of D. teissieri from phage λ Te5S in plasmid pUC8. This clone, pTe5S, has a 5S gene insert that migrates as the most abundant class of the D. teissieri 5S gene family, i.e. the fastest moving one (Fig.3).

The sequence of the 5S gene is shown in Fig.5. The mature 5S RNA region is identical to the corresponding regions of D. melanogaster and D. simulans, but two nucleotide variations exist in the precursor sequence. In the spacer,

```
   1
   G C C A A C G A C C A T A C C A C G C T G A A T A C A T C G G T T C T C G T C C

  41
   G A T C A C C G A A A T T A A G C A G C G T C G G G C G C G G T T A G T A C T T

  81
   A G A T G G G G G A C C G C T T G G G A A C A C C G C G T G T T G T T G G C C T

 121
   C G T C C A C A A C T T T T T G C T G C C T G C T G C C T G C T G C C T G C T G

 161                          T                              - - - - - - -
   C C T T C T T A G T T T T T A T T C T A G C A T T A T T G G C T A T T G G C T A

 201
   C A A A T C A G A A T G A A A A C T T T G T T C A C C T A A T T T C A A A T T T

 241          T
   T G T C T T C C A C T C A T T A A T C T T T T A T A T C T T T A T T A C G A T A

 281      G
   T C T A A A A C C C A A T G G T A C A A A A A C A G T C T A T T T C A G T C T A

 321
   T G G G C A T A A C T G A A T A T C A G A G T A T A A G G A C A C T G T T T A G

 361      C T
   C C C A G C G A C T T T C
```

Figure 4 - Sequence of a 5S gene unit of D. simulans (non coding strand). Nucleotides 1-120 : mature 5S RNA region, nucleotides 121- 135 : precursor specific sequence, nucleotides 136-373 : spacer. The 5 nucleotides differing from the D. melanogaster 5S gene (23) are written above the sequence. Symbol - above nucleotides 194-200 indicates that these nucleotides are missing in the D. melanogaster sequence. Repetitive heptamers are underlined.

the repetitive heptamer (GCTGCCT) common to the D. melanogaster and D. simulans sequences is present once, but not directly adjacent to the coding region (nucleotides 151-157, Fig.5). An altered copy of this heptamer, with an internal reiteration of 4 T residues, is present directly after the coding region (nucleotides 137-147, Fig.5). The spacer is very different from those of D. melanogaster and D. simulans. We have searched for homologous regions in D. melanogaster and D. teissieri 5S gene spacers with the EPURE program (24). Only two 30 nucleotide stretches turned out to have a highly significant homology. One is located not far from the 3'-end of the coding

