Vol. 166, No. 1, 1990 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
January 15, 1990 Pages 66-73

PRESENCE IN INVERTEBRATE GENOMES OF SEQUENCES CHARACTERIZED BY THE REPETITION OF THE TRIPLET CCPurine

Norbert BAKALARA, Joan COLLET, Richard PLANELLS, Yves THOUVENY and Michel FONTES

Laboratoire de Biologie de la Différenciation Cellulaire,
Faculté de Luminy, Case 901, 13388 Marseille Cedex 9, France

Received October 5, 1989

SUMMARY: In *Drosophila melanogaster* (Dm), polypeptidic domains have been found in different morphogenetic genes. Two types of them are characterized by the repetition of nucleotidic triplets: the M repeat $(CAX)_n$ and the paired repeat $(CAXCCX)_n$. In this paper we described a third type of repeat isolated from the genome of a Polychaete annelid: *Owenia fusiformis*. This repeat is characterized by the repetition of the triplet CCPurine. Phylogenetic studies showed the presence of this repeat in all the invertebrate genomes tested (eight copies in Dm genome) while we failed to detect it in vertebrate genomes.

Recent studies on gene structure have clearly demonstrated that a great number of genes can be subdivided into functional domains that are subjected to recombination and independent assortment (1,2). Thus any multifunctional gene might share sequences with several gene sets (3). A typical example is provided by the association of different domains in homeotic genes (4,5,6). Several domains have already been described, among which the first one was the "homeobox" (4,7). Then, other domains constituted by highly repeated sequences were evidenced, mainly in the *Drosophila melanogaster* genome:

The M repeat or OPA repeat formed by the repetition of CAX (4,8) was present in a great number of copies.

The "paired" repeat, present in ten genomic regions, was characterized by the repetition of CAXCCG. This one was detected in two segmentation genes, "paired" and "gooseberry" (9). Finally, a GCX and GAX repeats were described in "engrailed" (10).

Previously, Lewis (11) observed in his genetic study of *Drosophila melanogaster*'s bithorax complex, that such complex genes had probably evolved from a few ancestral genes. According to the earlier statements, morphogenetic genes might have evolved by different recombination of a few ancestral domains. Therefore it is interesting to follow the phylogenetic distribution of these domains as previously done for the homeobox (4).

To look for sequences similarity between morphogenetic genes of *Drosophila melanogaster* and genomic regions of more primitive animals, we decided to screen a genomic library of *Owenia fusiformis*. This one is a marine Polychaete annelid and one

0006-291X/90 $1.50

among most primitive metamerized animal living nowadays. We used as probes two distinct sequences from the Antennapedia complex: "Antennapedia" (Antp) and "Fushi-Tarazu" (Ftz) (12).

While Antp did not give any hybridization signal, Ftz allowed us to clone a new type of repeat characterized by the repetition of the triplet CCPu and its association to a putative double α helix protein structure. Phylogenetic distribution of this sequence was studied.

MATERIALS AND METHODS

Owenia fusiformis were obtained from Station Marine de Roscoff (France).

Probes were: i) a 650 bp Bam HI-Pvu II fragment from p903G clone (13) which contained a part of complementary DNA to an Antennapedia gene transcript (further refered as Antp). ii) a 1.1 Kb Pvu II fragment from pFS2, a genomic clone from the Fushi-Tarazu gene (6)(further refered as Ftz).

The genomic library was constructed by one of us (M.F.) in Dr. V. Pirotta's laboratory at the E.M.B.L. (Heidelberg, F.R.G.). This was performed by inserting partial Sau3A digestion fragments of *Owenia fusiformis* genome into bacteriophage L47 (14). This library was screened with [^{32}P]labelled, nick translated probes, under classic conditions. Nitrocellulose filters were washed in 2xSSC buffer, at 42°C. DNA of hybridizing clones was prepared and subcloned into pUC 18 and 19, as well as into M13 mp18 and mp19, according to standard methods (15). Sequencing was performed according to Sanger's method (16) and genomic DNA preparation according to Mc Ginnis' (17).

