Drosophila melanogaster Male Germ Line-Specific Transcripts With Autosomal and *Y*-Linked Genes

Steven R. H. Russell[1,2] and Kim Kaiser

Department of Genetics, University of Glasgow, Glasgow G11 5JS, Scotland

Manuscript received October 1, 1992
Accepted for publication January 16, 1993

ABSTRACT

We have identified of set of related transcripts expressed in the germ line of male *Drosophila melanogaster*. Surprisingly, while one of the corresponding genes is autosomal the remainder are located on the *Y* chromosome. The autosomal locus, at 77F on chromosome arm 3L, corresponds to the previously described transcription unit 18c, located in the first intron of the gene for an RI subunit of cAMP-dependant protein kinase. The *Y* chromosome copies have been mapped to region h18-h19 on the cytogenetic map of the *Y* outside of any of the regions required for male fertility. In contrast to *D. melanogaster*, where *Y*-linked copies were found in nine different wild-type strains, no *Y*-linked copies were found in sibling species. Several apparently *Y*-derived cDNA clones and one *Y*-linked genomic clone have been sequenced. The *Y*-derived genomic DNA shares the same intron/exon structure as the autosomal copy as well as related flanking sequences suggesting that it transposed to the *Y* from the autosomal locus. However, this particular *Y*-linked copy cannot encode a functional polypeptide due to a stop codon at amino acid position 72. Divergence among five different cDNA clones ranges from 1.5 to 6% and includes a large number of third position substitutions. We have not yet obtained a full-length cDNA from a *Y*-linked gene and therefore cannot conclude that the *D. melanogaster Y* chromosome contains functional protein-coding genes. The autosomal gene encodes a predicted polypeptide with 45% similarity to histones of the H5 class and more limited similarity to cysteine-rich protamines. This protein may be a distant relative of the histone H1 family perhaps involved in sperm chromatin condensation.

SPERMATOGENESIS, the production of highly differentiated sperm cells from stem cell progenitors, is a complex process that is apparently similar in organisms as diverse as insects and mammals. In the case of *Drosophila melanogaster*, single stem cell derivatives (gonial cells) each give rise to 16 primary spermatocytes by mitotic division [see LINDSLEY and TOKUYASU (1980) for a comprehensive review]. The latter grow rapidly in size, during which time they accumulate RNA and protein for the postmeiotic phase of development. Although translation can be detected postmeiotically, it is generally accepted that the majority of gene transcription is restricted to premeiotic stages (OLIVIERI and OLIVIERI 1966; DAS, KAUFMAN and GAY 1964a). Postmeiotic development can thus be viewed as being, for the most part, genetically preprogrammed. After meiosis, the 64 spermatids embark upon a complex program of cellular rearrangement and differentiation, eventually becoming motile sperm. Elucidation of the molecular processes underlying this extreme example of cellular differentiation represents a considerable challenge for developmental biologists.

As might be expected of such a complex process as spermatogenesis, many mutations are known that disrupt it. It has been estimated that 10% of *D. melanogaster* genes can be mutated by EMS to cause male-sterility (LINDSLEY and LIFSCHYTZ 1972). LIFSCHYTZ (1987) has divided these "fertility genes" into two broad classes: genes that express products specific to the male germ line, and genes whose function is required in both the germ line and soma. He suggests that the germ line is especially sensitive to perturbation of gene activity required for basic cellular functions. In support of this, some male-sterile mutations have been found to be alleles of lethal mutations [reviewed by LIFSCHYTZ (1987)]. Despite the availability of mutants, classical genetic analysis of sperm development can be problematic due to the preprogrammed nature of spermatogenesis. Pleiotropic effects of mutations in genes expressed premeiotically can cause a multiplicity of general defects later in sperm development, thereby thwarting a traditional epistatic analysis (HARDY, TOKUYASU and LINDSLEY 1981).

An alternative line of enquiry has been the characterization of genes identified solely on the basis of

The sequence data presented in this article have been submitted to the EMBL/GenBank Data Libraries under the accession numbers Z19565 (λgS9), Z19567 (cS1), and Z19566 (cS8).

[1] To whom correspondence should be addressed.
[2] Present address: Department of Genetics, University of Cambridge, Downing Street, Cambridge CB2 3EH, England.

their specific expression in the male germ line. Analysis of the *mst*87F gene family, for example, and of the gene encoding testis-specific β-tubulin, has identified sequences that confer germ line-specific transcriptional and translational regulation on reporter genes (SCHAFER *et al.* 1990; MICHIELS *et al.* 1989). There are also enhancer-trap lines in which germ line-specific genes have been tagged by β-galactosidase staining (GONCZY, VISWANATHAN and DINARDO 1992), and that in addition provide valuable markers for the various stages of sperm development. Coupled with immunological studies of germ line-specific proteins, studies such as these are likely to make an increasingly important contribution.

Only a very small proportion of the *D. melanogaster* "fertility genes" are male-specific in the sense of being on the *Y* chromosome. Despite accounting for approximately 12% of the male genome, the *Y* appears to be composed of mainly satellite DNA sequences (BONACCORSI and LOHE 1991). It is entirely heterochromatic (HEITZ 1933; KAUFMANN 1934; PIMPINELLI, SANTINI and GATTI 1978), and is dispensable for normal somatic development. It is, however, essential for male fertility (BRIDGES 1916), six relevant complementation groups being known. Four of these *fertility factors* (*kl-1*, *kl-2*, *kl-3* and *kl-5*) are on the long arm, and two (*ks-1* and *ks-2*) are on the short arm (KENNISON 1981; HAZELRIGG, FORNILI and KAUFMAN 1982; GATTI and PIMPINELLI 1983). At least three (*kl-3*, *kl-5* and *ks-1*) are associated with giant "lampbrush" loops reminiscent of those found in the spermatocyte nuclei of *Drosophila hydei* (BONACCORSI *et al.* 1988), a species that has been the focus of relatively intense study because of its large spermatocytes. The *Y* chromosome of *D. hydei* has between 7 and 16 complementation groups necessary for male fertility [see HACKSTEIN (1987) for a discussion], 5 of which are associated with lampbrush loops (HENNIG 1985).

In terms of its molecular structure the Drosophila *Y* chromosome remains poorly described. Most studies have focused on the *D. hydei* lampbrush loop containing fertility factors where it has been shown that the loops represent large single transcription units composed, in part, of loop-specific repetitive sequences (GROND, RUTTEN and HENNIG 1984; DE LOOS *et al.* 1985; HARAVEN, ZUCKERMAN and LIFSCHYTZ 1986; WLASCHEK, AWGULEWITSCH and BUNEMANN 1988). In *D. melanogaster* more limited studies suggest that approximately 80% of the *Y* chromosome is composed of simple satellite sequences. Here, however, the repetitive sequences are not loop-specific (BONACCORSI and LOHE 1991). LIVAK (1990) has studied the *Ste* and *Su(Ste)* loci where he has shown that the transcription and splicing of mRNA from the *X*-linked *Ste* locus is regulated by related *Su(Ste)* sequences located in region h11 of the *Y* chromosome. *Su(Ste)* contains regions of sequence identity with *Ste* interspersed with nonhomologus sequences (LIVAK 1990). DANILEVSKAYA and colleagues (1991) have sequenced DNA from the *Su(Ste)* region which shows homology to *Ste* and is also related to the He-T family of repetitive sequences (TAVERSE and PARDUE 1989). Thus it may be that *X*-linked *Ste* sequences have been transposed to the *Y* chromosome where they have diverged and been recruited to regulate *Ste*. Finally GOLDSTEIN, HARDY and LINDSLEY (1982) presented evidence to suggest that the fertility factors *kl-3* and *kl-5* encode polypeptide components of the sperm tail axoneme, the structure that provides the mechanical force for sperm movement. However, they could not entirely rule out the possibility that the relevant genes are elsewhere in the genome and are merely regulated in some way by *kl-3* and *kl-5*. In this respect it is of interest to note that male-sterile lesions in *D. melanogaster* loop-forming fertility factors have been identified which appear to have no effect on the morphology of the loops (BONACCORSI *et al.* 1988). Thus it is still an open question as to whether or not the Drosophila *Y* chromosome contains conventional protein coding genes. In summary, for neither *D. melanogaster* or *D. hydei* are we certain of the molecular structure or function of the majority of the *Y* chromosome and the role that it plays in spermatogenesis remains obscure.

This report describes the characterization of a family of *D. melanogaster* transcripts specific to the male germ line. Surprisingly, while one of the corresponding genes is on the third chromosome, the remainder are located on the *Y* chromosome. We have determined the sequence of genomic DNA including one *Y* chromosomal locus, and find that its structure is very similar to that of the autosomal locus. It appears unable to encode a functional polypeptide, however. The sequence of a near full-length cDNA clone derived from the autosomal locus (KALDERON and RUBIN 1988) translates as a polypeptide with similarity both to the vertebrate histone H1/H5 class of proteins and to mammalian cysteine-rich protamines.

MATERIALS AND METHODS

Drosophila methods: Drosophila stocks were maintained at 25° on standard cornmeal/yeast/agar medium. Mutant nomenclature is as described by LINDSLEY and ZIMM (1992). Agametic males were generated by crossing *Oregon R* males with virgin *bw tud*^{wc8}/*bw tud*^{wc8} females (BOSWELL and MAHAWOLD 1985). Genotypic females somatically transformed to maleness were generated by crossing *cn tra2*^{B} *bw*/*CyO*; *B*^{s}*Y* males with virgin *Df(2L)trix*/*CyO* females (BELOTE and BAKER 1983; GORALSKI, EDSTROM and BAKER 1989). Single-sex populations for Northern blot analysis were generated by use of the *cin* strategy (DIBENEDETTO *et al.* 1987). *XO* and *XXY* individuals were generated by crossing *Y*^{S}.*Y*^{L}, *In(1)EN, y w f* males with virgin *Oregon R* females. Elements of *T(X;Y)*s were separated by crossing *T(X;Y)* males with virgin *Oregon R* females (KENNISON 1981). For breakpoints in either *Y*^{L} or *Y*^{S}, the *Y*^{D}*X*^{P} element segregates to male

progeny and the $X^{D}Y^{P}$ element to female progeny. Males carrying *Y* chromosome deficiencies were generated by crossing attached *XY/Df(Y)* males to virgin *y w f* females, male progeny carry the *Y* chromosome deficiency. Wild-type flies were obtained from the Cambridge stock collection.

Nucleic acid isolation: Large and small scale preparation of Drosophila genomic DNA was carried out essentially as described by ASHBURNER (1989). Total Drosophila RNA from various developmental stages was isolated, with minor modifications, by the method of CHOMCZYNSKI and SACCHI (1987). Purification of poly(A$^+$) mRNA by oligo-dT cellulose chromatography, subcloning of cDNA and genomic DNA fragments, plasmid DNA and bacteriophage λ DNA isolation, were all carried out as described by SAMBROOK, FRITSCH and MANIATIS (1989).

