

---

**tRNA derived insertion element in histone gene repeating unit of *Drosophila melanogaster***

---

Yoshinori Matsuo\* and Tsuneyuki Yamazaki\*\*

---

Laboratory of Population Genetics, Department of Biology, Faculty of Science, Kyushu University, Fukuoka 812, Japan

---

Received September 7, 1988; Revised and Accepted November 23, 1988

---

**ABSTRACT**

Analysis of 41 histone homologous clones from an isogenic gene library of *Drosophila melanogaster* showed that non-histone fragments interrupt the histone repetitive clusters at several sites. Long (L) and short (S) forms of the repeating units are distinguished by the insertion of 240 bp into the spacer between H1 and H3 of the L units; Each form appears to be clustered with its own kind. The complete DNA sequence of the histone 5.0 kb repeating unit was determined. Five histone genes (H1, H2A, H2B, H3, H4) were identified in a repeating unit and several sequence blocks common to the five histone genes were found in the 5'- and 3'-regions. The insertion sequence of 240 bp was found to be similar to the Alu family, an element derived from tRNA.

**INTRODUCTION**

There are two types of histone gene families, a tandem cluster type and a dispersed type. The former is found in *Drosophila* (1, 2) and the latter, in chicken (3, 4, 5), mouse (6, 7) and human (8-10). In some species, such as *Xenopus* (11-17) and sea urchin (18-22, 61-62), both types are present together. The order of histone genes in a repeating unit, polarity of transcription and copy number often differ among species (2, 23-25). The orientation of transcription in the sea urchin is unidirectional in all five early histones (H1-H4-H2B-H3-H2A) (20), but not in yeast (H2A-H2B, H3-H4) or *Drosophila* (H1-H3-H4-H2A-H2B). In the latter species, H2A and H2B, H3 and H4 are transcribed in opposite directions (1, 26, 27). The number of copies in several species has been estimated as: 1-2 in yeast (26, 27), about 10 in chicken (5), 20-60 in *Xenopus* (12, 13), 110 in *Drosophila* (1, 28) and several hundred in the sea urchin (20, 22, 23).

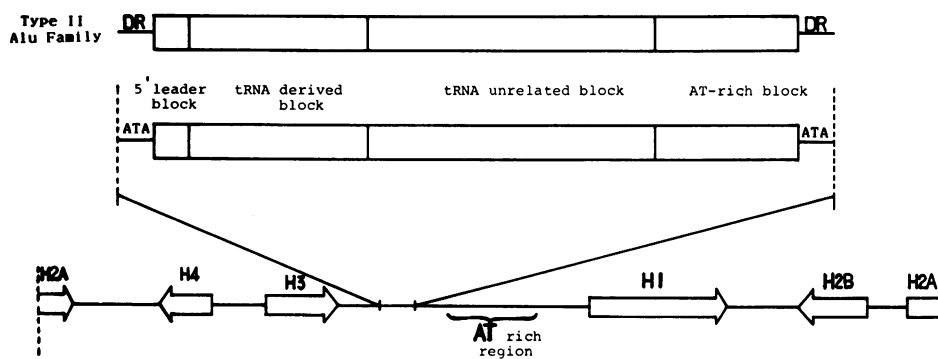


Fig. 1. The structure of the type II Alu family and insertion of histone gene are shown along with the location of insertion sites.

The histone genes of *D. melanogaster* repeat tandemly in the 39D-E region of chromosome IIR (28), where two kinds of repeating units, L (5.0 kb) and S (4.8 kb) are known to be present (29). The L unit has about 240 bp inserted between H1 and H3 (29) (see Fig. 1). Saigo et al. (30) analyzed the organization of the histone gene family by two dimensional electrophoresis and found that both of these units were clustered. Variant types also exist at low frequency (30). For instance, transposon 297 (a copia-like element) is inserted in the TATA box of H3 (31). As for the structure of the repeating unit, Goldberg (32) determined about 70 % of the DNA sequence of the S repeating unit, but the 3'-coding region and intergenic spacer have yet to be characterized completely.

In this study, restriction enzyme mapping and histone homologous fragments were determined for forty-one histone positive clones from the lambda library of a *D. melanogaster* isogenic strain. The organization of the histone gene family in a genome was deduced. The complete DNA sequence of the histone L unit was determined so as to examine the structure of histone genes.

## MATERIALS AND METHODS

### *D. melanogaster* library

To study the organization of the histone gene family in a chromosome, the isogenic strain, AK-194, was used. AK-194 was

previously constructed by extracting both the second and third chromosomes simultaneously from an isofemale line using a balance-lethal system (33). The gene library was constructed by either EcoRI partial digestion using a charon 4 vector (34) or MboI partial digestion using a  $\lambda$ EMBL4 vector (35). In order to minimize the production of the artificially ligated DNA of small size (<10 kb) the DNA with the molecular weight of 12-20 kb was purified by sucrose gradient centrifugation. In addition, the *Drosophila* DNA was treated with alkaline phosphatase prior to ligation to vector DNA. Plaque hybridization was performed to screen the histone genes (36).