RESULTS AND DISCUSSION

We screened a genomic library of *Owenia fusiformis* with Ftz and Antp as probes. Spreading more than three genomes, we could isolate six recombinant clones, named λ1 to λ6. We found by restriction mapping and hybridization analysis that λ1 and λ6, as well as λ2 and λ5 were identical. Each of the four remaining clones exhibited two Hind III fragments able to be recognized by Ftz probe, under our stringency conditions (0.5 x SSC, 50°C). Both fragments hybridized each other. Moreover, from one clone to another, these fragments could cross-hybridize. Thus, we focused our work on one of them, namely λ1. Its restriction map is presented on Fig. 1. We subcloned in pUC 18 the two hybridizing fragments, the 1.8 Kb Pst I-Hind III one and the 1.8 Hind III-Hind III one. These subclones were respectively named PH 18 and H 18. The major part of the PH 18 sequence is presented on Fig. 2. We noted the presence of 49 triplets CCPu. This repeat was surrounded by AT-rich regions. Sequence of H 18 displayed a similar CCPu, although shorter (25 triplets)(Fig. 3); This second repeat was also surrounded by AT-rich regions. Comparison of these repeats showed a high similarity percentage between them: among 68 nucleotides, only 5, all located at the Pu position, were differing. In the *Drosophila melanogaster* genome, previously described repeats (M repeat = CAX, and "paired" repeat = CAXCCG) exhibit some analogous features. They are GC-rich and the third variable base of the triplet is a Purine base. Concerning Ftz that we used as probe, this gene displays a CCX region (extending from nucleotide 1167 to nucleotide 1244) (6) which does not constitute a monotonous repetition. Nevertheless we compared this region from Ftz with

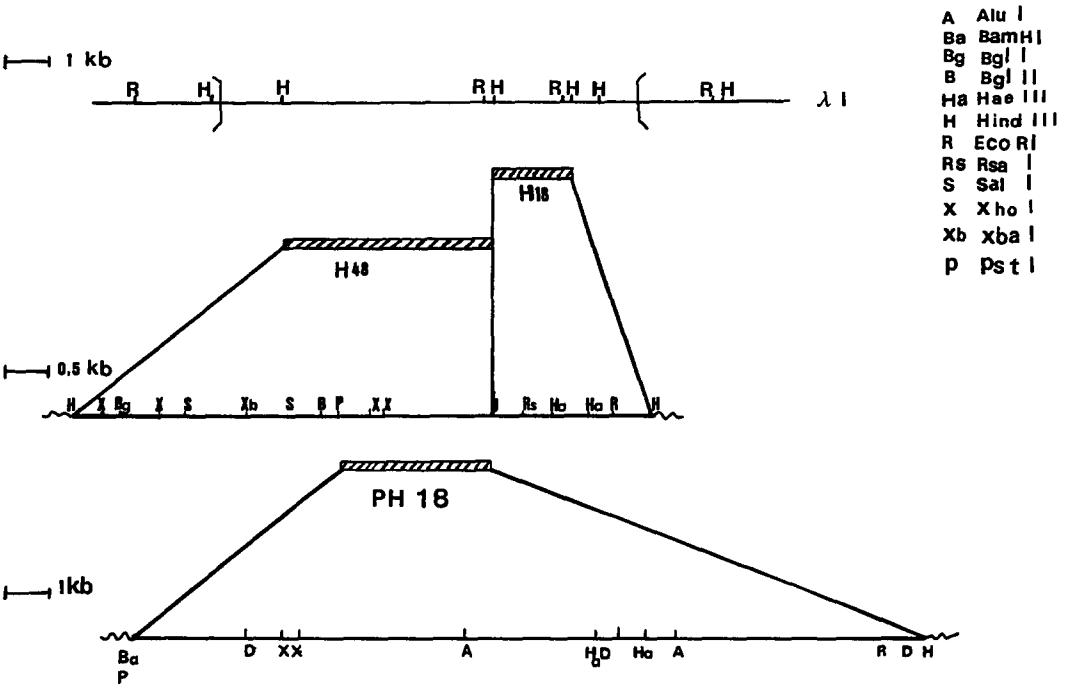

FIGURE 1: RESTRICTION MAP OF λI AND SUBCLONES.
The restriction map was established with the following enzymes: Alu I (A), BamH I (Ba), Bgl II (Bg), Dra I (D), Hae III (Ha), Hind III (H), EcoR I (E), Rsa I (R), Sal I (S), Xho I (X), Xba I (Xb) and Pst I (P). The position of the different fragments subcloned into Puc vectors (H18, PH18 and H48) is indicated below the restriction map by dashed boxes. These boxes indicated the regions of homology between the λI fragments and Ftz. Brackets indicate the limits of the insert.

the CCPu sequence, (Fig. 4). We found two regions of high similarity. The richness in GC content of these similar regions was likely to account for the stabiliy of the hybridization that we observed between Ftz and PH 18. Furthermore it might indicate the presence in the Ftz gene of an ancestral CCPu repeat which could have diverged during evolution.