Gel electrophoresis, blotting and hybridization: DNA and RNA were separated by electrophoresis on agarose and formaldehyde/agarose gels respectively, and were blotted onto Hybond-N membranes, as described by SAMBROOK, FRITSCH and MANIATIS (1989). Hybridization was carried out at 42° in 50% formamide, 1% sodium dodecyl sulfate (SDS), 5 × SSPE, 10 × Denhardt's reagent 100 μg/ml denatured salmon sperm DNA. ^{32}P-labeled DNA probes were prepared from gel-purified restriction fragments by random-priming (FEINBERG and VOGELSTEIN 1983). Blots were washed in 0.1 × SSC, 1% SDS at 65° (high stringency), or in 2 × SSC, 1% SDS at 55° (low stringency).

DNA sequencing: Double-stranded plasmid DNA sequencing was carried out by the Sequenase 2 method, according to the manufacturers recommendations (U.S. Biochemical Corp.). Nested deletions of plasmids were generated as described by HENIKOFF (1984). DNA sequence analysis was performed with the assistance of the GCG suite of computer programs (DEVEREUX, HAEBEREL and SMITHIES 1984).

Male-specific genes: λgS8 and λgS9, two EMBL3 derivatives containing fragments of *D. melanogaster* genomic DNA, were isolated in a screen for phage that hybridize with cDNA representing male, but not female, third instar larvae (RUSSELL 1989). Although the inserts are non-overlapping, restriction fragments that hybridize with male cDNA cross-hybridize with one another under conditions of high stringency. Two cDNA clones (cS1, 530 bp; cS8, 479 bp) were isolated from a male third instar larval cDNA library (RUSSELL 1989) via homology with a 3.2-kb *Hin*dIII fragment of λgS9 that encompasses the male-specific region (Figure 1).

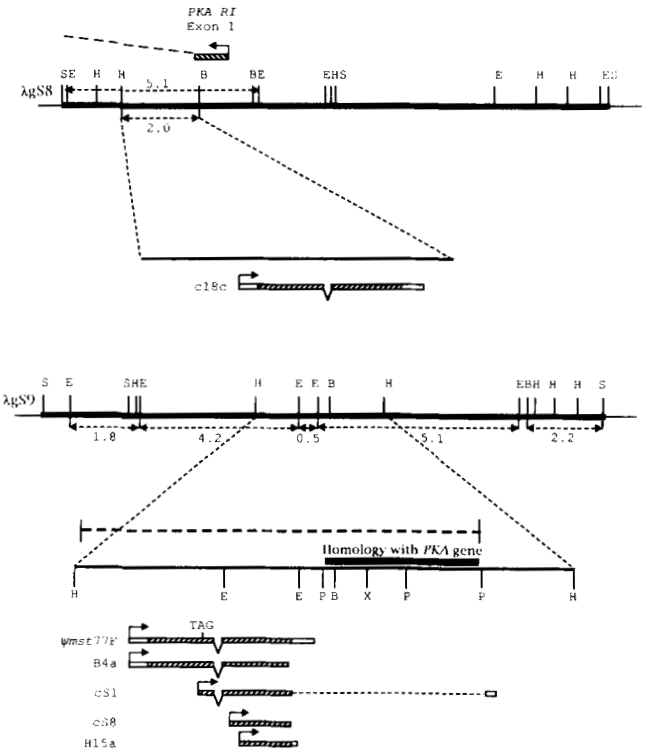

FIGURE 1.—The Drosophila genomic DNA clones λgS8 (autosomal; 77F) and λgS9 (*Y*-linked), their relationships to the cDNA clones c18c and cS1, and to the gene encoding the RI subunit of PKA (see also KALDERON and RUBIN 1988). Exons are shown as boxes (with the proviso discussed in the text concerning what appears to be an extra exon of cS1). Relative directions of transcription are indicated by arrows. c18c appears to represent an authentic transcript of the autosomal locus defined by λgS8 (see text). λgS8 also contains a component of the functional RI gene from 77F as shown. Other relationships are not intended to imply the origin of a particular cDNA from the genomic locus in question, merely a correspondence in terms of overall DNA sequence. Thin lines in the maps of genomic DNAs represent vector arms. Shaded boxes in the maps of cDNA clones represent putative coding sequences, open boxes represent 5′ and 3′ untranslated sequences. ψmst77F represents the conceptual mRNA from the λgS9 clone with the position of the stop codon indicated. The hatched box above λgS8 represents the first exon of the RI gene (note that a small proportion of RI transcripts have sequences internal to the box spliced out). In the case of the 77F locus, this exon is separated from subsequent exons by an intron ≈12 kb in length (beyond the limits of the Drosophila DNA component of λgS8). The 3.2-kb *Hin*dIII fragment shown in expanded form below the map of λgS9 includes all of the homology with male cDNA. It has been sequenced in its entirety. The region of *known* homology between λgS9 and the 77F locus begins at nucleotide position 47 and extends for 2411 bp and is shown as a dotted line above the 3.2-kb *Hin*dIII fragment (see also Figure 6). The extent of homology with the PKA RI subunit is indicated with a black bar. The location of the 3′ end of cS1 is shown and is also included in Figure 6. B, *Bam*HI; E, *Eco*RI; H, *Hin*dIII; P, *Pst*I; S, *Sal*I. Sizes of restriction fragments are given in kilobases.

RESULTS

In the course of screening a *D. melanogaster* genomic DNA library for genes expressed in male but not female third instar larvae, we isolated two clones (λgS8 and λgS9; Figure 1) with substantial homology between their male-specific regions, despite having non-overlapping inserts (RUSSELL 1989). As we show below these two clones contain related sequences which show male restricted expression. Furthermore we demonstrate that in the genome these sequences are found at a single autosomal site and at several copies on the *Y* chromosome. The sequence of a 2.0-kb *Bam*HI-*Hin*dIII fragment including the entire male-specific region of λgS8 was found to be that of a previously characterized locus, 18c (KALDERON and RUBIN 1988; EMBL accession no. X16962), mapping to cytological region 77F on the left arm of chromosome *3* (see Figure 6 for the sequence). 18c is contained within the first intron of a gene encoding a regulatory (RI) subunit of cAMP-dependent protein kinase (PKA). As

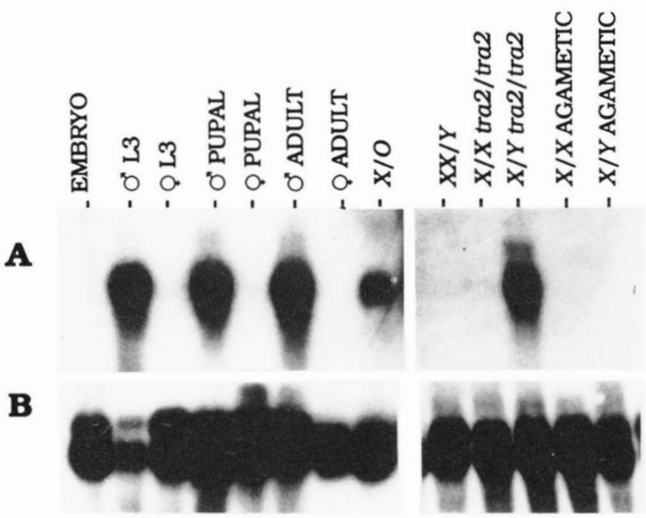

FIGURE 2.—Northern blots representing approximately 2 μg of poly(A⁺) mRNA per lane from the indicated stages and genotypes. (A) Putative *Y*-linked cS1 probe and (B) same blot reprobed with cloned DNA from the actin 5C gene (FYRBERG *et al.* 1981). L3, third instar larvae.

the transcription of 18c is restricted to males we propose to rename it *mst*77F (*m*ale-*s*pecific *t*ranscript) in accordance with accepted practice.

Transcriptional analysis: The cDNA clone cS1 was used to probe Northern blots of poly(A⁺) mRNA prepared from various developmental stages, as well as from agametic and sexually transformed individuals (Figure 2). A broad band, suggestive of a heterogeneous collection of mRNAs of average size 1400 nucleotides, is seen in lanes representing wild-type male, but not female, third instar larvae, pupae and adults. The band is not seen in lanes representing adult male offspring of females homozygous for the mutation *tudor* (*tud*^{wc8}). Such males have no germ line (BOSWELL and MAHAWOLD 1985). Nor is it seen in adult genotypic females homozygous for the *tra*2 mutation (pseudomales) which, though transformed somatically to maleness, have a female-like or undifferentiated germ line in which few spermatocytes can be detected (STEINMANN-ZWICKY, SCHMIDT and NOTHIGER 1989). In both agametic males and pseudomales the somatic component of the gonad is phenotypically male, albeit underdeveloped (BOSWELL and MAHAWOLD 1985; BROWN and KING 1960). Taken together these data suggest that the transcript(s) detected by cS1 are either expressed in the male germ line, or are dependent upon the male germ line for their expression. Support for the former hypothesis is provided by preliminary results from *in situ* hybridization to larval testis, in which cS1 detects transcripts present in spermatocytes (S. R. H. RUSSELL, unpublished data).

Figure 2 also shows the relative levels of transcript(s) in *XXY* females (none detectable), and in *XO* males (detectable but at reduced intensity). Three possibilities could account for the latter observation: (a) a general reduction in cell metabolism, associated with disruption of spermatogenesis in *XO* individuals, results in reduced transcription or stability of germ-line transcript(s); (b) *Y*-linked copies are in some way required for maximal transcription of the autosomal locus; (c) *Y*-linked copies are themselves transcribed, giving rise to poly(A⁺) RNAs similar in size to the autosomal transcript. Isolation of cDNA clones appearing to represent *Y*-linked copies suggests that the latter are indeed transcribed (see below). Finally, KALDERON and RUBIN (1988) presented evidence that the *mst*77F family is transcribed in adult heads, although at relatively low levels compared with those in adult bodies. We have not formally ruled out such a possibility, but given the above data it is probable that their result was due to a contamination of heads with bodies (D. KALDERON, personal communication).

***Y*-linked sequences with homology to *mst*77F:** KALDERON and RUBIN (1988) noted that *mst(77)F* sequences were repeated in the genome and to investigate their distribution carried out *in situ* hybridization to squashes of salivary gland polytene chromosomes. Hybridization was seen only at 77F, suggesting to them that the sequences related to *mst*77F are all clustered within this region of chromosome *3*. Hybridization to the *Y* chromosome would not have been observed in these experiments. We have obtained evidence for *Y*-linked copies as follows. The cDNA clone cS1 was used to probe Southern blots of DNA isolated from *Oregon R* males and females. In addition to a common band, due to a 5.1-kb *Eco*RI fragment that spans the *mst*77F locus, seven extra bands are detected in lanes representing male DNA (10 kb, a 5.1-kb band which comigrates with the autosomal band, 4.5, 4.2, 3.8, 3.1 and 0.5 kb; Figures 3 and 4). Thus, sequences related to cS1 are present on the *Y* chromosome, a conclusion that is supported by similar analysis of DNA isolated from *XO* males and *XXY* females (Figure 4A). In the case of three of the male-specific bands (0.5, 4.2 and 5.1 kb) fragments of the same sizes and hybridization characteristics occur within the genomic clone λgS9. Identical results are obtained when the KALDERON and RUBIN c18c clone is used as a probe (not shown). The lack of overlap between the Drosophila DNA components of λgS8 and λgS9 genomic clones is therefore explained by the latter being derived from the *Y* chromosome.