#### Cloning

DNA of a plasmid or phage was prepared by the Alkali-SDS and liquid culture methods, respectively (37). Digested DNA was ligated into the poly-linker site of a pUC9 plasmid (38). Transformation was conducted according to the  $\text{CaCl}_2$  method (39) using the strain JM83 or TB1.

#### Labelling of DNA

An EcoRI fragment of pKSL100 (histone 5.0 kb unit of *D. melanogaster*) was labelled by nick translation (40) using  $\alpha$ - $^{32}\text{P}$ dCTP.

#### Southern blotting

Digested DNA was separated by agarose gel electrophoresis and transferred to a nylon membrane filter (Pall) by the method of Southern (41). Hybridization was conducted with a histone gene probe for 24-36 hrs at 68°C (37). The filters were washed with 6xSSC once, 4xSSC twice, and 1xSSC once at 68°C.

#### DNA sequencing

We determined the DNA sequence by dideoxy chain termination (42, 43) using a denatured plasmid as the template (44). Takara sequencing kit and Amersham's universal primer or oligonucleotide primer (17mer 5'-GCGATGACGCTTGGCG-3') were used for the sequencing reactions.

### RESULTS

#### Organization of the histone gene family in *D. melanogaster*

The organization of this family in *D. melanogaster*, that is, the distribution and structure of members of this family in the genome, was deduced by analyzing many independent histone clones

---

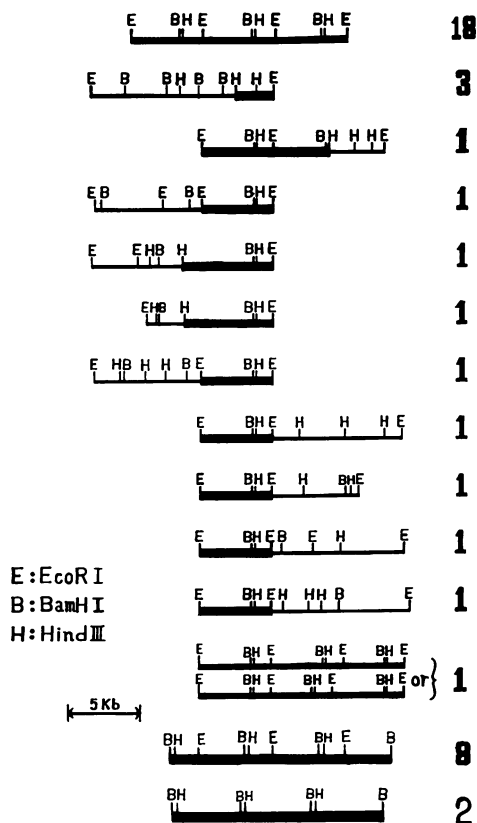


Fig. 2. Restriction mapping of 31 clones from the AK-194 EcoRI library and 10 clones from the BamHI library. In the right column is shown the number of clones with the same mapping patterns. The thick line shows the histone repeating homologous fragments.

from an isogenic strain (AK-194). To avoid biased selection of a cloning enzyme, thirty-one and ten clones were obtained from EcoRI and BamHI libraries, respectively. In Fig. 2, the histone homologous repetitive fragments (thick line) with and without the EcoRI site are considered to be the histone L (5.0 kb) and S (4.8 kb) units, respectively. As expected from the tandemly clustered structure, most of the clones (28 out of 41: 18 clones of 3L units from the EcoRI library, 8 clones of 3L units from the BamHI library and 2 clones of 3S units from the BamHI library) each had 3 units of the same length in tandem, indicating that each unit was grouped with its own kind. This is consistent with the

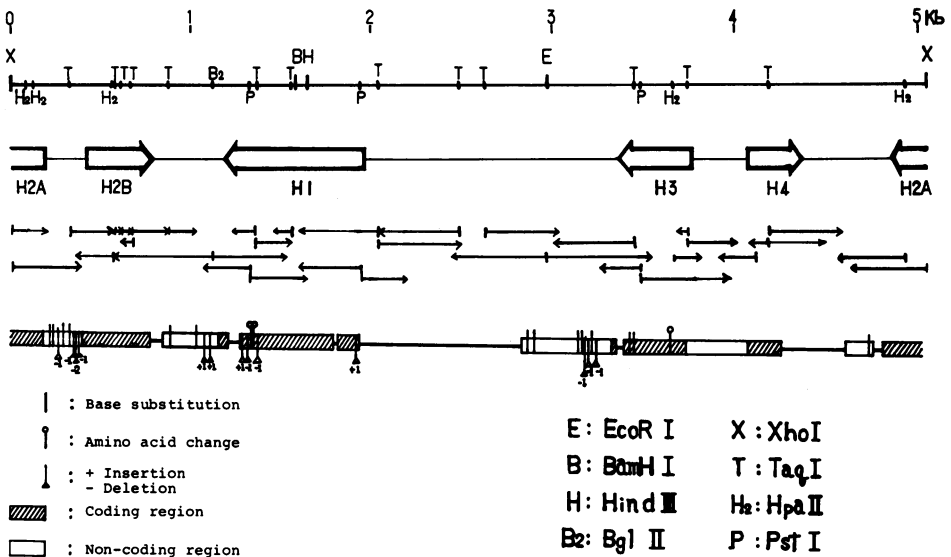


Fig. 3. Sequencing strategy of the histone 5 kb unit. DNA sequences were determined mainly from PstI, TaqI, and HpaII sites. Boxed regions at the bottom of the figure indicate those sequenced previously by Goldberg (32).