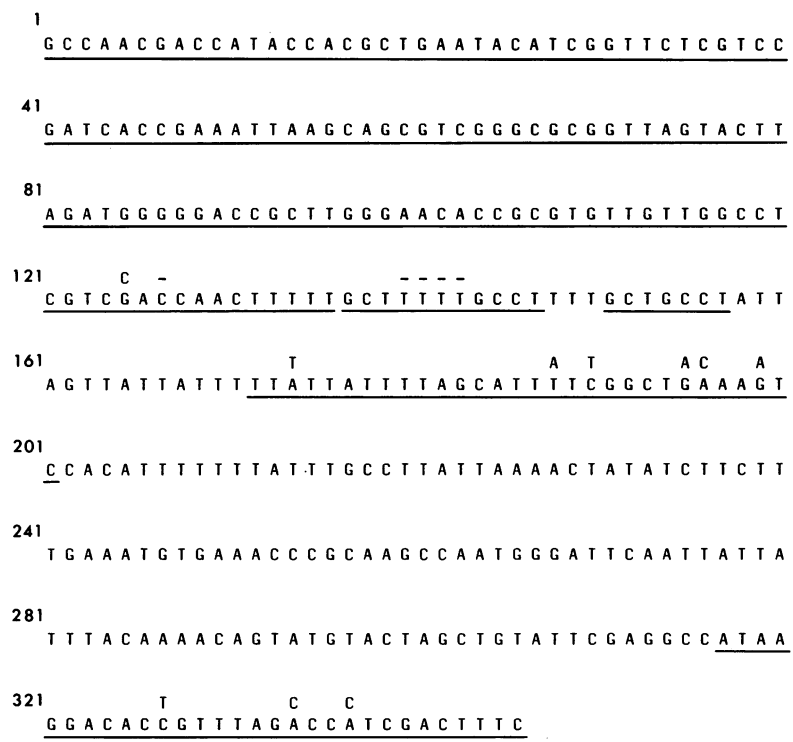

Figure 5 - Sequence of a 5S gene unit of D. teissieri (non coding strand). Nucleotides 1-120 : mature 5S RNA region, nucleotides 121-136 : precursor specific sequence, nucleotides 137-346 : spacer. Regions homologous to the D. melanogaster 5S gene are underlined. Differences with the D. melanogaster sequence are written only above otherwise homologous regions. Symbol - indicates that the nucleotide is missing in the D. melanogaster gene.

region (nucleotides 172-201, Fig.5). The other is directly adjacent to the 5'-end of the coding region (nucleotides 317-346, Fig.5).

5S genes of other drosophilid groups

The sizes of the 5S gene units in D. takahashii, D. virilis, D. nasuta and Zaprionus collarti are 400, 460, 460 and 375 base pairs, respectively. The digestion of genomic DNA from the last three species with Hae III does not produce 5S hybridization patterns with fragments of the same size as, or smaller than the 5S gene unit, but fragments of high molecular weight (Fig.6). There is thus no Hae III site in most of the 5S genes of these species. This result indicates that there is at least one sequence difference between the 5S RNA of these three species and the 5S RNA of D. melanogaster,

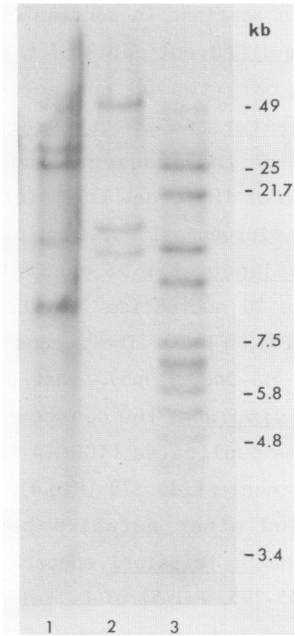

Figure 6 - Autoradiogram of embryonic Drosophila DNA digested with Hae III, fractionated on a 0.3 % agarose gel and hybridized to a ^{32}P-labelled 5S probe. Electrophoresis was carried out 3 days at 1V/cm in acetate buffer. DNA lanes : (1) D. nasuta, (2) D. virilis, (3) Zaprionus collarti. Molecular weight markers are indicated on the right.

D. simulans and D. teissieri since the Hae III site is located at the end of the coding region (23, nucleotides 116-119: Fig.4 and 5).

DISCUSSION

Evolution of the 5S genes in the D. melanogaster subgroup

The differences between D. melanogaster and D. simulans 5S genes are remarkably few. There are only 5 nucleotide substitutions between the two sequences, and a duplication of an heptamer in the D. simulans sequence (Fig.4). Tschudi and Pirrotta (23) detected 3 nucleotide variations in D. melanogaster 5S gene units present in a 20 units array. Differences between D. melanogaster and D. simulans 5S genes and differences between members of the D. melanogaster 5S gene family appear thus to be almost of the same order of magnitude.

The sequence of the D. teissieri 5S gene, when compared to its D. melanogaster counterpart, shows that the coding and non coding regions are evolving at different rates. The regions coding for the mature 5S RNA are

exactly the same in the two species, in contrast with the spacers which are very different (Fig.5). The different parts of the spacer are also evolving at different rates since only two 30 nucleotide stretches are clearly homologous in D. melanogaster and D. teissieri (Fig.5). These regions, located one at each side of the transcription unit, are most probably implicated in the function of the gene. It is interesting to note that the only region susceptible to micrococcal nuclease digestion in the B model of Louis et al. (25) for D. melanogaster 5S chromatin coincides with one of these two regions, i.e. the 30 nucleotide stretch located downstream from the gene (nucleotides 172-201, Fig.5). The second region, directly upstream from the gene (nucleotides 317-346, Fig.5), most probably interacts with RNA polymerase III (25). In D. simulans, the conserved segment which starts at nucleotide 170 contains the duplicated TTGGCTA heptamer and shows another difference at the level of nucleotide 178 (Fig.4).