To investigate the generality of our observations, we examined DNA prepared from *D. melanogaster* isolated from a range of geographical locations, as well as from various sibling species. In the case of nine additional *D. melanogaster* strains, both male and female DNAs contain a 5.1-kb *Eco*RI fragment that hybridizes with cS1, while additional fragments are evident in male DNA. The hybridization patterns of

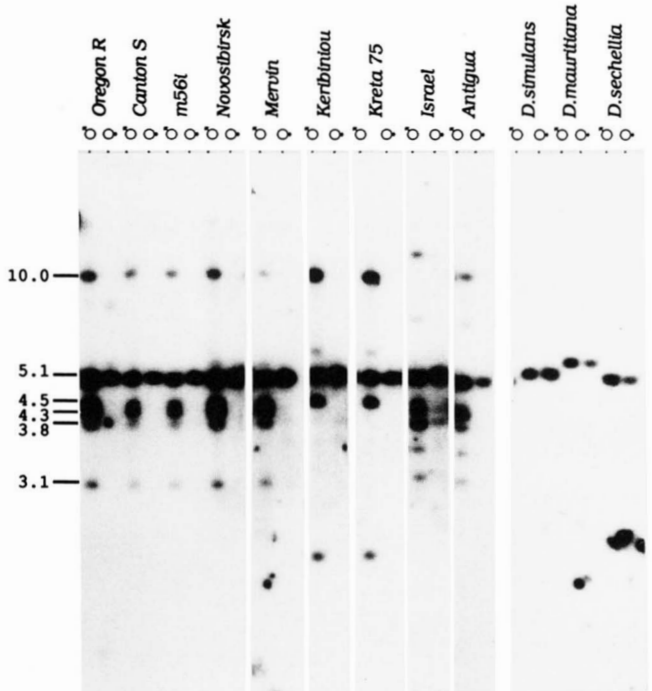

FIGURE 3.—Southern blot of *Eco*RI-cleaved male and female genomic Drosophila DNA probed with the putative *Y*-linked cS1 cDNA clone. The first 18 lanes are DNAs from *D. melanogaster* strains; the blot was washed at high stringency. The remaining lanes, containing DNA from sibling species, were washed at low stringency. All lanes derive from the same blot, which has been rearranged for clarity. Sizes in kilobases refer to the male-specific bands of *Oregon R* DNA. (Note that the 0.5-kb fragment referred to in the text is not visible on this exposure. The relevant band can be seen in Figure 4A.)

Canton S, *m56i*, *Novosibirsk* and *Mervin* DNAs (Figure 3) are all similar to the *Oregon R* pattern. *Keribiniou* and *Kreta 75* DNAs exhibit a simpler pattern with four male-specific bands, two of which appear to have *Oregon R* counterparts (Figure 3). *Israel* and *Antigua*, as well as the attached-*XY* strain used for generating *XO* and *XXY* individuals, the *T(X;Y)* strains and the *Df(Y)* strains described below, exhibit a more complex pattern of at least 10 male-specific bands, some of which appear to have *Oregon R* counterparts (Figures 3 and 4). In contrast, none of the three sibling species we have tested (*Drosophila simulans*, *Drosophila sechellia* and *Drosophila mauritiana*) shows evidence of *Y*-linked copies (Figure 3), their DNAs exhibit only a single band in both males and females. *In situ* hybridization to *D. simulans* polytene chromosomes indicates a single site of hybridization at 77F (not shown). In summary, the *Y*-linked sequences appear to be unique to *D. melanogaster*, and thus may have "transposed" to the *Y* chromosome after present day *D. melanogaster* and its sibling species diverged from a common ancestor.

We have also prepared DNA from males lacking different regions of the *Y* chromosome. The entirely heterochromatic *Y* of *D. melanogaster* consists of a long

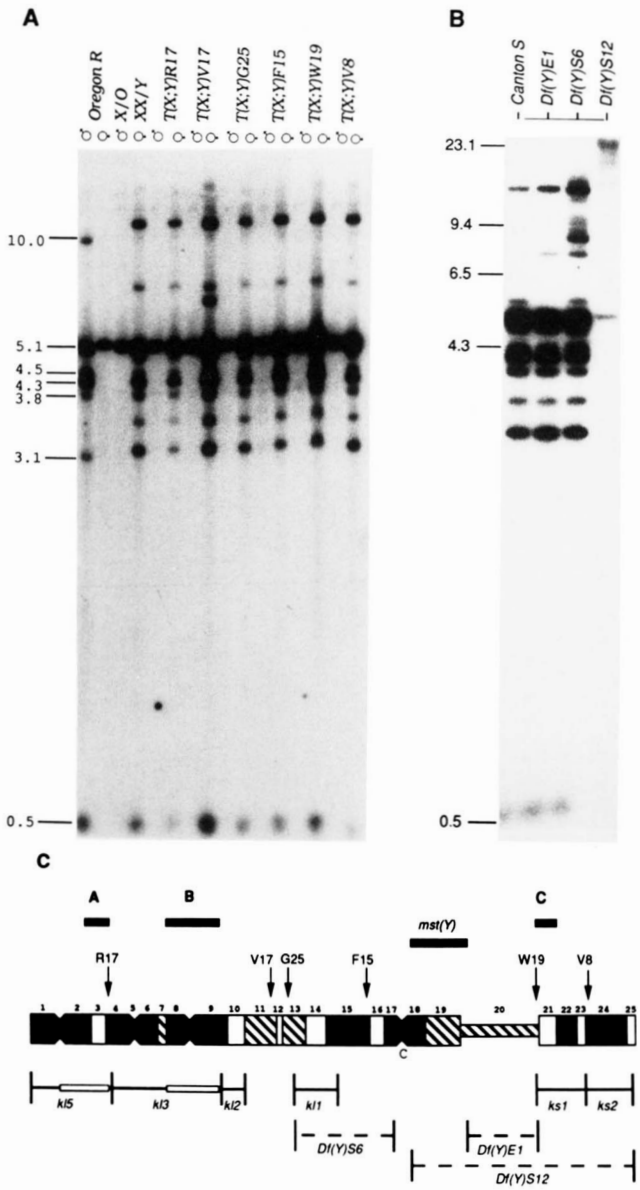

FIGURE 4.—Restriction site and cytogenetic analyses. (A) Southern blot of *Eco*RI-cleaved male and female DNA isolated from *D. melanogaster* strains as indicated and probed with the putative *Y*-linked cS1 cDNA clone. The blot was washed at high stringency. Sizes in kilobases refer to the male-specific bands of *Oregon R* DNA. (B) Southern blot of *Eco*RI-cleaved DNA from males carrying the indicated *Y* chromosome deficiencies. The blot was probed with the autosomally derived cDNA clone c18c (KALDERON and RUBIN 1988) and washed at high stringency. The intensity of the 5.1-kb fragment in *Df(Y)S12* is of lower intensity than expected whether the c18c or cS1 probes are used; however, we note a substantial portion of the DNA remains undigested, and we have obtained the same result on two separate occasions and believe that it is an experimental artifact. (C) A cytogenetic map of the *Y* chromosome (redrawn after BONACCORSI *et al.* 1988), showing the *T(X;Y)* and *Df(Y)* breakpoints. Lampbrush loop-forming regions (A, B and C) are indicated above the map. The locations of fertility factors are shown with the lines representing their maximal extent and the open boxes their minimal extent. The centromere and the region containing the *mst*77F homology are also indicated.

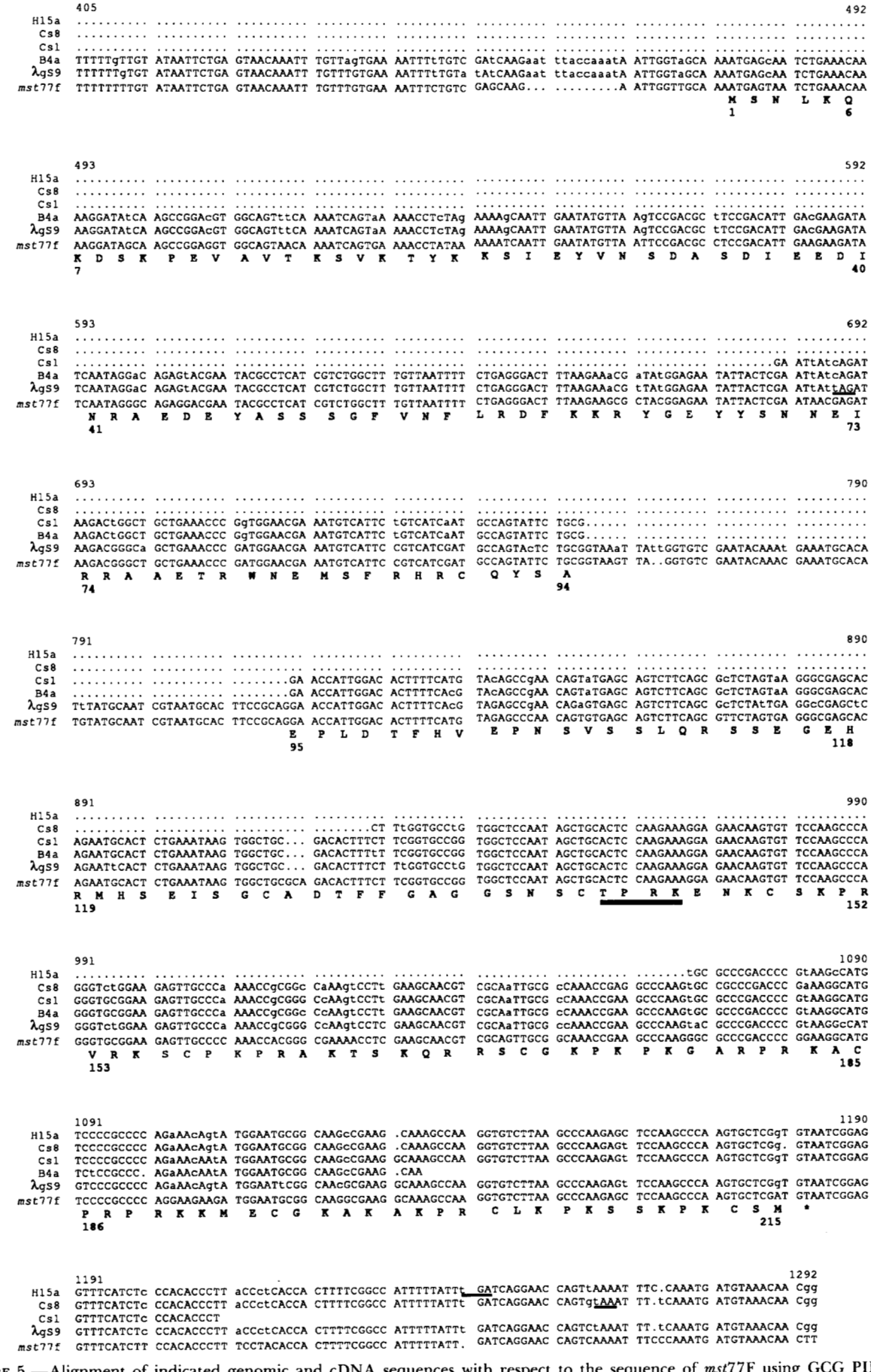

FIGURE 5.—Alignment of indicated genomic and cDNA sequences with respect to the sequence of *mst*77F using GCG PILEUP with default parameters. Differences from *mst*77F are indicated in lower case. Numbering is taken from KALDERON and RUBIN (1988). The 5′

arm, Y^L, and a short arm, Y^S (Figure 4C). Radiation-induced translocation between the *Y* and the *X* generates hybrid chromosomes [*T(X;Y)*'s] of two types. Y^DX^P chromosomes (*D*, distal; *P*, proximal) have a centromere-less portion of the *Y* appended to a centromere-containing portion of the *X*. Conversely, X^DY^P chromosomes have a centromere-containing portion of the *Y* appended to a centromere-less portion of the *X*. We analysed a series of strains carrying different reciprocal translocations between either the long or the short arm of the *Y* chromosome and centromeric heterochromatin of the *X* (KENNISON 1981). Males having both hybrid chromosomes lack no genetic information and are fertile if the *Y* breakpoint does not interrupt a fertility factor (see Figure 1 of GOLDSTEIN, HARDY and LINDSLEY 1982 for a diagrammatic representation). Crossing such males with wild-type females, however, causes Y^DX^P chromosomes to segregate in male progeny and X^DY^P chromosomes to segregate in female progeny. Six *T(X;Y)*'s were studied in all, representing breakpoints throughout the *Y* chromosome (Figure 4C). We followed the sequences of interest by hybridization as above to Southern blots of the respective DNAs (Figure 4A). In the case of Y^L breaks (R17, V17, G25 and F15), homology segregated with female progeny. Thus, all of the *Y*-linked members must lie proximal to (*i.e.*, on the centromere side of) the F15 breakpoint. In the case of Y^S breaks (W19 and V8) homology again segregated with the female progeny. Thus all of the *Y*-linked members must lie proximal to the W19 breakpoint.