conclusion arrived at by a different method (30). Twelve of the remaining clones, however, had histone non-homologous fragments as well as one or two histone units. Although at most two different clones of these fragments may be at the ends of a histone cluster, some may be situated in the middle of histone loci. This is because the histone genes were found to be located in the 39D-E region of chromosome IIR by *in situ* hybridization (28). Thus, some parts of the histone gene cluster may possibly be interrupted by non-histone DNA. Southern blotting experiments of genomic DNA digested with BamHI showed the presence of histone homologous fragments with various sizes when blotted against the histone probes, confirming the above interpretation about the structure of histone gene families (data not shown).

Complete DNA sequence of the 5.0 kb histone gene repeating unit

One L unit was cloned (AK-194-19) and its DNA sequence was determined. The sequencing strategy and 5041 bp sequence determined are shown in Figs. 3 and 4, respectively. Five histone coding regions were identified within a repeating unit by

Nucleic Acids Research

computer analysis and amino acid sequence (45-48).

Units of the histone repeats have 200-1400 bp of nontranscribed spacer sequence. The largest spacer between H1 and H3 is 1400 bp and contains the AT-rich region (about 77 % of

Table with 10 columns representing positions 10-110. Each row contains a sequence of amino acids and their corresponding three-letter codes. The sequence is highly repetitive, showing variations in the order and identity of amino acids across the different units.