In a further step, we looked for putative coding sequences, and found in pH18 two open reading frames (ORF). The first ORF (ORF1) included the CCPu repeat and encoded for a polypeptidic chain partly formed by the repetition of amino-acid Proline. The second ORF (ORF2) encoded for a highly basic polypeptidic chain. If these two ORFs are present on the same transcript, then the CCPu repeat should be read as a polyproline peptide. This assumption comes from the observation that, these two ORFs are separated by two intron consensus (18)(Fig. 5a). The putative secondary structure of the polypeptide possibly encoded by ORF2 was examined by Garnier's method (19). According to it, potential existence of two α-helix separated by a short β-turn was evidenced. (Fig. 5b). Such a structure has previously been described in several DNA binding proteins (for instance Mat α1 and α2, λ cro and rep, homeobox containing genes, ...)(6,20). To go further in the interpretation of this data, it would be

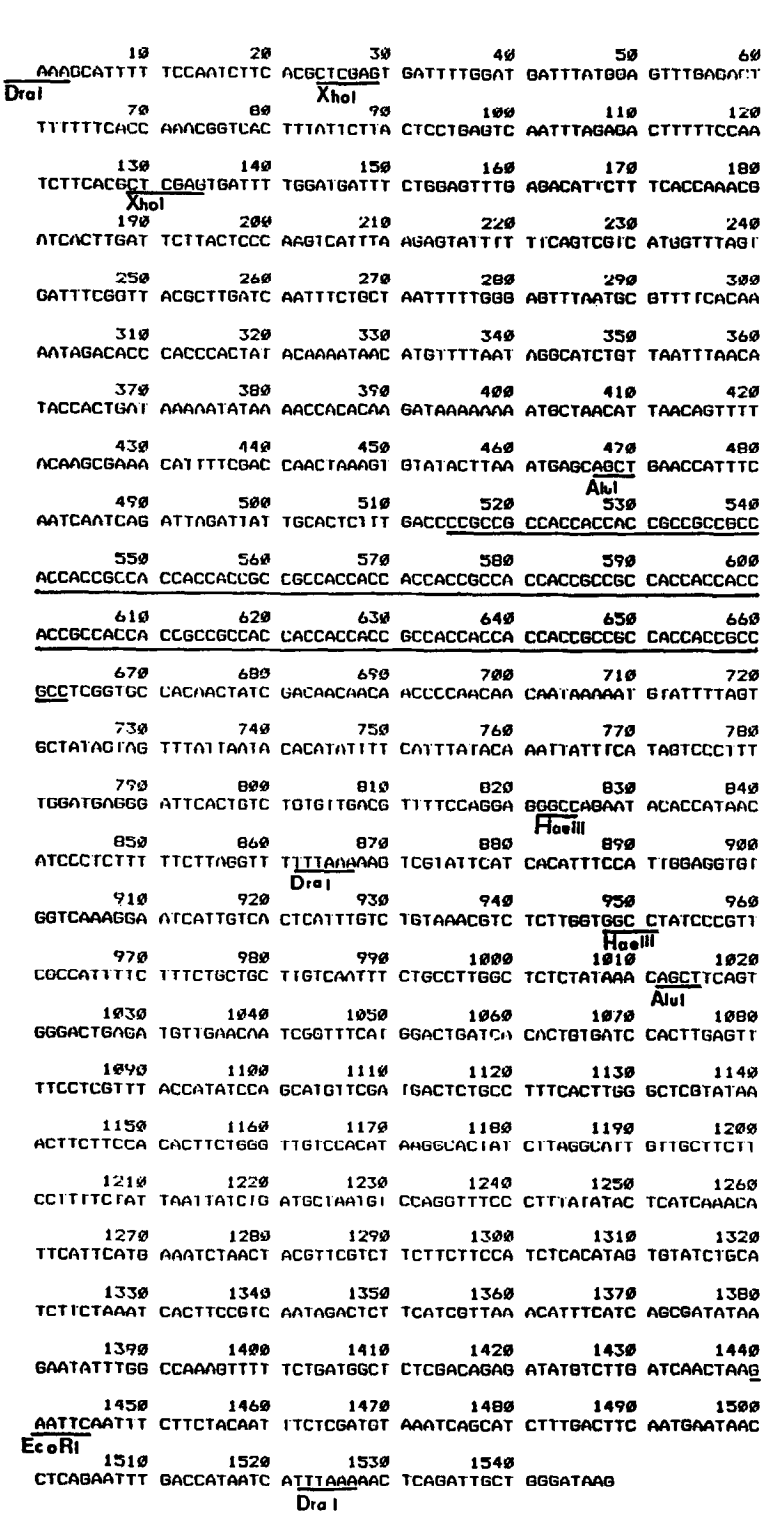

FIGURE 2: NUCLEOTIDE SEQUENCE OF PH18.

Nucleotide sequence of PH18 is shown between the first Dra I site and the Hind III site. Restriction sites are indicated and the CCPu repeat underlined.

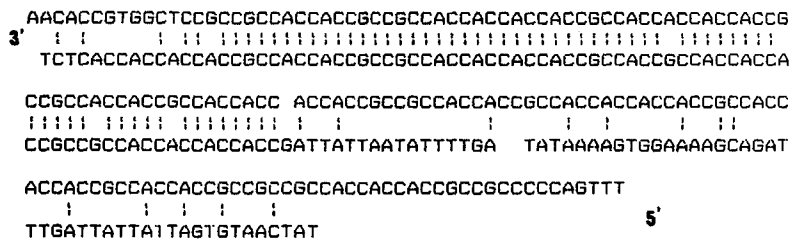

FIGURE 3: COMPARISON BETWEEN PH18 AND H18 REPEATS.

interesting to determine the existence of this polypeptide *in vivo* and look for its putative DNA binding function.

To determine the occurrence of the CCPu repeat in owenia genome, genomic Southern blots were undertaken, by using DNA from a single individu. As seen in Fig. 6, ten bands hybridized under high stringency conditions with A4 (Fig. 1) as probe. Assuming two repeats by regions (as demonstrated by the hybridization pattern of all clones we isolated), we may conclude that CCPu repeat is located in 4 to 5 distinct regions