To localize the *Y* linked copies more precisely we investigated males bearing three different *Y* chromosome deficiencies (Figure 4C). Compound *XY/Df(Y)* males were crossed with virgin *y w f* females and the DNA from male progeny analyzed by Southern blotting (Figure 4B). While the normal complement of male-specific fragments is present on the *Df(Y)E1* and *Df(Y)S6* chromosomes none is present on the *Df(Y)S12* chromosome. In summary, all of the *Y*-linked members of the *mst*77F family reside on the short arm of the *Y* chromosome, between the proximal boundaries of *Df(Y)S12* and *Df(Y)E1*. This region, h18-h19 on the cytogenetic map (GATTI and PIMPINELLI 1983), is indicated in Figure 4C. It does not include any known fertility determinants.

A gene "family": That KALDERON and RUBIN's (1988) clone c18c represents an authentic transcript of the autosomal *mst*77F locus is suggested by the fact that the two sequences differ at only a single, noncoding, position. They also sequenced two other cDNAs (B4a and H15a) that, although clearly related to *mst*77F, are sufficiently different both from it and from each other that they seemed likely to represent independent loci. We have also isolated and sequenced two cDNA clones (cS1 and cS8), and have sequenced a segment of λgS9 DNA (see legend to Figure 1) that includes the entire male-specific region. Taken together these sequences appear to represent five additional loci which our Southern blot data indicate derive from the *Y* chromosome. An alignment of the various sequences with respect to that of *mst*77F, together with the sequence of the 215-amino acid product deduced for *mst*77F (via the sequence of c18c), is shown in Figure 5.

The member of the family represented by genomic clone λgS9 retains the intron/exon structure of the autosomal locus (the single intron comprises residues 757–818), but is unable to encode a full-length polypeptide due to a stop codon at amino acid residue 72 (equivalent to residues 688–690 of the *mst*77F sequence and underlined in Figure 5). There is also an insertion of 12 bp just 5′ to the putative initiation codon, an in-frame deletion of nucleotide residues 918–920 (that would cause deletion of an alanine residue if a full-length translation product was generated), and a number of single-base substitutions.

cDNA B4a (698 bp) extends 5′ to include the *mst*77F translation initiation codon, and the 12-bp insertion observed in λgS9. In the other direction the homology with *mst*77F extends to nucleotide position 1134 where the B4a cDNA terminates (KALDERON and RUBIN 1988). It is not known whether the cDNA showed evidence of a poly(A) tail since the 3′ end of the clone was not sequenced in full (D. KALDERON, personal communication). As in the case of λgS9, B4a has an in-frame deletion of residues 918–920, together with a number of single-base substitutions. There are also deletions of residues 1100 and 1131, the effect of which is a change of reading frame beyond amino acid residue 188. According to KALDERON and RUBIN, the putative B4a product would terminate after a further 10 amino acids. Taking into account the in-frame deletion, the B4a product would thus be 18 amino acids shorter than the *mst*77F product.

boundary represents residue 48, the 3′ boundary residue 863 of the polyadenylated cDNA clone c18c (total length 1031 bp). The sequence of c18c can be derived merely by removing the intron (residues 757–818) from the *mst*77F sequence. Homology between *mst*77F and λgS9 extends considerably further both 5′ and 3′ (see Figure 6). B4a extends 5′ to position 399, and 3′ to position 1124. cS1 diverges from *mst*77F beyond position 1209 and terminates after a further 60 bp with evidence of a poly(A) tail. cS8 diverges beyond position 1340 and terminates after a further 69 bp with no evidence of a poly(A) tail. H15a diverges both upstream of position 1068, and beyond position 1280 (terminating after a further 27 bp). The sequence of the putative *mst*77F translation product (deduced from c18c) is also shown. The translation stop codon is indicated by an asterisk (*), and the putative phosphorylation site is underlined. Also underlined is the premature stop codon in the λgS9 sequence (residues 688–690). See text for other details.

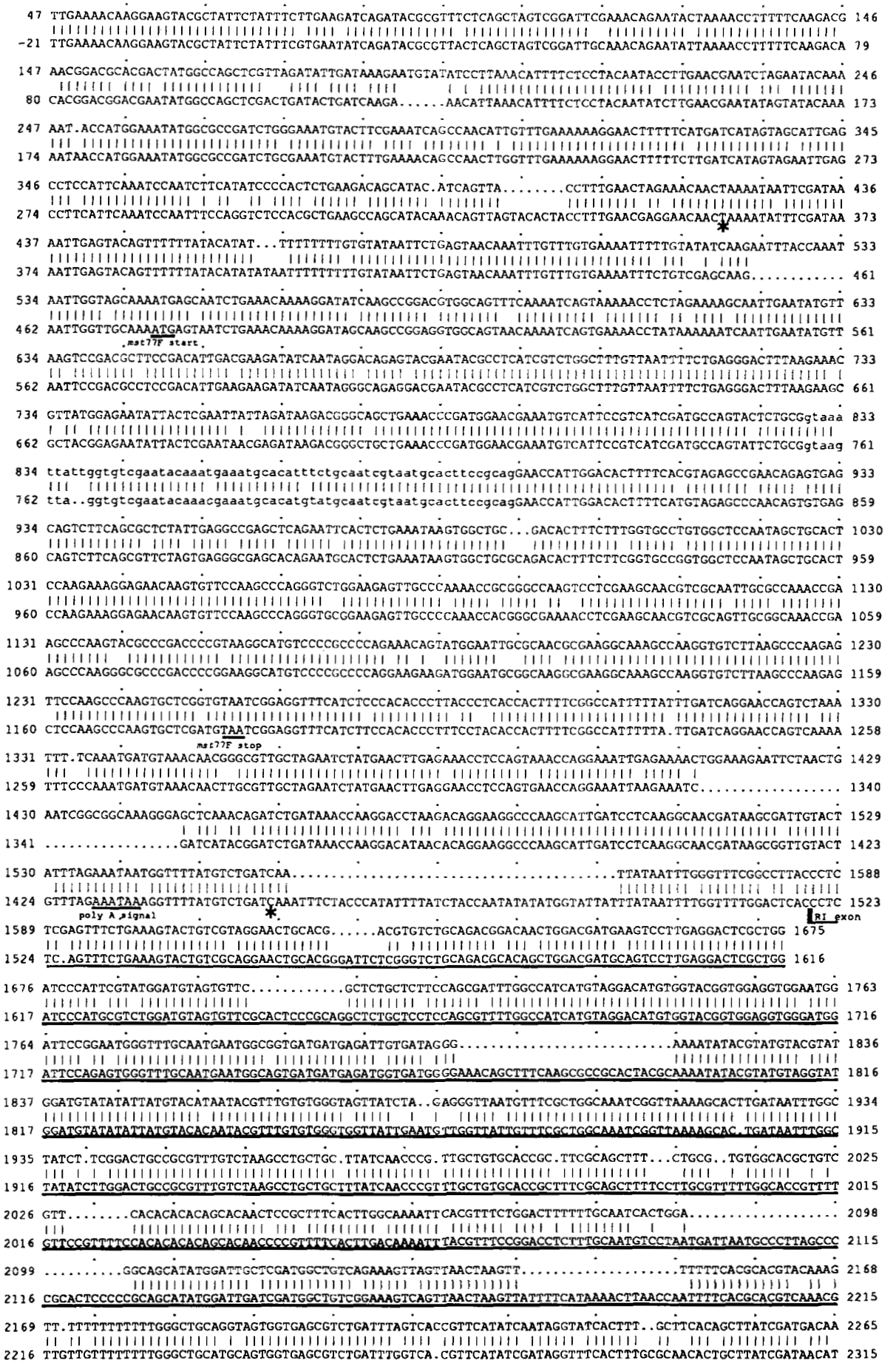

FIGURE 6

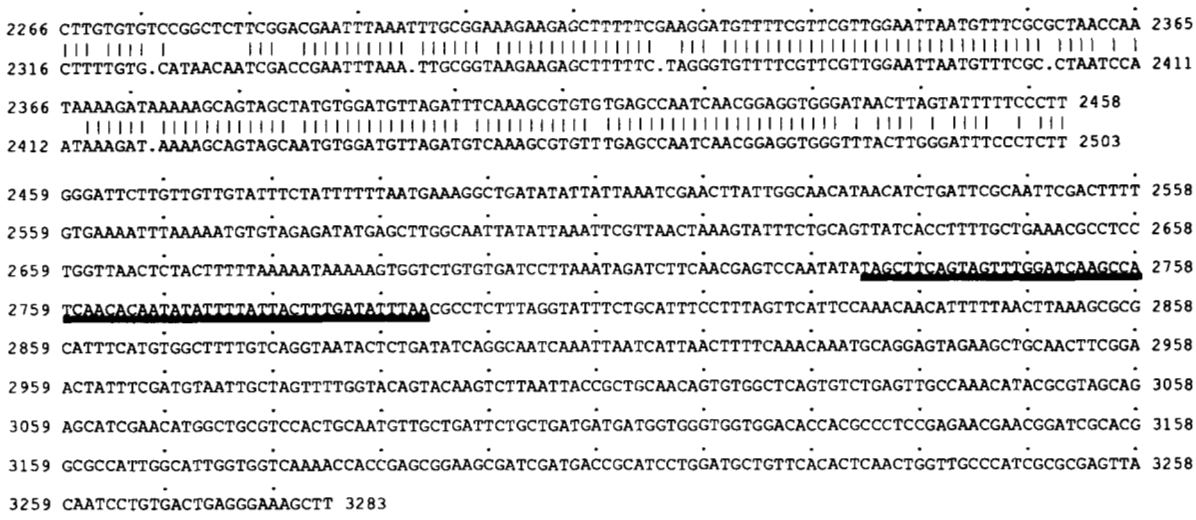

FIGURE 6.—Extended homology between the 77F locus (bottom strand, KALDERON and RUBIN 1988) and λgS9 (top strand) derived with the GCG BESTFIT program using default parameters. Numbering is from left to right with respect to the sequence of the 3.2-kb *Hin*dIII fragment of λgS9 shown in expanded form in Figure 1. Negative numbers denote nucleotides upstream of the limit of KALDERON and RUBIN's sequence in that direction. The *mst*77F intron is shown in lower case. Translation initiation and termination codons for the putative *mst*77F translation product (*mst* start and *mst* stop), and the putative polyadenylation signal of c18c (AATAAA) are underlined (KALDERON and RUBIN 1988). The asterisks under the T at position 421 and the C at position 1558 are respectively the 5′ and 3′ boundaries of c18c, the most extensive cDNA of the *mst*77F family (KALDERON and RUBIN 1988). The gene for the RI subunit of PKA is transcribed in opposite orientation to the *mst*77F gene and the extent of the longest cDNA for this gene is indicated by the thin underlining. The end of the homology between 77F and λgS9 (position 2189) lies within a region that includes heterogeneous RI transcription initiation sites (KALDERON and RUBIN 1988). Homology between the 3′ end of the cS1 cDNA clone and the genomic DNA is indicated by the thick underlining note that the sequence of this portion of cS1 is identical to that of λgS9, downstream of this position cS1 contains A residues.