	2540	2550	2560	2570	2580	2590	2600	2610	2620	2630	2640
	TAATAATA	ATAATAAT	TCTTAATA	TTAATGTG	TCGTTGTT	TAATGTGT	TCATGCTAT	TTTAAATA	TTTGTCTGA	AATCAACT	TTCTCTGC
	ATTATATT	TTATTATA	AGAATTA	AATTACCA	ACGAACA	ATTACACAA	AGTACGTA	AAATTTTAT	AAACAGACT	TTAGTTAGC	AAAGGACCG
	2650	2660	2670	2680	2690	2700	2710	2720	2730	2740	2750
	ACATATGC	TACCGCTA	GTAQCAAT	AAATAAAT	AAATAAAAT	AAAAGAAG	AATTTTGT	TTATCTTA	TTATGTA	TATATCAT	TCGGCAACA
	TGATACGA	ATGCGAGTA	CATCAGTT	TTTATTAG	TTTTCCTG	TTTTCCTG	TTTAAACAT	AATGAGAA	AATACATAT	ATAAATGTA	AGCGCTTGT
	2760	2770	2780	2790	2800	2810	2820	2830	2840	2850	2860
	AAATACCA	TTTCGTGT	GAATTTAC	TAGGTTAT	TTTTTACG	CGTTGAAG	GACAGTGC	ATTGTCGG	ACGACCTCT	CAATAAAC	ACATCTTCA
	TTAATCGT	AAAGGCAG	CTAAATAT	ATCCAAAA	AAAAATGT	GCAAACTTC	CTGTACAG	TAAACGGCA	TCTGTGAGA	GTTATTATG	TGTAAGAAT
	2870	2880	2890	2900	2910	2920	2930	2940	2950	2960	2970
	GTTAACAT	GAAAAATA	TAAATAG	AGTTTAT	TATTTAT	TGTTATTT	TAAATATG	CGTATCAT	GTAATTTAT	TATTTTCAA	CTGTCTTCA
	CAATGTGT	CTTTTTAT	ATTTTATG	TCAAATAA	ATAATAAA	ACAATAAA	ATTATATG	CGAGTAGTA	CTTAAGATA	ATTAAATTA	CTGTCTTCA
	2980	2990	3000	3010	3020	3030	3040	3050	3060	3070	3080
	CAGAAATG	ATTTTTTT	TAGATGTA	AGGTTACG	TCCTTGCC	CGCACGAC	TTTACTGT	TTAAAGTA	TTCCGCAAA	GAACATAA	ATATATAT
	GCTTAAGT	TAAAAAAA	ATCTACAA	TCCCAAGT	AGGGAAGG	GGTGTGTT	AAATTTATG	AATTTGTAC	AAAGCTTGT	GCTTTTAT	TATATAT
	3090	3100	3110	3120	3130	3140	3150	3160	3170	3180	3190
	CTTGTATTA	ACCTCTCT	TAAATTA	AATGTGAG	TGACAATA	AATAAGATA	TTTTTATG	TTTCTGCA	GAATTAAT	TTTACATAT	GATTTAAGC
	GAACATAT	TGGGAAGA	ATTAATAT	TTACACTG	ACTGTAAT	TTATTTAAT	AAAAATGG	AAGTAGCGT	CTTAATGTA	AAAATGATA	CTAAATCTG
	3200	3210	3220	3230	3240	3250	3260	3270	3280	3290	3300
	CGCATAGC	TTGGCGTA	AATAATGA	TATGAATT	GTTCATTT	TTTTTAAT	AFAAATGT	GGCAACAGA	AATTTATCT	CATTTAGCT	GGTTTGTGT
	GGCATAGCA	ACCGCGTA	TTTATATCT	ATVCTAAA	CAAGAATA	AAAAATTT	TTTTTACG	CCGTTGCT	TTAAATATG	GTAATTTGA	CCAACAACA
	3310	3320	3330	3340	3350	3360	3370	3380	3390	3400	3410
	CCTGAAGC	ACCGAATA	GACTAGCT	GGCTTAT	TGAAGTTA	TCCGTTGT	GCTTATG	GCTCGGCG	AATGCGTC	GCTACTGA	TGCTTTTGG
	GGACTTTTC	TGCTAATA	CTCATGCT	CGGCAATG	AGTTTAA	ATGGCACT	CGAATGTC	CGAGCGCG	TTACGAGCG	CGATTGACT	ACAGAATCC
	3420	3430	3440	3450	3460	3470	3480	3490	3500	3510	3520
	CATTATGTC	ATACCGTT	CATGAAGG	ACACAAGT	GTATCTGA	AGAGCAAC	CAGGTAGCT	TCGCTAGCT	CCTGCAGGC	CATAACCGC	GAGCTCTGA
	GTAATATG	TATGGAGC	GCTCTAGC	TCTGTTAC	CATTAAGT	TCTTGTGT	GTCGATGTA	AGCATGTA	GGACTCTCG	GATTTGAGC	CTTGAGCT
	MetIleThrLys	AlaHisIleAla	LysLeuSerHis	ArgPhe	LeuGlyVal	LeuArgLys	AlaGlu	GlnLeu	AlaMetVal	AlaSer	SerGlnPhe
	3530	3540	3550	3560	3570	3580	3590	3600	3610	3620	3630
	ATCCGAAT	CGCTTAA	TGGTAGCA	TTTACAGC	CAGACGTC	AAAGCGCT	TGGGTTAT	AACTCGGT	CTCTTTGT	ACGGCAAT	TTACCGCA
	TAGCTTCA	CGAGAATT	ACTGCTCT	AAATGCTG	GTCTGATC	TTTTCTGA	ACGCTAAT	TTGGAGCG	GAGAAACCA	TCGCTGTT	AACTGCTCT
	ArgLeuSer	ThrLysPhe	AspGlnAla	IleGlu	ArgValLeu	ArgGlnPhe	ProLeuLys	ArgIleLeu	GluThr	ValArg	ArgLysLeu
	3640	3650	3660	3670	3680	3690	3700	3710	3720	3730	3740
	GCACGGTC	CAGGGGTA	CGGTGGGC	TTCTCAC	ACTCCGTC	TGGAGCTC	TTGGCAGCG	CCCTTAGT	CAGTTTGT	CGTGGGCT	TTCCACCT
	CGGTCGCA	GTCGCTAT	CGCCACCC	AGAAAGTG	GAGGCAAC	ACTCTGTG	AAAGCTCG	CGAATATG	GTCACAAA	CGACGAGCA	AAAGTGTTA
	AlaVal	ThrGlyPro	ArgTrp	ArgHis	ProLysLys	ValGlyThr	AlaAla	ThrAla	SerLys	ArgAla	LysGlyThr
	3750	3760	3770	3780	3790	3800	3810	3820	3830	3840	3850
	CGATTGCA	CGAATTTCT	TGGTACAG	CATCTGCT	TTGGTTCA	CTAAATTC	CGTTCACT	TTCAAGCT	AAAAACAT	AAACGATC	AGCATTGCT