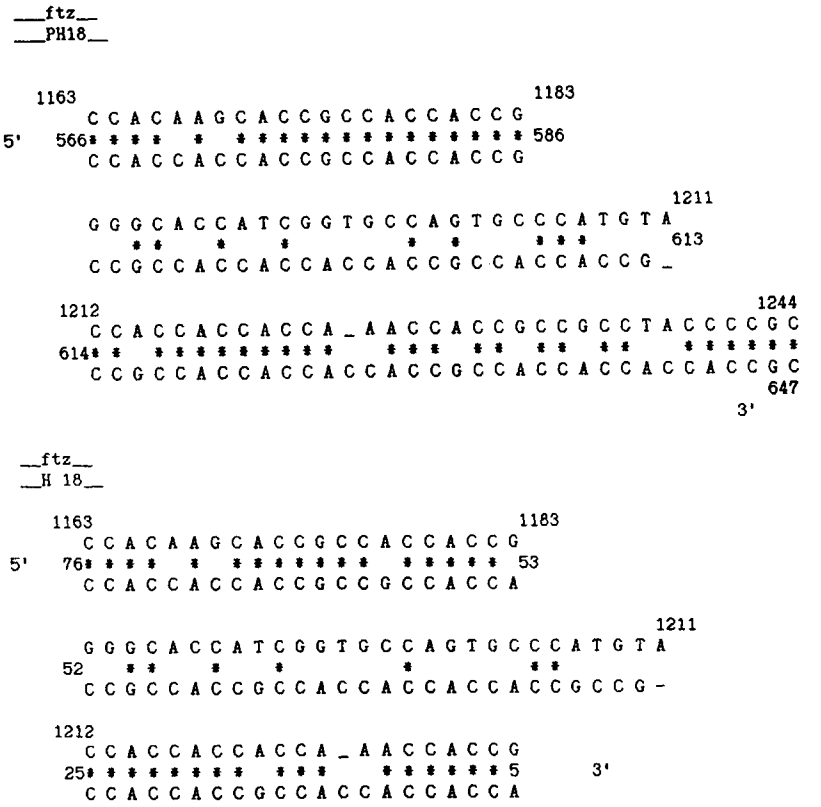

FIGURE 4: SEQUENCE HOMOLOGY BETWEEN H18, PH18 AND FTZ.
Numbers indicated above Ftz sequence refer to the Laughon and Scott's paper (6). Numbers indicated above H18 and PH18 sequence refer to Figs 2 and 3 of the present paper.