cDNA cS1 (530 bp) is incomplete at its 5′ end. It has the 918–920 deletion and a number of single-base substitutions. Correspondence of cS1 with *mst*77F ceases beyond nucleotide position 1209, within the *mst*77F 3′ untranslated region (see also Figure 1). Thereafter, cS1 diverges for a further 60 nucleotides before it terminates with evidence of a poly(A) tail. The 60 divergent 3′ nucleotides of cS1 correspond to a region of λgS9 approximately 1.6 kb downstream of the boundary of homology with c18c (Figures 1 and 6). Since there is no evidence of a sequence related to the 5′ splice junction consensus at the point of divergence it is simpler to think of cS1 as representing a transcript of an internally deleted locus, rather than to invoke alternative splicing. At 25 bp upstream of the cS1 poly(A) tail is the sequence AATATA, not the most common of polyadenylation signals but one that is used on occasion (MANLEY 1988).

cDNA cS8 (479 bp) is also incomplete at its 5′ end. It has the 918–920 deletion, the 1131 deletion, a deletion of nucleotide 1180, and a number of single-base substitutions. Any cS8 polypeptide would terminate 24 amino acid residues downstream of the *mst*77F product (Figure 5). cS8 diverges from the *mst*77F nucleotide sequence beyond position 1340 (not shown), and terminates after a further 69 nucleotides with no evidence of a poly(A) tail. As the clone weakly detects an additional transcript on Northern blots and an additional band in genomic Southern blots (not shown), it is possible that its 3′ end represents a cDNA cloning artifact.

cDNA H15a (213 bp) is described as diverging from *mst*77F upstream of nucleotide position 1068, as having a change of reading frame at amino acid residue 199 and as diverging from *mst*77F for 27 nucleotides beyond nucleotide position 1280 (KALDERON and RUBIN 1988). Although KALDERON and RUBIN did not specify the nucleotide position corresponding to the frameshift, for the sake of data presentation we have taken it to be position 1131. The effect of the frameshift would be to cause the putative translation product to terminate after a further 19 amino acid residues. It was not reported whether the cDNA showed evidence of a poly(A) tail.

Meaningful pairwise comparison of the six nucleotide sequences of Figure 5 is complicated by their various degrees of completeness, as well as the significant divergence from the "canonical" *mst*77F sequence at some 5′ and 3′ ends. Looking closely at the differences between the six sequences, however, it is apparent that some nucleotide substitutions, insertions and deletions are shared between different family members, whereas others are unique. In particular, the λgS9, B4a, cS1, cS8 and H15a sequences (the "*Y* group"; see DISCUSSION) share a number of changes with respect to the *mst*77F sequence. Over the region corresponding to the putative *mst*77F coding domain, a significant number of these are third base substitutions.

How many of the sequences represent functional loci? Assuming that *mst*77F is indeed functional, and that c18c represents its transcript, the locus represented by λgS9 is unlikely to encode a functional polypeptide due to the premature stop codon at amino acid residue 72. Both H15a and cS8 show no homology with *mst(77)F* at their 3′ ends and If translated, both would result in polypeptides having carboxy termini that are longer and substantially different in character than that of the *mst*77F polypeptide. Until we have full-length cDNA clones we cannot comment further on the genes encoding these cDNAs. cS1 which may be incomplete at its 5′ end due to premature termination of cDNA synthesis shares the same termination codon and basic structure as *mst*77F, and thus may represent a functional locus. Finally, B4a, despite having a frameshift at amino acid residue 189 and diverging from *mst*77F within the coding domain, is able to encode a polypeptide that is neither radically different in length from the *mst*77F product, nor altogether different in terms of the nature of its terminal residues. Thus it too may encode a polypeptide with functional similarity to *mst77F*.

Structure of the locus represented by λgS9: To learn more about the nature of the event(s) that gave rise to *mst*77F-like genes on the *Y* chromosome, we have sequenced the entire 3.2-kb *Hin*dIII fragment of λgS9, shown in expanded form in Figure 1. This fragment, containing all of the homology with male cDNA and only a single member of the *mst*77F family, has more extensive similarity with DNA in and around the *mst*77F locus than might have been expected. The extended 77F and λgS9 sequences, aligned in Figure 6, have approximately 92% sequence identity. 5′ to the *mst(77)F* gene the homology extends at least 500 nucleotides upstream of the initiation codon. In addition the intron-exon structure of the male-specific gene is conserved, there are only 5 base changes within the 62-bp intron (Figure 5). This indicates, at least in the case of this particular *Y*-linked copy, that presence on the *Y* chromosome is not due to integration of a reverse-transcript of a *mst*77F mRNA. At the 3′ end, homology between the 77F locus and λgS9 extends for several hundred base pairs beyond the poly(A) addition site of c18c, and into transcribed sequences of the gene for the RI subunit of PKA. Polymorphism (notably insertion/deletion) is most evident 3′ to the *mst*77F gene. It may be that sequence drift in the 5′ region has been subject to more constraint due to functional selection of some kind, for example important regulatory sequences may need to be conserved for appropriate expression. Assuming that c18c represents a prematurely terminated cDNA of a 1.4-kb polyadenylated mRNA (see below), some of the 5′ region is likely to correspond to 5′-untranslated sequences. A search for promotor motifs upstream of the c18c 5′ end provides little compelling evidence for standard promoter elements. A 5/8 match to the TATA-box consensus is found centered around nucleotide 320, 30 bp upstream of the c18c 5′ end and a 7/9 match to the CAAT-box consensus is present around position 197, 153 bp upstream of the c18c 5′ end. In the absence of primer extension data, however, we cannot firmly identify the transcription start-site of *mst*77F. We have also searched for promoter elements present in other male germ line-specific genes such as β_2-tubulin and *mst*87F but failed to detect any similarities.

We have also investigated the nature of λgS9 sequences which flank the region with autosomal homology. The 2.2-kb *Bam*HI-*Sal*I fragment of Figure 1 detects a family of repetitive DNA sequences containing a few *Y*-linked and many autosomal copies (Figure 7). The two relatively intense bands visible in lane 1 (male DNA) have not been investigated further. The 1.8-kb *Eco*RI fragment of Figure 1 detects a fragment of its own size in both male and female DNAs (Figure 7). The intensity of the corresponding band suggests that the 1.8-kb fragment lies internal to a repetitive DNA sequence. Data from limited restriction mapping are consistent with the repetitive sequence being a transposable element of the *calypso* class (COTE *et al.* 1986). In summary λgS9 DNA, outside of the region of extended homology with the 77F locus, has characteristics that might be expected of the heterochromatic *Y* chromosome.

The *mst(77)F* gene product: The c18c cDNA clone sequenced by KALDERON and RUBIN (1988) contains a long open reading frame (ORF) which translates as a 215-amino acid polypeptide. They assigned no function to this putative translation product but noted that it contained a basic C-terminal domain. We have used the computer programs Prosrch and BLAST to search the protein databases with the sequence of the *mst*77F polypeptide. Using Prosrch (COLLINS, COULSON and LYALL 1988) to search the Pir28 protein database, highest scores (100–120) were obtained against various examples of the vertebrate histone H1/H5 family. Chicken erythrocyte H5, in particular, can be aligned with the *mst*77F polypeptide to display 26% identity and 45% similarity when conservative changes are allowed (Figure 8A). It should not be overlooked, however, that correspondence over the C-terminal 75 amino acids is in large part due to the predominance of basic residues in this region of both proteins. The contribution of amino acid constitution to the match is further emphasized by relatively high GCG-BESTFIT quality scores between a randomized *mst*77F polypeptide and chicken H5. Finally, the *mst*77F polypeptide is only slightly more like chicken H5 than like chicken H1 or Drosophila H1 (≈20% identity, 35% similarity). In contrast, likeness between

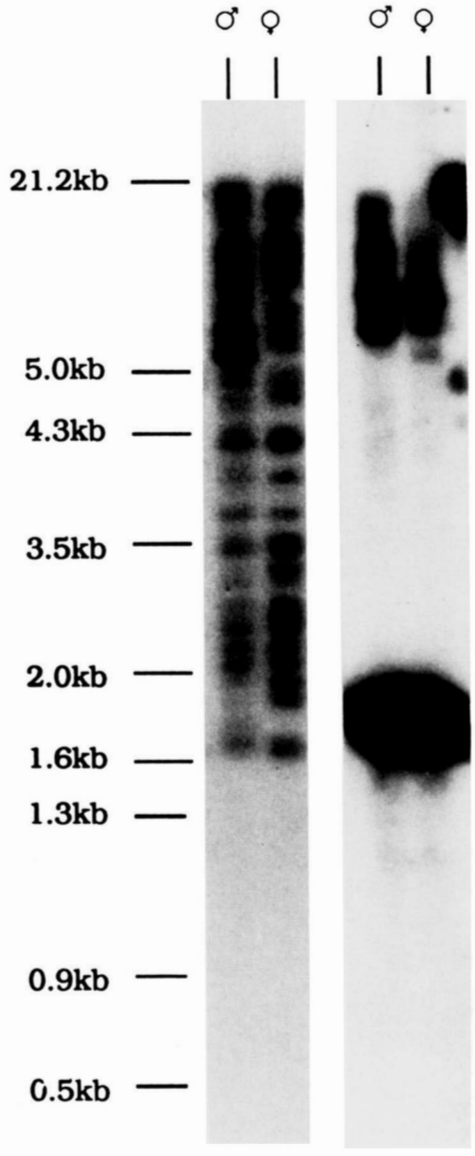

FIGURE 7.—Southern blots of *Eco*RI-cleaved male and female *Oregon R* DNA washed at high stringency. The left hand two lanes were probed with the 2.2-kb *Bam*HI-*Sal*I fragment of λgS9, and the right hand lanes with the 1.8-kb *Eco*RI fragment (Figure 1). The two probes were of comparable specific activity and the two blots, each representing the same amount of the same pair of cleaved DNAs, were exposed for autoradiography for the same time.