	GCTAAAGCT	CGTAAAGCC	ACCAATGCT	GTGAGGCT	AACCCAAAT	GATTTCAAT	CGAATGAT	AACTGCAAC	TTTTGCTTA	TTTGTAGT	TGTAAGCA
	Ser	LysArg	AlaThr	GlnLys	ThrArg	AlaHis					
	3860	3870	3880	3890	3900	3910	3920	3930	3940	3950	3960
	ACCTACTT	ATAGCAAT	TGCAAGCT	AGAGGAGA	GAGAGAGA	GAAAACAG	GGCCATTA	TTTGCAGC	GAAAACGAA	CGAATCAT	CTTGCATC
	TGGATGATA	TATCGTAG	ACTCGCCAT	CTTCTCTCT	CTTCTCTCT	CTTCTCTCT	CCGGTAAAT	AACTCGTC	TTTTCTGCT	CTGTTGAT	CTGCATGAG
	3970	3980	3990	4000	4010	4020	4030	4040	4050	4060	4070
	TCTCGGTT	TTTGTGTT	ACCTATA	AGGGGACG	AGAACTGT	AAATTAGT	TTTAGTAT	TTCTGCTG	CGGTGATA	TAGTATAGA	CAGTGAATA
	AGAGCCAAA	AACAGCAAA	TGGATATA	TCCCGGTG	TCTTTCAG	TTAATCAAG	AAATCACT	AAGCAGCA	CGCAATAT	ATCATATCT	GTCACTTTT
	H4	4080	4090	4100	4110	4120	4130	4140	4150	4160	4170
	L	ACTGCTG	TGGTAAAG	GGCAAGCT	TGGGAAAGG	TGGCGCAAG	CGTATGCA	AAAGTCTG	TGATAACAT	CAAGGTATC	GGAAGCTG
	ACTGCAAGC	ACCATTTCT	CGTTCCGA	ACCTTTCC	ACGCGGTC	CGATAGCT	TTCAACGAC	ACTATTGT	GTTCATAT	GTCTGAGC	ATAGCGGCA
	ThrGly	ArgGlyLys	GlyLys	GlyLys	GlyAla	LysArgHis	ArgLys	ValLeu	ArgAsp	AsnIle	GlnIle
	4180	4190	4200	4210	4220	4230	4240	4250	4260	4270	4280
	TTGCGCTG	GAGCGGTT	GAAAGGCA	TCTGAGCA	TATACAGCA	ACGCGTGC	GTTCTGAG	TTTTCTGG	GAACTGAT	CGATGCGC	TGACTACAC
	ACCGGGCC	CTCCGCCA	CTCCCGAT	AGACCTGAT	ATATCTCT	TTGGCACG	CAAGCTTC	AAAGAAGC	GTCGTTAA	GCACTACCG	ACTGGTGTG
	LeuAla	ArgArg	GlyGly	ValLys	ArgIle	SerGlyLeu	ThrGlu	ThrArg	GlyVal	LeuLys	ValPhe
	4290	4300	4310	4320	4330	4340	4350	4360	4370	4380	4390
	GGACACGC	AAGGGAAG	CAGTTACG	CATGATGT	GTGTACCT	TGAAGGCA	AGGCGCAC	CTCTAGCT	TTGGCGTA	AAAAGTAT	ATCCTGTG
	CCTGTGGG	TCTCTCT	GTAATGTC	GTCACAGAA	CACAAGAG	ACTCTCTG	TCCGCGTG	GAGATGCT	AACCCCAAT	TTTTCACT	TAGCACAT
	GluHis	AlaIle	AlaArg	LysArg	LysThr	ValThr	AlaMet	AspVal	ValThr	AlaLeu	LysArg
	4410	4420	4430	4440	4450	4460	4470	4480	4490	4500	4510
	CCCTATAA	GCAATCGT	CTTTACGA	CCACCATCA	GTTTTAAA	GGAGGAAG	TTTCAAAA	TATTTAAT	TATTTGCT	GACTCCAAA	AAATATGT
	GGGTAAAT	CGTTAGCC	GAAAGTCT	GGTGTAGT	CAAAAATTT	CCTCTCTTA	AAAGTTTT	ATAAATTA	ATAACAGC	CTGAGGTTT	TTTATACAT
	4520	4530	4540	4550	4560	4570	4580	4590	4600	4610	4620
	AACTAAAT	AAATGCTG	CGTTCGAT	CTTTCAATA	GAATGTGT	TTCTAACAT	GATTTGTA	GACTGTCA	ATGGTTTC	ACACCGCAT	TACAGCATC
	TTGATTTTA	TTTACAGCA	GCAAGCAT	GAAAGTTAT	CTTACACATA	AAAGTTGAT	CTAACATGT	CTGCATGAT	TACCAAACT	TGTGGCTAC	ATGCTGTAT
	4630	4640	4650	4660	4670	4680	4690	4700	4710	4720	4730
	CATCGTGT	TGGATGTT	TAAAGAACA	AATTAATG	TATACCGAT	ATAGTAGAA	ACCCCTACT	CATTACACT	TTTACGTTA	TATTTCTTA	TGATTTTAA
	GTAGCACC	ACGCTAAC	ATTTCTCT	TAAATATCA	AATGGCGTA	TATCATCTT	TGGGAATG	GTAATGGAT	AAATGCAAT	ATTAAAGAA	ACTAAATAT
	4740	4750	4760	4770	4780	4790	4800	4810	4820	4830	4840
	ATTACACAA	ATTCGCAAG	TAAAAATTT	TCCATATT	TGCTGAAT	ATTTGCTC	TGAAAAGG	GGTTTGAT	ACGATTTAT	GATCATGTG	GTTTTAGCT
	TAAATGTTT	TAAATGTTT	AAATTTTTA	AGCATTAGA	ACCTATTA	AAACAGCAG	ACTTTTCTG	CCAAAGTTC	GCTTAAATA	CTATGATC	CAAAAATCG
	4850	4860	4870	4880	4890	4900	4910	4920	4930	4940	4950
	TAGCCTTGA	AACGTTTAG	CCTTCTCT	GGTCTTCT	GGCAACGA	CAGCCTGT	ATTAGGCA	ACGCCACT	GTGCAATT	GACCGCGG	AGCAGCTGT
	ATCGAAACT	TTGCAATC	GAAGAAGG	CGAGAGAA	CGCTGCTT	GTGGACATA	TAACTGTT	TGGGTGGA	CAGTTTAA	CTGGGCTC	TCTGTCA
	4960	4970	4980	4990	5000	5010	5020	5030	5040	5050	5060
	TAACTGACT	GTCGTTGG	TACCGGTT	CGAGATGCT	CGGAATAT	CTAGTCTT	TGTTGATC	ACGACATCT	CCAGCGACT	C	
	AATTTAGG	CAGCAAGCC	ATGCGCTAA	CTCTACTG	CGCTTATA	GATCAGAA	ACAACAGCT	CTGCTGAT	GCTCGTTGA	G	
	LeuGlu	AlaAsp	AsnArg	IleAla	LeuGln	LeuHis	ArgPro	IleIle	ArgThr	LysLys	
											AsnAsp
											ArgIleAlaAsn
											GlyAlaLeuGlu