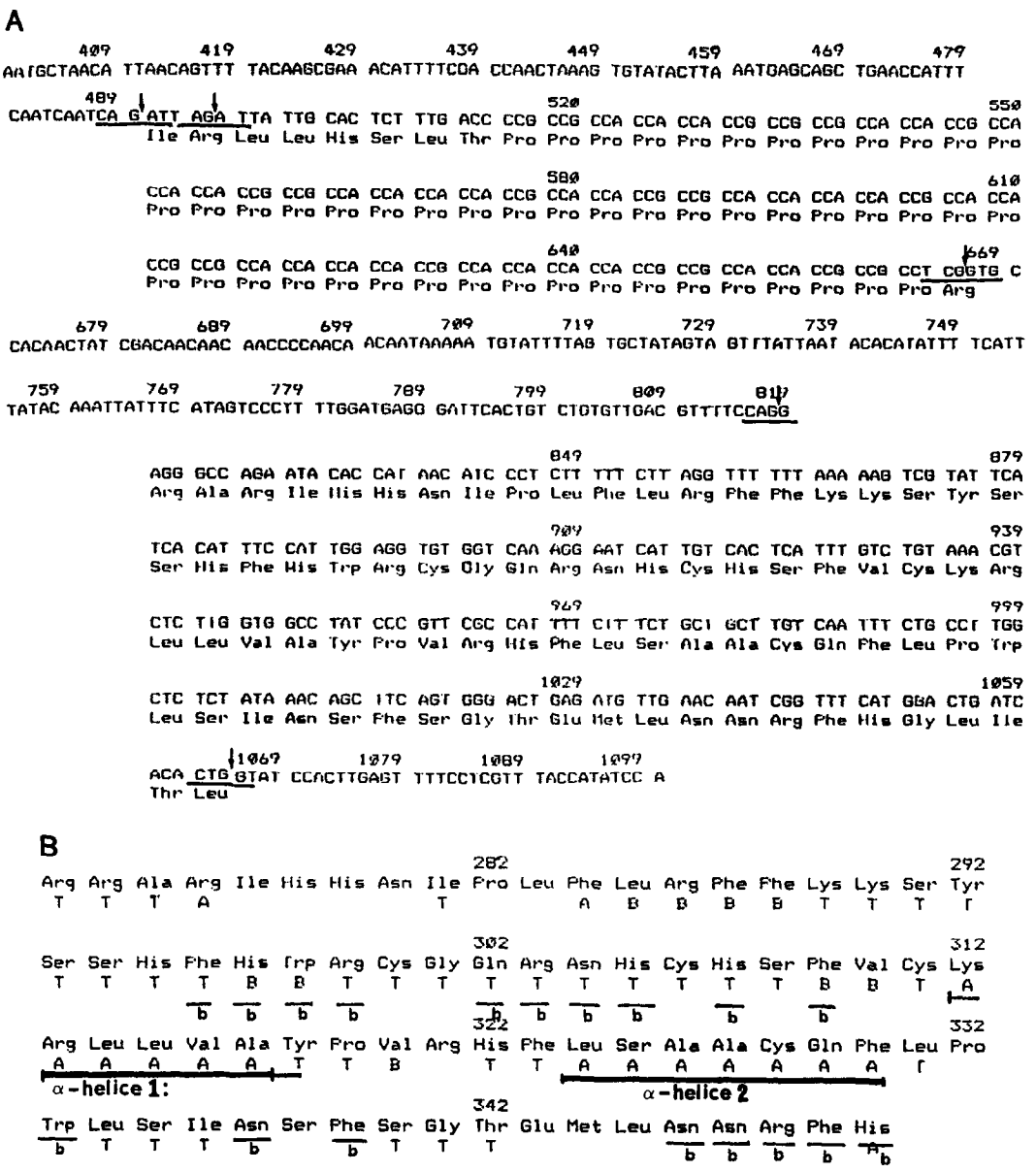

FIGURE 5: POTENTALLY CODING REGIONS OF PH18.

(A). Translation into amino-acids of the unique potentially coding region of PH18 (Intron consensus sequences are underlined and break points indicated by an arrow). Numbers above the sequence refer to the nucleotide sequence presented in Fig. 2. (B). Possible secondary structure of the second open reading frame: when a letter is indicated under an amino-acid, A means that the amino-acid is implicated in a α-helix structure, B in β-sheets and T in turns, as displayed by a Garnier's computer analysis of the DNA sequence.

of the *Owenia fusiformis* genome. Therefore we can conclude that the CCPu repeat is closer in its genomic distribution to the "paired" repeat than to OPA repeat.

In order to determine the phylogenic distribution of the CCPu repeat, DNA from several species was prepared and submitted

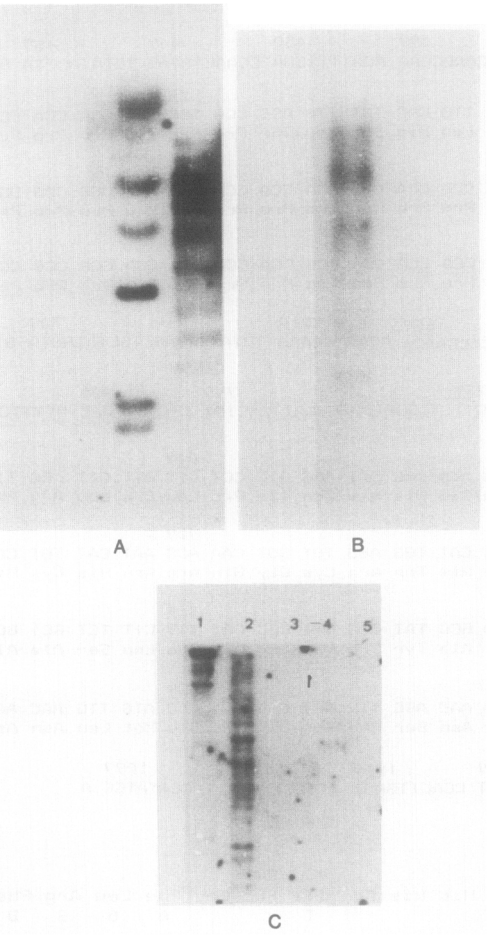

FIGURE 6: DETECTION OF CCPu REPEAT RELATED SEQUENCES IN THE GENOME OF VARIOUS ANIMALS.

Blots of Eco RI-digested DNA from *Owenia fusiformis* (A), *Drosophila melanogaster* (B), *Ciona intestinalis*, a tunicate (C, lane 1), Amphoxius (C, lane 2), Xenope (C, lane 3), Mouse (C, lane 4) and Man (C, lane 5) were hybridized with the Alu I fragment of PH18 (see Fig. 2) as probe. This CCPu repeat-containing fragment had been [^{32}]P-labelled by oligopriming. Blots were washed by 0.1xSSC, at 68°C and exposed for 12 hours. Hind III digested λ DNA was used as a size marker. 0.5 to 1 µg of Owenia DNA, 2 µg of Drosophila, Ciona and Amphoxius DNA and 10 µg of Xenope, mouse and man DNA, have been deposed. All DNAs were extracted from a single animal.

to hybridization with A4 as probe. Results are shown on Fig. 6. All tested invertebrate genomes displayed a comparable number of bands (about 10). The greater number of bands exhibited by Amphoxius pattern is likely to be due to the triplication of the genome between Urochordates and Cephalochordates, as proposed by Atkin and Ohno (21). Concerning the hybridization pattern of *Drosophila melanogaster*, we observed under our washing conditions (0.1 X SSC, 68°C) 7 to 8 bands. Nevertheless, none of them corresponded to Fushi-Tarazu. Therefore we have to assume that A4 displayed with some drosophila genomic regions a closer similarity that it did with Fushi-Tarazu gene. Surprisingly it

should be noted that no vertebrate genome displayed bands, even after a longer exposure of the film (10 fold). As a control, an actin probe was successfully hybridized, under the same stringency conditions (data not shown). This result cannot allow us to know whether the CCPu repeat has either diverged or is absent in vertebrate genomes.