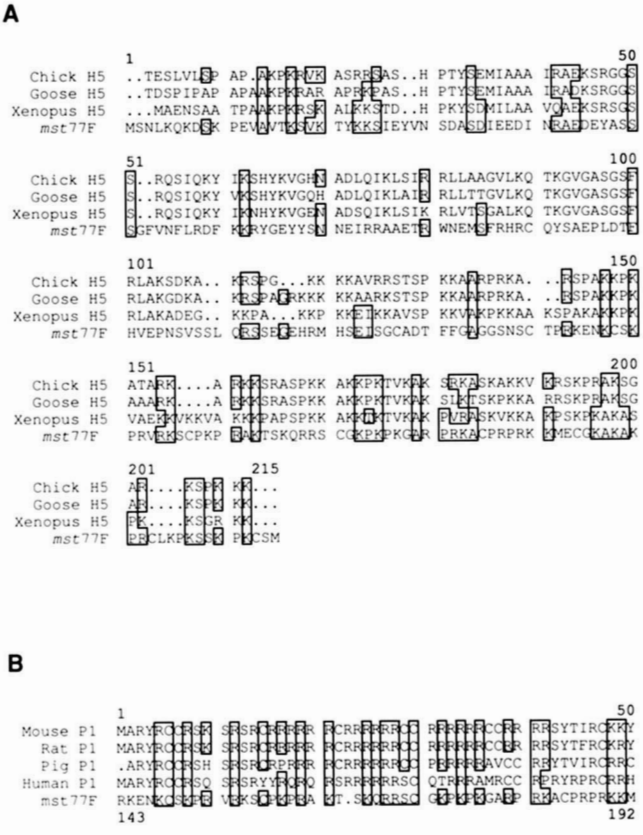

FIGURE 8.—Alignments of polypeptide sequences. (A) Alignment of the indicated polypeptide sequences using the GCG PILEUP program with default parameters. The *mst*77F sequence, to which the numbering refers, is the putative translation product of cDNA clone c18c [KALDERON and RUBIN (1988); Swissprot accession number: P16909]. Identities between the c18c translation product and the other polypeptides are boxed. PIR accession numbers are as follows: chicken H5, A29179; goose H5, A02588; Xenopus H5, JT0403. (B) Alignment of the putative *mst*77F polypeptide with mature mammalian protamine 1 sequences using GCG PILEUP with default parameters. Cysteine and basic (arginine and lysine) residues shared with the *mst*77F polypeptide are boxed. Numbering above the figure refers to rat protamine P1. Numbering below the figure refers to the *mst*77F polypeptide. Swissprot accession numbers are as follows: rat, P10118; mouse, P02319; human, P04553; pig, P19757.

chicken H5 and either chicken H1 or Drosophila H1 is significantly greater (≥35% identity, 50% similarity).

If it is a relative of H1 and H5, the *mst*77F polypeptide is a rather distant one. Secondary structure predictions, however, made using the computer programme ALB (PTITSYN and FINKELSTEIN 1989), tend to support the identification of the *mst*77F polypeptide as a member of the H1/H5 family. The first 10 amino acids are very polar. Subsequent sequence suggests a globular head region (residues 11–135) in which the dominant secondary structure is α-helix. This is followed by a basic tail (residues 136–215) in which lysine, arginine and proline are frequently clustered, and that would be unstructured in free solution. Overall the structure is reminiscent of the those of H1 and H5, both of which have disordered N- and C-terminal domains flanking a central globular domain [see ISENBERG (1979) and BRADBURY (1992) for reviews]. Phosphorylation of members of the H1 family, occurring at serine or threonine residues in coordination with the cell division cycle is thought to activate their chromatin-condensing functions (HOHMANN 1983; BRADBURY 1992). In addition a consensus kinase motif of the form S(T)PK(R)K(R) has been identified in the N and C termini of histone H1, it is expected that phosphorylation of this site would abolish the interaction between H1 and DNA (CHURCHILL and SUZUKI

1989). The *mst*77F polypeptide contains one such sequence (residues 141–144, underlined in Figure 5), close to the predicted boundary between the globular domain and the basic tail.

Using the computer program BLAST (ALTSCHUL *et al.* 1990) to search the Swissprot protein database with the sequence of the *mst*77F polypeptide gave slightly different results. As well as identifying members of the H1/5 class high scores were also obtained between a segment of the basic tail of the *mst*77F polypeptide and cysteine-rich protamines from rat, pig, mouse and man. Similarities were approximately 50% (Figure 8B), in large part reflecting the predominance of basic residues but also involving cysteines. Protamines are short and predominantly basic polypeptides that complex with DNA. They are the major component of the sperm chromatin of many organisms (BLOCH 1969), their role being to assist compaction of the genome into the restricted volume of the sperm head. The cysteine-rich variety is typical of protamines found in mammalian sperm (SUBIRANA 1983; YELICK *et al.* 1987). The cysteines are thought to stabilize compacted chromatin via disulfide bridges (BEDFORD and CALVIN 1974). Although it is rare for cysteine residues to occur within histones H1 and H5, the *mst*77F polypeptide has 10, 8 of which are fairly regularly spaced throughout its basic tail (see Figure 8A).

Phosphorylation of serine residues by protamine kinase is thought to stimulate the binding of protamines to DNA where they replace histones or transition proteins, become dephosphorylated and condense DNA (BELLVE, ANDERSON and HALEY-BOWDOIN 1975; OLIVA and DIXON 1991). Although serine residues make up 10% of the basic tail and 12% of the full-length *mst*77F polypeptide, a relatively high serine content is also a feature of the HI/H5 family as a whole.

In summary, the *mst*77F polypeptide has characteristics suggestive of a distant member of the histone H1/H5 family. The C-terminal basic tails of H1 and H5 ("linker" histones) play a role in chromatin condensation via interaction with DNA. In H1 the basic residue is almost exclusively lysine. In H5, associated in chicken erythrocytes with chromatin inactivation, there are significant amounts of the more basic residue, arginine. In this respect the *mst*77F polypeptide is more like H5 than H1. The relationship between the *mst*77F polypeptide and cysteine-rich protamines seems relatively tenuous. It does, however, serve to point out the interspersion of cysteine residues within the basic tail, a feature that distinguishes the *mst*77F polypeptide from all other members of the H1/H5 family.

DISCUSSION

Male-specific transcripts with autosomal and *Y*-linked genes: We have identified a family of genes transcribed in male but not female *D. melanogaster*. Transcription of any of the copies requires the presence of a male germ line, and preliminary *in situ* hybridization data suggest that this is restricted to spermatocytes (S. R. H. RUSSELL, unpublished). The member of the family that has been characterized in greatest detail is *mst*77F, present at cytological position 77F on the left arm of chromosome *3* (KALDERON and RUBIN 1988). It is contained within the large intron of the gene for an RI subunit of cAMP-dependent PKA (Figure 1). Southern blot analysis of male and female DNAs (Figures 3 and 4) points to *mst*77F-like sequences being present on the *Y* chromosome of *D. melanogaster*. Evidence that these represent members of an *mst*77F gene family is provided by genomic DNA clone λgS9, and by cDNAs B4a, cS1, cS8 and H15a (the *Y* group). The insert of λgS9 (Figure 1), clearly derived from the *Y* chromosome, contains a single *mst*77F-like gene (≈8% divergence) having the intron-exon structure of the autosomal locus embedded within related flanking sequences (Figures 5 and 6). Each of the four cDNA sequences differs sufficiently from the sequence of the autosomal transcript (≈5% divergence) that they too must represent *Y*-linked members (Figure 5).

As to the number of *Y*-linked loci we can consider the number of bands detected by cS1 in Southern blots of male *Oregon R* DNA (Figures 3 and 4). There are eight bands in total, one of which represents the autosomal *mst*77F locus (5.1 kb). Three of the remaining bands (0.5, 4.2 and 5.1 kb) represent the family member present within λgS9 (Figure 1), leaving four (3.1, 3.8, 4.5 and 10 kb) to accommodate further members of the family. Thus we estimate that there are five *Y*-linked loci assuming that each non-λgS9 band represents a single locus. It appears be the case that some *Canton S* derivatives have additional *Y*-linked members, since the *Y* chromosomes of the *XXY*, *T(X;Y)* and *Df(Y)* individuals represented in Figure 4 were derived from a *Canton S* background (M. GATTI, personal communication).

Although members of the *Y* group are all more closely related to one another than they are to the autosomal *mst*77F locus (apparent at the DNA sequence level in terms of shared nucleotide substitutions, insertions and deletions), divergence still varies between 1.5 and 6% over the region of available overlap between any two members of the group (Figure 5). Can this be taken to indicate that each DNA clone from the *Y* group represents an independent locus, or might it be argued that some of these clones represent strain differences at a single locus? Our own sequences (λgS9, cS1 and cS8) all represent an *Oregon R* population maintained in Glasgow while the cDNA clones c18c B4a and H15a and the genomic clone 18c are derived from a *Canton S* background (KALDERON

and RUBIN 1988). Although the pattern generated by hybridization of cS1 to *Canton S* DNA resembles that of *Oregon R* DNA (Figure 4A), sequence polymorphisms of this level may still be present which would not be revealed by the sensitivity of the Southern blotting experiments. Thus we cannot determine how many *Y* loci are represented by the cloned DNA. In conclusion, although there is more that one family member on the *Oregon R Y* chromosome, strain differences at a single locus may account for the differences between some of our cloned DNA.

Are the *Y*-linked members functional? In all of the preceding discussion, the assumption has been made that the autosomal *mst*77F locus is functional, and we have thus used its sequence as a standard against which to compare sequences representing *Y*-linked members. This assumption cannot be proven on the basis of the available data. It is relevant, however, that of several sibling Drosophila species we have investigated, all except *D. melanogaster* have a single member of the *mst*77F family (Figure 3), and in the case of *D. simulans* it resides at the cytological position equivalent to 77F of *D. melanogaster*. For reasons of karyotype similarity the same is likely to be true of *D. mauritiana* and *D. sechellia*. Presumably it is a functional locus in these species.

Accepting *mst*77F as functional, we must conclude that the locus represented by λgS9 would be unable to encode a functional polypeptide due to its premature stop codon (Figure 5). Moreover, for reasons of either truncation or divergence, none of the cDNAs representing *Y*-linked members contains a full-length coding domain (Figure 5). We may reasonably question, therefore, whether any of the *Y*-linked members are capable of producing a functional polypeptide. Only in the case of the loci represented by cS1, and possibly that represented by B4a, does this seem likely.

Expression of *Y*-linked members: KALDERON and RUBIN (1988) discussed a second gene within the large RI intron, though they did not report its sequence. This is 13a, transcribed in the same direction as *mst*77F. Its 3′ end lies approximately 350 bp upstream of the most 5′ nucleotide present in 18c, a cDNA representing the autosomal *mst*77F locus (KALDERON and RUBIN 1988). This places the 3′ end of 13a very close to the 5′ boundary of the sequence shown in Figure 5, and delimits the initial 350 bp as a region that one might reasonably investigate for *mst*77F promoter activity. Motifs suggestive of known promoter elements are not convincing; however, a complete analysis awaits primer extension and transformation studies.