Fig. 1. Complete DNA sequence of the histone 5 kb unit. Position No. 1 is the XhoI site. Five histone structural genes are indicated by underlines. The boxed sequences show tRNA derived insertions (Alu-like family).

the sequence is A or T in Nos. 2115-2927 of Fig. 4) and 240 bp insertion (Nos. 2927-3167 of Fig. 4). The AT-rich region may have a specific structure and perform certain functions (49). The repeat of a simple sequence (GA)<sub>10</sub> was found in the spacer between H3 and H4 (Nos. 3882-3902 of Fig. 4). A similar repeating structure was found in the histone repeating unit of sea urchin (50) and in the gammer-globin genes of human (51).

t-RNA derived insertion sequences (240 bp) in the L unit

The structure and sequence of the insertion in the spacer between H1 and H3 were examined in detail. The insertion site was inferred from the sequences of the S (4.8 kb) and L units (5.0 kb) (32, 52 and Fig. 4). These sequences have several features in common with elements derived from t-RNA. First, the length of the insertion of each of these sequences (240 bp) is comparable to that of the Alu-like family (approximately 70-300 bp). Secondly, the structure of the sequence is quite similar to that of the Alu-like sequences (53) as described below. Thirdly, direct repeat (DR) and insertion sites are very AT-rich as suggested by Daniels and Deininger in Alu-like sequences (54). A typical structure of Alu-like sequences is shown in Fig. 1. There is a DR block at either end and inbetween, 57-leader, t-RNA derived, t-RNA unrelated and AT-rich blocks. Insertion sequences found in histone L repeating unit have similar structures and putative DR (ATA) (Fig. 1). The DNA sequence of the insertion corresponding to the tRNA derived block is shown in Fig. 5B and the conserved sequence in the tRNA family, in Fig. 5A. Twelve out of 14 positions in the latter were found in the former (55). t-RNA is transcribed by RNA polIII and contains its promoter within the molecule (boxed in Fig. 5A) (56). The general promoter for polIII proposed by Sharp et al. (56) is indicated at the bottom of Fig. 5 as well as the insertion sequences corresponding to this region. Conserved sequences of the polIII promoter were found in the insertion sequences at the same position (15 out of 16 were matching). From these findings the structure of these insertion sequences appears similar to that of the Alu-like family and the insertion sequences may possibly be derived from t-RNA.

Common sequences among five histone genes in *D. melanogaster*

The 5'-upstream positions of five histone genes in *D.*







Fig. 6. 5' regions and 3' regions are compared in 5 histone genes in *Drosophila melanogaster*. Only 200 bp upstreams from the start codon and 250 bp downstreams from the stop codon are shown in the figure. Common sequences are boxed.

to *Xenopus* (23) and yeast (24). The hair pin loop structured block (AATCGTCTCTTTTCAGGACCACAA) found at about 40-100 bp downstream from the stop codon appears important for the 3'-end formation of the m-RNA (57, 58). This sequence block was conserved in most of the histone 3' region (23). The AATAAA sequence thought to be a polyadenylation signal, is rarely present in published DNA sequences of histone genes (47).

**DISCUSSION**

Several common sequence blocks were found in the 5'- and 3'-regions of five histone genes of *D. melanogaster*. The AGTGAAA sequence block was found in the 5' region at almost the same position, but a homologous sequence could not be detected in other species such as sea urchin (18-22), yeast (26, 27), chicken (3-5), or human (8-10). Thus, this sequence block may be important for the specific expression of *Drosophila* histone genes. In addition to the "hair pin loop" structure block, the AATAAA se-

quence block was found in the 3'-region. From this, it is conceivable that *Drosophila* histone genes produce two kinds of mRNA, one carrying poly(A) and one that does not.

Recently, the Alu-like family, rodent type 2 Alu family, rat ID sequences, rabbit C family, and the bovine or goat 73-base-pair repeat were found to be similar to tRNA (53, 54, 59). The insertion sequence in the AT-rich spacer region of histone repeat unit is also similar to t-RNA. This insertion could be transcribed by RNA polIII, since the conservative sequence of the RNA polIII promoter is in this insertion. The insertion site is very close to the region reported as the binding sites for protease sensitive components (49). Thus, the element may perform some role in gene expression or constructing the chromatin structure. The functions of these sequences (AGTGAAA, AATAAA, Alu-like insertion) remain to be determined by future research.