In conclusion, we cloned, from the genome of an annelid, regions formed by the repetition of CCPu triplet. Like the "paired" repeat, this new repeat is only present in a few copies in the annelid genome, as well as in other invertebrates genomes. This sequence is absent or divergent in vertebrates. The biological role of this repeat might be enlighted by the isolation in the *Drosophila melanogaster* genome, of one gene which contained it. Moreover, because of it structural analogies with the "paired" repeat, our sequence should be checked for its involvement in traumatic regeneration of this annelid. This biological system provides an alternative model for the study of morphogenetic events (22).

Acknowledgments: We are indebted to Doctors M. Akam and S. Kerridge for providing us with probes and for helpful discussion. We are thankful to Dr. V. Pirrotta for giving us the opportunity to construct the Owenia's library in his lab.

REFERENCES

1. Gilbert,W. (1978) Nature *271*:501-505.
2. Gilbert,W., Marchionni,M. and Mac Knight,G. (1986) Cell *46*:151-154.
3. Frigerio,G., Burri,M., Bopp,D., Baumgartner,S. and Noll,H. (1986) Cell *47*:735-746.
4. Mac Ginnis,W., Levine,M.S., Hafen,E., Kuroiwa,A. and Gehring,W.J. (1984) Nature *308*:428-433.
5. Regulsky,M., Harding,K., Kostriken,R., Karch,F., Levine,M.S. and Mac Ginnis,W. (1985) Cell *43*:71-80.
6. Laughon,A. and Scott,M.P. (1984) Nature *310*:25-31.
7. Mac Ginnis,W., Garber,R.L., Wirz,S., Kuroiwa,A. and Gehring,W.J. (1984) Nature *311*:403-408.
8. Wharton,K.A., Yedvobnick,V.G.F.and Aatavinis-Tsakona,S. (1985) Cell *40*:55-62.
9. Bopp,D., Burri,M., Baumgartner,S., Frigerio,G. and Noll,M. (1986) Cell *47*:1033-1040.
10. Poole,S.J., Kauvar,L.M., Drees,B. and Kornberg,T. (1985) Cell *40*:37-43.
11. Lewis,E.B. (1978) Nature *276*:565-567.
12. Scott,M.P., Weiner,A.S., Hazelrigg,T.I., Polisky,B.A., Pirrota,V., Scalenghe,F. and Kaufman,T.C. (1983) Cell *35*:763-766.
13. Garber,L.G.R., Kuroiwa,A. and Gehring,W.J. (1983) EMBO.J. *2*:2027-2037.
14. Loenen,W.A. and Brammar,W.J. (1981) Gene *10*:249-251.
15. Maniatis,F., Fritsch,E.F. and Sambrook,J. (1982) *in* Molecular Cloning - A Laboratory Manual, Ed. C.S.H. Lab. Public., Cold Spring Harbor, USA.
16. Sanger,F., Nicknen,S. and Coulson,A.R. (1977) Proc. Natl. Acad. Sci. USA *74*:5463-5468.
17. Mac Ginnis,W., Shenmoen,A.W. and Beckendorf,S.K. (1983) Cell *34*:75-84.
18. Mount,M.S. (1982) Nucl. Ac. Res. *10*:459-470.
19. Garnier,J., Osguthorpe,D.J. and Robson,B. (1978) J. Mol. Biol. *120*:97-120.
20. Pabo,C.U. and Sauer,R.T. (1984) Ann. Rev. Biochem. *53*:293-321.
21. Atkin,N.B. and Ohno,S. (1967) Chromosoma *23*:10-13.
22. Thouveny,Y. (1967) Arch. Zool. Exp. et Gen. *108*:347-386.