Isolation of cDNA clones representing *Y*-linked members demonstrates that at least some of the latter are transcribed. The prevalence of the clones within cDNA libraries implies transcription at levels comparable to the autosomal locus, in agreement with the reduced levels of transcripts in *XO* males (Fig. 2). Moreover, a Northern blot of poly(A^+) RNA isolated from *XY* individuals has no more bands than does the blot of *XO* RNA, suggesting that *Y*-linked transcripts are all polyadenylated mRNAs of approximately the same size as transcripts of the autosomal gene (Figure 2). There is evidence for polyadenylation from at least some cDNA sequences. Although the *Y*-linked member represented by λgS9 has a domain corresponding to the putative *mst*77F promoter region, there are a number of insertions and deletions (Figure 6). Single base substitution, by comparison, is not significantly more apparent than within the coding domain. These data tend to suggest that *Y*- linked members are expressed conventionally from *mst*77F-like promoters, and with the same temporal and developmental specificity as the autosomal locus.

Genetic screens for male-sterile mutations (either deletions or EMS induced) have provided no evidence for essential fertility functions in the region of the chromosome where the *Y*-linked members map (Figure 4C; KENNISON 1981, 1983; GATTI and PIMPINELLI 1983). Even if one or more *Y*-linked members were able to give rise to a functional polypeptide, however, this could be accounted for by full or partial complementation of fertility by the autosomal locus. We are initiating a mutational analysis of the autosomal locus in order to address this question.

Finally, the apparently normal expression of the autosomal copy in *XO* males observed on Northern blots (Figure 2) would tend to rule out any role for *Y*-linked copies in splicing transcripts of the autosomal locus. This is in contrast with the *Y*-linked *Su(Ste)* locus of *D. melanogaster*, which is required for appropriate splicing of transcripts of the *X*-linked *Stellate* locus (LIVAK 1990).

Origin of *Y*-linked members: Members of the *Y* group are all more closely related to one another than to *mst*77F suggesting that they result from gene duplication following the transposition of a single copy of the *mst*77F locus to the *Y*. This is particularly apparent in the case of the in-frame 918–920 deletion which is present in all of the putative *Y* copies (Figure 5). In this respect it is worth noting that the autosomal locus is itself in an unusual genomic location within the intron of another gene. As the autosomal location appears to be the same in the case of *D. melanogaster* and *D. simulans* it is probable that the autosomal copy transposed to the 77F locus in a common ancestor and then to the *D. melanogaster Y* after divergence. The significant proportion of third base changes over the *mst*77F coding domain, reflecting a high proportion of synonymous/conservative changes, implies that there may have been at one time selection for function of the *Y* loci.

Transposition to the *Y* involved not merely the *mst*77F locus. The family member represented by λgS9, for example (Figures 1 and 6), is embedded within a 77F-like domain that extends downstream to include the first exon of the gene for the RI subunit of PKA. In the case of other family members, correspondence with 77F may extend at least several hundred base pairs further in this direction (S. F. GOODWIN, unpublished). Upstream correspondence between λgS9 and the autosomal locus extends for at least several hundred bp beyond the most 5′ nucleotide found in cDNAs of the *mst*77F family (Figure 6).

Although transposition must have played a role in generating the *Y*-linked members of the *mst*77F family, there is a fairly high degree of conservation of their genomic organization across a range of *D. melanogaster* isolates from different geographic locations. Such conservation would not be expected for actively transposing sequences, which tend to show considerable variation between strains (RUBIN 1983; ISING and BLOCK 1984), and might be taken to suggest some form of functional constraint. On the other hand, since we are ignorant of the mechanism by which transposition has taken place, it may merely be an accident of history. This situation is in sharp contrast to the *Y*-associated sequences isolated from *D. hydei* which are not conserved between different *hydei* strains (BRAND and HENNIG 1989; HARAVEN, ZUCKERMAN and LIFSCHYTZ 1986).

Y-linked copies are not present in the sibling species *D. simulans*, *D. mauritiana* and *D. sechellia*. *D. simulans* and *D. mauritiana* also have a substantially lower copy number of *Y*-linked *Su(Ste)* sequences than does *D. melanogaster* (LIVAK 1984), and the *D. simulans Y* is devoid of functional *rRNA* genes (LOHE and ROBERTS 1990). Thus the *Y* chromosomes of even very closely related species can be substantially different. Such differences may in some way drive or consolidate reproductive isolation.

The *mst*77F gene product: The putative *mst*77F product has a sequence and predicted secondary structure suggestive of a distant relative of the histone H1/H5 ("linker" histone) family, complemented by an array of cysteine residues in its basic tail that invites comparison with mammalian cysteine-rich protamines (Figure 8). We are not aware of such a structure having been described previously. Indeed, cysteines are notable by their absence from all but the occasional member of the H1/H5 family, and have never been previously observed in the basic tail (WELLS and MCBRIDE 1989). By analogy with the cysteine-rich protamines, their role in the *mst*77F polypeptide may be tight compaction of chromatin via disulfide bridges. In this context it is also worth noting the relatively high arginine:lysine ratio of the basic tail. Such a feature is characteristic of H5 rather than H1, and again may reflect a role in relatively tight compaction of chromatin since chicken erythrocyte H5, the archetypal member of the H5 group, is implicated in chromosome inactivation.

A role in sperm chromatin condensation? Prior to this report, only two *D. melanogaster* histone genes lying outside the somatic histone gene cluster on chromosome 2 had been described. These are the H2A variant, H2vd (VAN DAAL *et al.* 1988), and the H3 variant, H3.3 (FRETZIN *et al.* 1991). For many other organisms, in contrast, several variants of both core and linker histones are known, often developmentally regulated (STEIN, STEIN and MARZLUFF 1984; WU *et al.* 1986). The range of variation is generally greater for linker histones than for core histones (ISENBERG 1979). Though the physiological basis of this variation is, on the whole, poorly understood, a substitution of somatic histones for sperm-specific DNA-binding proteins during the latter stages of spermatogenesis has been described for many species (SUBIRANA 1983). Sperm-specific substitutes range from relatively slight variants of the somatic histones as seen in sea urchins and starfish, through to the short basic protamines found in mammals and some species of fish. They appear to mediate the compaction of chromatin into the small volume of the sperm head. In the case of *D. melanogaster*, several studies have suggested a histone transition during spermatogenesis (DAS, KAUFMANN and GAY 1964a,b; HAUSCHTECK-JUNGEN and HARTL 1982). In *D. hydei*, moreover, somatic H1 immunoreactivity is undetectable during postmeiotic stages of spermatogenesis (KREMER, HENNIG and DIJKHOF 1986), suggesting replacement by some other protein(s). The *mst*77F polypeptide is the first candidate for a specialized DNA binding protein involved in compaction of Drosophila sperm chromatin.

Clearly further experimentation if required to determine the precise role of the *mst*77F polypeptide. One possibility is that, like the cysteine-rich protamines, it is involved in the terminal stages of sperm chromatin condensation. Alternatively, it may function as a transition protein to assist partial chromatin condensation prior to the appearance of protamines. Such proteins have been described for rat (GRIMES *et al.* 1977). Yet another possibility is that the *mst*77F polypeptide is a precursor that is subsequently processed to yield an active polypeptide, as in the case of mouse and human protamine 2 (YELICK *et al.* 1987; AMMER, HENSCHEN and LEE 1986). A combination of biochemical and immunohistochemical studies will help to resolve these questions.

We wish to thank M. GATTI, M. MCKEOWN, B. TAYLOR and M. WOLFNER for *D. melanogaster* strains; D. KALDERON for DNA and helpful discussion; M. ASHBURNER, A. CARPENTER, G. HEIMBECK and F. HUNDLEY for critical reading of the manuscript; and C. CRANE-ROBINSON for help with protein structure prediction. We would also like to acknowledge the anonymous reviewers for critical

comments which improved the manuscript. This work was supported entirely by the United Kingdom Medical Research Council with grants to S.R.H.R, K.K., and M. ASHBURNER.

LITERATURE CITED

ALTSCHUL, S. F., W. GISH, W. MILLER, E. W. MYERS and D. J. LIPMAN, 1990 Basic local alignment search tool. J. Mol. Biol. **215:** 403–410.

AMMER, H., A. HENSCHEN and C. H. LEE, 1986 Isolation and amino-acid sequence analysis of human sperm protamines P1 and P2. Biochemistry **367:** 515–522.

ASHBURNER, M., 1989 *Drosophila: A Laboratory Manual.* Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

BEDFORD, J. M., and H. I. CALVIN, 1974 The occurrence and possible significance of –S–S– crosslinks in sperm heads with particular reference to eutherian mammals. J. Exp. Zool. **188:** 137–156.

BELLVE, A. R., E. ANDERSON and L. HANLEY-BOWDOIN, 1975 Synthesis and amino-acid composition of basic proteins in mammalian sperm nuclei. Dev. Biol. **47:** 349–365

BELOTE, J. M., and B. S. BAKER, 1983 The dual function of a sex determination gene in *Drosophila melanogaster*. Dev. Biol. **95:** 512–517.

BLOCH, D. P., 1969 A catalogue of sperm histones. Genetics **61** (Suppl.): s93–s111.

BONACCORSI, S., and A. LOHE, 1991 Fine mapping of satellite DNA sequences along the *Y* chromosome of *Drosophila melanogaster*: relationships between satellite sequences and fertility factors. Genetics **129:** 177–189.

BONACCORSI, S., C. PISANO, F. PUOTI and M. GATTI, 1988 *Y* chromosome loops in *Drosophila melanogaster*. Genetics **120:** 1015–1034.

BOSWELL, R. E., and A. P. MAHAWOLD, 1985 *tudor*, a gene required for assembly of the germ plasm in *Drosophila melanogaster*. Cell **43:** 97–104.

BRADBURY, E. M., 1992 Reversible histone modifications and the chromosome cell cycle. Bioessays **14:** 9–16.

BRAND, R. C., and W. HENNIG, 1989 An abundant testis RNA species shows sequence similarity to *Y* chromosomal and other genomic sites in *Drosophila hydei*. Mol. Gen. Genet. **215:** 469–477.

BRIDGES, C. B., 1916 Non-disjunction as proof of the chromosomal theory of heredity. Genetics **1:** 1–52, 107–163.

BROWN, E. H., and R. C. KING, 1960 Studies on the expression of the *transformer* gene of *Drosophila melanogaster*. Genetics **46:** 143–156.

CHOMCZYNSKI, P., and N. SACCHI, 1987 Single step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal. Biochem. **162:** 156–159.

CHURCHILL, M. E. A., and SUZUKI, M., 1989 SPKK motifs prefer to bind to DNA at A/T-rich sites. EMBO J. **8:** 4189–4195

COLLINS, J. F., A. F. W. COULSON and A. LYALL, 1988 The significance of protein-sequence simularities. Comput. Appl. Biosci. **4:** 67–71.

COTE, B., W. BENDER, D. CURTIS and A. CHOVNICK, 1986 Molecular analysis of the *rosy* locus of *Drosophila melanogaster*. Genetics **112:** 769–783.

DANILEVSKAYA, O. N., E. V. KURENOVA, M. N. PAVLOVA, D. V. BEBEHOV, A. J. LINK, A. KOGA, A. VELLEK and D. L. HARTL., 1991 He-T family DNA sequences in the *Y* chromosome of *Drosophila melanogaster* share homology with the *X*-linked Stellate genes. Chromosoma **100:** 118–124.