In regard to the organization of the histone gene family in *D. melanogaster*, the L and S units are polymorphic and clustered, respectively. Variant types having non-histone fragments are sometimes present in the histone cluster (31, 60). In some *Drosophila* species such as *D. mauritiana*, *D. teissieri*, *D. erecta*, and *D. orena*, the length type is constant (61, 62). It appears that variant types accidentally increase or become fixed in the histone clusters of certain *Drosophila* species and several subclusters segregate in the polymorphic state in *D. melanogaster*.

The interruption of a histone cluster by non-histone fragments may at times reduce the rate of genetic exchange among the members of a family. Disruption of homogenization would cause the histone clusters to differ from those of the main cluster. If these "orphan" genes (63) come to have different function(s) as in the case of the late histone genes in sea urchin (18, 19), the processes responsible for hindering the exchange of genetic information among the members of a family, will also be important for the progressive evolution of organisms.

#### ACKNOWLEDGMENTS

We are grateful to Dr. K. Saigo for his considerable advice and suggestions and providing the histone probe. Thanks are also due to Drs. T. Mukai and S. Yokoyama for reading the manuscript,

---

to Dr. M. Hattori for his comments on sequencing DNA, to Dr. R. Yoshioka for preparing the synthetic DNA, and to Dr T. Gojobori and Miss E. Nakamura for assisting in the DNA sequence analysis. This work was supported in part by a research grant from the Yamada Science Foundation, Kudo Science Foundation and Ministry of Education, Japan.

\* Present address: Department of Ecology, Ethology and Evolution, University of Illinois, Shelford Vivarium, 606 E. Heaky St., Champaign, Illinois 61820, USA

\*\* To whom correspondence should be addressed.

### REFERENCES

1. Lifton, R. P., Goldberg, M. L., Karp, R.W. and Hogness, D. S. (1977) Cold Spring Harbor Symp. Quant. Biol. 42, 1047-1051.
2. Kedes, L.H. (1979) Ann. Rev. Biochem. 48, 837-870.
3. Engel, J.D., Sugarman, B.J. and Dodgson, J.B. (1982) Nature 297, 434-436.
4. Grandy, D.K., Engel, J.D. and Dodgson, J.B. (1982) J. Biol. Chem. 15, 8577-8580.
5. Harvey, R. R., Krieg, P. A., Robins, A. J., Coles, L. S. and Wells, J. R. E. (1981) Nature 294: 49-53.
6. Sittman, D., Chiu, I., Pan, C., Cohn, R., Kedes, L. and Marzluff, W. (1981) Proc. Natl. Acad. Sci. USA 78, 4078-4082.
7. Seiler-Tuyns, A. and Birnstiel, M.L. (1981) J. Mol. Biol. 151, 607-625.
8. Heintz, N., Zernik, M. and Roeder, R.G. (1981) Cell 24, 661-668.
9. Sierra, F., Lichtler, A., Marashi, F., Rickles, R., VanDyke, T., Clark, S., Wells, J., Stein, G. and Stein, J. (1982) Proc. Natl. Acad. Sci. USA 79, 1795-1799.
10. Zhong, R., Roder, R.G. and Heintz, N. (1983) Nucleic Acid Res. 11, 7409-7425.
11. Turner, P.C. and Woodland, H.R. (1982) Nucleic Acid Res. 10, 3769-3780.
12. Old, R. W., Woodland, H.R., Ballentine, J. E. M., Aldridge, T. C., Newton, C. A., Bains, W. A. and Turner, P. C. (1984) Nucleic Acid Res. 10, 7561-7580.
13. Moorman, A. F. M., DeLaff, R. T. M. and Destree, O. H. J. (1980) Gene 10, 185-193.
14. Ruberti, I., Fracapane, P., Pierandrei-Amaldi, P., Beccari, E., Amaldi, F. and Bozzoni, I. (1982) Nucleic Acid Res. 10, 7543-7559.
15. Van Dongen, W., de Laaf, L., Moorman, A. and Destree, O. (1981) Nucleic Acid Res. 9, 2297-2311.
16. Zernik, M., Heintz, N., Boime, I. and Roeder, R. (1980) Cell 22, 807-815.
17. Perry, M., Thomsen, G.H. and Roeder, R.G. (1985) J. Mol. Biol. 185, 479-499.