It is possible to find other putative homologous regions to the melanogaster sequence in the D. teissieri spacer. For example, the sequence TACAAAACAGT (nucleotides 283-293, Fig.5) of D. teissieri is also found in the D. melanogaster sequence, but with a stretch of 5 rather than 4 A. The very high proportion of A and T residues in the spacers makes it difficult to ascertain whether these sequences are truly homologous.

The ACT-TA box described as highly conserved among polymerase III specific genes (26), is the same in D. melanogaster and D. simulans. It is located 45 base pairs upstream from the structural gene (ACT-GAA-TA, nucleotides 329-336, Fig.4). In D. teissieri, the middle portion of the box is very different (ACT-AGCTG-TA, nucleotides 298-307, Fig.5). An other box (ACT-GTT-TA, nucleotides 352-359, Fig.4) is present in the D. simulans sequence 22 nucleotides before the gene, but is not found in the D. teissieri sequence (Fig.5).

Some length heterogeneity has been demonstrated inside the 5S gene family of D. melanogaster (23). It is due to variable numbers (4 to 7) of a repetitive heptamer that follows the coding region. As shown in Fig.3, some variation in the length of the 5S genes also occurs in the other species of the subgroup. This heterogeneity occurs in individual 5S gene clusters since it was detected in cloned arrays of 5S genes in D. melanogaster (23) as well as in D. teissieri (our experiments). We do not know yet if different numbers of the heptamer account for this size variability in all species of the subgroup. This could be the case if one considers that the sequenced 5S unit of D. teissieri, which comes from the most abundant but smallest size class

of the D. teissieri 5S gene family (Fig.3) has only one copy of the heptamer (or two if taking account of the modified heptamer, Fig.5). If the size variations in Drosophila 5S gene families were always correlated with different numbers of the heptamer, this region could appear as a hot spot for the recombination events leading to concerted evolution. Brownlee et al. (27) have already proposed that the repetitive 15-oligomers present in the X. laevis 5S gene spacers could enhance crossing-over.

Phylogeny of the sibling species of D. melanogaster

Phylogenic relationships within the melanogaster species subgroup have already been proposed on the grounds of chromosomal inversions (28), ribosomal and histone gene structures (29,30), enzymes distribution (31) and satellite DNA homologies (32). All of these data show that the four species D. melanogaster, D. simulans, D. mauritiana and D. sechellia form an homogeneous class. Our mapping (Fig.2) and sequencing (Fig.4) results also prove these species to be closely related. The main difficulty in establishing phylogenies from our 5S gene data is the problem of homology between restriction sites located at approximately the same position. For example, are the Alu I sites present in D. teissieri, D. erecta and D. orena homologous , i.e. are their adjacent sequences homologous ? No answer can be given to this question without knowing the sequences. The only sequencing data we possess at the present time concern the Dde I sites located in the spacers of D. melanogaster and D. simulans 5S genes and the Taq I sites located before the 5S coding region in D. melanogaster and D. teissieri (Fig.2). The sequences of the genes (Fig.4 and 5) prove these sites to be homologous. Sequencing of the other species 5S genes will allow to propose a phylogeny for the subgroup members.

5S genes of other drosophilid groups

The D. melanogaster species group is composed of twelve subgroups (33). One of these is the melanogaster subgroup which includes the eight sibling species studied in this report. D. takahashii belongs to another subgroup of the melanogaster group. By restriction analysis, we found no differences in the 5S gene coding regions of all these nine species. In more distantly related species, such as D. virilis, D. nasuta and Zaprionus collarti which belong to other drosophilid groups, there is no Hae III site in the 5S coding region. This implies a different sequence for the 5S RNA of these species. A systematic search for the occurrence of the Hae III site in the 5S gene coding region could give interesting information about relationships among the drosophilid groups.