DAS, C. C., B. P. KAUFMANN and H. GAY, 1964a Autoradiographic evidence of synthesis of an arginine-rich histone during spermiogenesis in *Drosophila melanogaster*. Nature **204:** 1008–1009.

DAS, C. C., B. P. KAUFMANN and H. GAY, 1964b Histone protein transition in *Drosophila melanogaster*. Exp. Cell Res. **35:** 507–514.

DE LOOS, F., R. DIJKHOF, C. J. GROND and W. HENNIG, 1985 Lampbrush chromosome loop-specificity of transcript morphology in spermatocyte nuclei of *Drosophila hydei*. EMBO J. **3:** 2845–2849.

DEVEREUX, J., P. HAEBEREL and O. SMITHIES, 1984 A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12:** 387–395.

DIBENEDETTO, A. J., D. M. LAKICH, W. D. KRUGER, J. M. BELOTE, B. S. BAKER and M. F. WOLFNER, 1987 Sequences expressed sex-specifically in *Drosophila melanogaster* adults. Dev. Biol. **119:** 242–251.

FEINBERG, A. P., and B. VOGELSTEIN, 1983 A technique for radiolabelling DNA restriction fragments to high specific activity. Anal. Biochem. **132:** 6–13.

FRETZIN, S., B. D. ALLAN, A. VANDAAL and S. R. C. ELGIN, 1991 A *Drosophila melanogaster* H3.3 cDNA encodes a histone variant identical with vertebrate H3.3*. Gene **107:** 341–342.

FYRBERG, E. A., B. J. BOND, N. D. HERSHEY, K. S. MIXTER and N. DAVIDSON, 1981 The actin genes of Drosophila are highly conserved but intron positrons are not. Cell **24:** 107–116.

GATTI, M., and G. PIMPINELLI, 1983 Cytological and genetic analysis of the *Y* chromosome of *Drosophila melanogaster*. I. Organisation of the fertility factors. Chromosoma **88:** 349–373.

GOLDSTEIN, L. S. B., R. W. HARDY and D. L. LINDSLEY, 1982 Structural genes on the *Y* chromosome of *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA **79:** 7405–7409.

GONCZY, P., S. VISWANATLHAN and S. DINARDO, 1992 Probing spermatogenesis in *Drosophila* with P-element enhancer detectors. Development **114:** 89–98.

GORALSKI, T., J. E. EDSTROM and B. S. BAKER, 1989 The sex determination locus *transformer-2* of *Drosophila* encodes a polypeptide with similarity to RNA binding proteins. Cell **56:** 1011–1018.

GRIMES, S. R., M. L. MEISTRICH, R. D. PLATZ and L. S. HNILICA, 1977 Nuclear protein transitions in rat testis spermatids. Exp. Cell Res. **110:** 31–39.

GROND, C. J., R. G. RUTTEN and W. HENNIG, 1984 Ultrastructure of the *Y* chromosomal lampbrush loops in primary spermatocytes of *Drosophila hydei*. Chromosoma **89:** 85–95.

HACKSTEIN, J. H. P., 1987 Spermatogenesis: Genetic aspects. Results Probl. Cell Differ. **15:** 63–116.

HARAVEN, D., M. ZUCKERMAN and E. LIFSCHYTZ, 1986 Origin and evolution of the transcribed repeated sequences of the *Y* chromosome lampbrush loops of *Drosophila hydei*. Proc. Natl. Acad. Sci. USA **83:** 125–129.

HARDY, R. W., K. T. TOKUYASU and D. L. LINDSLEY, 1981 Analysis of spermatogenesis in *Drosophila melanogaster* bearing deletions for *Y* chromosome fertility genes. Chromosoma **83:** 593–617.

HAUSCHTECK-JUNGEN, E., and D. HARTL, 1982 Defective histone transition during spermiogenesis in heterozygous *Segregation Distorter* males of *Drosophila melanogaster*. Genetics **101:** 57–69.

HAZELRIGG, T., P. FORNILI and T. C. KAUFMAN, 1982 A cytogenetic analysis of X-ray induced male steriles on the *Y* chromosome of *Drosophila melanogaster*. Chromosoma **87:** 591–610.

HEITZ, E., 1933 Zytologische Untersuchungen an Dipteren. III. Die somatische Hyteropyknose bei *Drosophila melanogaster* und ihre genetische Bedeutung. Z. Zellforsch. Mikrosk. Anat. **20:** 237–287.

HENIKOFF, S., 1984 Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene **28:** 351–359.

HENNIG, W., 1985 *Y* chromosome function and spermatogenesis in *Drosophila hydei*. Adv. Genet. **23:** 179–234.

HOHMANN, P., 1983 Phosphorylation of H1 histones. Mol. Cell. Biochem. **57:** 81–92.

ISENBERG, I., 1979 Histones. Annu. Rev. Biochem. **48:** 159–191.

ISING, G., and K. BLOCK, 1984 A transposon as a cytogenetic marker in *Drosophila melanogaster*. Mol. Gen. Genet. **196:** 6–16.

KALDERON, D., and G. M. RUBIN, 1988 Isolation and characterisation of *Drosophila* cAMP-dependant protein kinase genes. Genes Dev. **2:** 1539–1556.

KAUFMANN, B. P., 1934 Somatic mitoses of *Drosophila melanogaster*. J. Morphol. **56:** 125–155.

KENNISON, J. A., 1981 The genetic and cytological organisation of the *Y* chromosome of *Drosophila melanogaster*. Genetics **98:** 529–548.

KENNISON, J. A., 1983 Analysis of *Y*-linked mutations to male sterility in *Drosophila melanogaster*. Genetics **103:** 219–234.

KREMER, H., W. HENNIG and R. DIJKHOF, 1986 Chromatin organisation in the male germ-line of *Drosophila hydei*. Chromosoma **94:** 147–161.

LIFSCHYTZ, E., 1987 The genetic control of spermatogenesis. Int. Rev. Cytol. **109:** 211–258.

LINDLSEY, D. L., and E. LIFSCHYTZ, 1972 The genetic control of spermatogenesis in *Drosophila*, pp. 203–222 in *The Genetics of Spermatozoon: Proceedings of the International Symposium*, Edinburgh, edited by R. A. BEATTY and S. GLUECKSOHN-WAELSCH.

LINDLSEY, D. L., and K. T. TOKUYASU, 1980 Spermatogenesis, in *The Genetics and Biology of Drosophila*, edited by M. ASHBURNER and T. R. F. WRIGHT. Academic Press, London.

LINDLSEY, D. L., and G. ZIMM, 1992 *The Genome of Drosophila melanogaster*. Academic Press, New York.

LIVAK, K. J., 1984 Organisation and mapping of a sequence on the *Drosophila melanogaster X* and *Y* chromosomes that is transcribed during spermatogenesis. Genetics **107:** 611–634.

LIVAK, K. J., 1990 Detailed structure of the *Drosophila melanogaster stellate* genes and their transcripts. Genetics **124:** 303–316.

LOHE, A., and P. A. ROBERTS, 1990 An unusual *Y* chromosome of *Drosophila simulans* carrying amplified rDNA spacer without rRNA genes. Genetics **125:** 399–406.

MANLEY, J. L., 1988 Polyadenylation of mRNA precursors. Biochim. Biophys. Acta **950:** 1–12.

MICHIELS, F., A. GASCH, B. KALTSCHMIDT and R. RENKAWITZ-POHL, 1989 A 146 bp promoter element directs the testis specificity of the *Drosophila* β_2-tubulin gene. EMBO J. **8:** 1559–1565.

OLIVA, R., and G. H. DIXON, 1991 Vertebrate protamine genes and the histone-to-protamine replacement reaction. Progr. Nucleic Acid Res. Mol. Biol. **40:** 49–52.

OLIVIERI, G., and A. OLIVIERI, 1966 Autoradiographic study of nucleic acid synthesis during spermatogenesis in *Drosophila Melanogaster*. Mutat. Res. **2:** 366–380.

PIMPINELLI, S., G. SANTINI and M. GATTI, 1978 ^{3}H-Actinomycin-D binding to mitotic chromosomes of *Drosophila melanogaster*. Chromosoma **66:** 389–395.

PTITSYN, O. B., and A. V. FINKELSTEIN, 1989 Prediction of protein secondary structure based on physical theory–histones. Protein Eng. **2:** 443–447.

RUBIN, G. M., 1983 Dispersed repetitive DNA's in *Drosophila*, pp. 329–361 in *Mobile Genetic Elements*, edited by J. A. SHAPIRO. Academic Press, New York.

RUSSELL, S. R. H., 1989 Isolation of male-specific genes from *Drosophila melanogaster*. Ph.D. Thesis, Glasgow University.

SAMBROOK, J., E. F. FRITSCH and T. MANIATIS, 1989 *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

SCHAFER, M., R. KUHN, F. BOSSE and U. SCHAFER, 1990 A conserved element in the leader mediates post-meiotic translation as well as cytoplasmic polyadenylation of a *Drosophila* spermatocyte mRNA. EMBO J. **9:** 4519–4525.

STEIN, G. S., J. L. STEIN and W. F. MARZLUFF, 1984 *Histone Genes: Structure, Organisation and Regulation*. John Wiley & Sons, New York.

STEINMANN-ZWICKY, M., H. SCHMIDT and R. NOTHINGER, 1989 Cell autonomous and inductive signals can determine the sex of the germline of *Drosophila melanogaster* by regulating *Sex-lethal*. Cell **57:** 157–166.

SUBIRANA, J. A., 1983 Nuclear proteins in spermatozoa and their interaction with DNA, pp. 197–213 in *The Sperm Cell*, edited by J. ANDRE. Martinus Nijhoff, The Haag.

TRAVERSE, K. L., and M. L. PARDUE, 1989 Studies of He-T DNA sequences in the pericentric regions of *Drosophila* chromosomes. Chromosoma **97:** 261–271.

VAN DAAL, A., E. M. WHITE, M. A. GOROVSKY and S. C. R. ELGIN, 1988 *Drosophila* has a single copy of the gene encoding a highly conserved histone H2A variant of the H2A.F/Z type. Nucleic Acids Res. **16:** 7487–7497.

WELLS, D., and C. MCBRIDE, 1989 A comprehensive compilation and alignment of histones and histone genes. Nucleic Acids Res. **17** (Suppl.): r311.

WLASCHEK, M., A. AWGULEWITSCH and H. BUNEMANN, 1988 Structure and function of *Y* chromosomal DNA 1: sequence organisation and localisation of four families of repetitive DNA on the *Y* chromosome of *Drosophila hydei*. Chromosoma **96:** 145–158.

WU, R. S., H. T. PANUSZ, C. L. HATCH and W. M. BONNER, 1986 Histones and their modifications. CRC Crit. Rev. Biochem. **20:** 201–263.

YELICK, P. C., R. BALHORN, P. A. JOHNSON, M. CORZETT, J. A. MAZRIMAS, K. C. KLEENE and N. B. HECHT., 1987 Mouse protamine 2 is synthesised as a precursor whereas mouse protamine P1 is not. Mol. Cell. Biol. **7:** 2173–2179.

Communicating editor: M. T. FULLER