18. Childs, G., Nocente-McGrath, C., Lieber, T., Holt, C. and Knowles, J. A. (1982) *Cell* 31, 383-393.
19. Maxon, R., Mohun, T., Gormezano, G., Childs, G. and Kedes, L. (1983) *Nature* 301, 120-125.
20. Cohn, R.H., Lowry, J. C. and Kedes, L.H. (1976) *Cell* 9, 147-161.
21. Overton, G.C. and Weinberg, E.S. (1978) *Cell* 14, 247-257.
22. Cohn, R. H. and Kedes, L. H. (1979) *Cell* 18, 855-864.
23. Hentschel, C.C. and Birnstiel, M.L. (1981) *Cell* 25, 301-313.
24. Maxon, R., Cohn, R. and Kedes, L. (1983) *Ann. Rev. Genet.* 17, 239-277.
25. Old, R. W. and Woodland, H.R. (1984) *Cell* 38, 624-626.
26. Smith, M. M. and Andresson, O. S. (1983) *J. Molec. Biol.* 169, 663-690.
27. Wallis, J.W., Hereford, L. and Grunstein, M. (1980) *Cell* 22, 799-805.
28. Pardue, M. L., Kedes, E. S., Weinberg, E. S. and Birnstiel, M. L. (1977) *Chromosoma* 63, 135-151.
29. Karp, P. W. (1980) Ph. D. Thesis, Stanford Univ.
30. Saigo, K., Millstein, L. and Thomas, C.A. (1981) *Cold Spring Harbor Symp. Quant. Biol.* 45, 707-710.
31. Ikenaga, H. and Saigo, K. (1982) *Proc. Natl. Acad. Sci. USA* 79, 4143-4147.
32. Goldberg, M. (1979) Ph. D. thesis, Stanford Univ.
33. Yamazaki, T., Matsuo, Y., Inoue, Y. and Matsuo, Y. (1984) *Jpn. J. Genet.* 59, 33-49.
34. Maniatis, T., Hardison, R.C., Lacy, E., Lauer, C., O'Connell, C., Quon, D., Sim, G.K. and Efstratiadis, A. (1978) *Cell* 15, 687-701.
35. Karn, J., Brenner, S. and Barnett, L. (1983) *Methods in enzymology* 101, 3-19.
36. Benton, W. and Davis, R. (1977) *Science* 196, 180-182.
37. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) *A laboratory manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
38. Vieira, J. and Messing, J. (1982) *Gene* 19, 259-268.
39. Mandel, M. and Higa, A. (1970) *J. Mol. Biol.* 53, 159-162.
40. Rigby, P.W.J., Diekmann, M., Rhodes, C. and Berg, P. (1977) *J. Mol. Biol.* 113, 237-251.
41. Southern, E. (1975) *J. Mol. Biol.* 98, 503-518.
42. Sanger, S., Nicklen, S. and Coulson, A. (1977) *Proc. Natl. Acad. Sci. USA* 74, 5463-5467.
43. Sanger, F., Coulson, A.R., Barrell, B.G., Smith, J.H. and Roe, B.A. (1980) *J. Mol. Biol.* 143, 161-178.
44. Hattori, M., Hidaka, S. and Sasaki, Y. (1985) *Nucleic Acid Res.* 13, 7813-7827.
45. Alfageme, C.R., Zweidler, A., Mahowald, A. and Cohen, L.H. (1974) *J. Biol. Chem.* 249, 3729-3736.
46. Isenberg, I. (1979) *Ann. Rev. Biochem.* 48, 159-191.
47. Wells, D.E. (1986) *Nucleic Acid Res.* 14, r119-r149.
48. Elgin, S. C., Schilling, J. and Hood, L.E. (1979) *Biochemistry*, 5679-5685.
49. Gasser, S. M. and Laemmli, U. K. (1986) *The Embo Journal* 5, 511-518.
50. Sures, I., Lowry, J. and Kedes, L.H. (1978) *Cell* 15, 1033-1044.

## Nucleic Acids Research

---

51. Slighyom, J.L., Blechl, A.E. and Smithies, O. (1980) *Cell* 21, 627-638.
52. Matsuo, Y. and Yamazaki, T. (1989) *Genetics* (in press).
53. Sakamoto, K. and Okada, N. (1985) *J. Mol. Evol.* 22, 134-140.
54. Daniels, G.R. and Deininger, P. L. (1985) *Nucleic Acid Res.* 13, 8939-8954.
55. Lawrence, C.B., McDonnell, D.P. and Ramsey, W. J. (1985) *Nucleic Acid Res.* 13, 4239-4252.
56. Sharp, S., DeFranco, D., Dingermann, T., Farrell, P. and Soll, D. (1981) *Proc. Natl. Acad. Sci. USA* 78, 6657-6661.
57. Birchmeier, C., Grosschedl, R. and Biernstiel, M.L. (1982) *Cell* 28, 739-745.
58. Kreig, P.A. and Melton, D.A. (1984) *Nature* 308, 203-206.
59. Endoh, H. and Okada, N. (1986) *Proc. Natl. Acad. Sci. USA* 83, 251-255.
60. Liebermann, D., Hoffman-Liebermann, B., Weinthal, J., Childs, G., Maxon, R., Mauron, A., Cohen, S.N. and Kedes, L. (1983) *Nature* 306, 342-347.
61. Cohn, E., Strachan, T. and Dover, G. (1982) *J. Mol. Biol.* 158, 17-35.
62. Dover, G. (1982) *Nature* 299, 111-116.
63. Childs, G., Maxon, R., Cohn, R.H. and Kedes, L. (1981) *Cell* 23, 651-663.