ACKNOWLEDGEMENTS

We thank Drs. L. Tsacas and J. David for providing Drosophila strains; Drs. J.C. Le Hégarat and J.A. Lepesant for providing phage and bacterial strains; Drs. V. Pirrotta and C. Tschudi for providing pBR7A plasmid. We are indebted to Dr. N. Randsholt for valuable comments on the manuscript. This work was supported by a grant from the C.N.R.S. (ATP Biologie Moléculaire du Gène).

REFERENCES

1. Erdmann, V., Huysmans, E., Vandenberghe, A. and De Wachter, R. (1983), Nucleic Acids Res., **11**, r105-r133.
2. Fedoroff, N.V. (1979), Cell, **16**, 697-710.
3. Brown, D.D., Wensink, P.C. and Jordan,E. (1972), J. Mol. Biol., **63**, 57-73.
4. Zimmer, E.A., Martin, S.L., Beverley, S.M., Kan, Y.W. and Wilson, A.C. (1980), Proc. Natl. Acad. Sci. USA, **77**, 2158-2162.
5. Smith, G.P. (1976), Science, **191**, 528-535.
6. Carroll, D. and Brown, D.D. (1976), Cell, **7**, 477-486.
7. Szostak, J.W. and Wu, R. (1980), Nature, **284**, 426-430.
8. Slightom, J.L., Blechl, A.E. and Smithies, O. (1980), Cell, **21**, 627-638.
9. Jackson, J.A. and Fink, G.R. (1981), Nature, **292**, 306-311.
10. Egel, R. (1981), Nature, **290**, 191-192.
11. Dover, G. (1982), Nature, **299**, 111-117.
12. Smith, G.E. and Summers, M.D. (1980), Anal. Biochem., **109**, 123-129.
13. Wahl, G.M., Stern, M. and Stark, G.R. (1979), Proc. Natl. Acad. Sci. USA, **76**, 3683-3687.
14. Vieira, J. and Messing, J. (1982), Gene, **19**, 259-268.
15. Miller, J.H. (1972), Experiments in Molecular Genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
16. Maxam, A.M. and Gilbert, W. (1980), in Methods in Enzymology, Vol. **65**, pp. 499-560.
17. Mandel, M. and Higa, A. (1970), J. Mol. Biol., **53**, 159-162.
18. Grunstein, M. and Hogness, D.S. (1975), Proc. Natl. Acad. Sci. USA, **72**, 3961-3965.
19. Maniatis, T., Hardison, R., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, G.K. and Efstratiadis, A. (1978), Cell, **15**, 687-701.
20. Karn, J., Brenner, S., Barnett, L. and Cesareni, G. (1980), Proc. Natl. Acad. Sci. USA, **77**, 5172-5176.
21. Birnboim, H.C. and Doly, J. (1979), Nucleic Acids Res., **7**, 1513-1523.
22. Jacq, B., Jourdan,R. and Jordan,B.R.(1977), J. Mol. Biol., **117**, 785-795.
23. Tschudi, C. and Pirrotta, V. (1980), Nucleic Acids Res., **8**, 441-452.
24. Michel, F., Jacquier, A. and Dujon, B. (1982), Biochimie, **64**, 867-881.
25. Louis, C., Schedl, P., Samal,B. and Worcel,A.(1980), Cell, **22**, 387-392.
26. Indik, Z.K. and Tartof, K.D. (1982), Nucleic Acids Res., **10**, 4159-4172.
27. Brownlee, G.G., Cartwright, E.M. and Brown, D.D. (1974), J. Mol. Biol., **89**, 703-718.
28. Lemeunier, F. and Ashburner, M. (1976),Proc. R. Soc. Lond.B., **193**, 275-294.
29. Tartof, K. (1979), Cell, **17**, 607-614.
30. Coen, E., Strachan, T. and Dover, G. (1982), J. Mol. Biol., **158**, 17-35.
31. Eisses, K.T., Van Dijk, H. and Van Delden, W. (1979), Evolution, **33**, 1063-1068.
32. Strachan, T., Coen, E., Webb, D. and Dover, G. (1982), J. Mol. Biol., **158**, 37-54.
33. Bock, I.R. (1980), Systematic Entomology, **5**, 341-356.