Genet. Res., Camb (1996), **68**, *pp.* 191–202 *With 7 text-figures*

A 1·5 kb direct repeat sequence flanks the *suppressor of forked* gene at the euchromatin–heterochromatin boundary of the *Drosophila melanogaster* X chromosome

MARK TUDOR[1], ANDREW MITCHELSON AND KEVIN O'HARE*
Department of Biochemistry, Imperial College of Science, Technology & Medicine, London SW7 2AZ, UK

(*Received 30 January 1996 and in revised form 17 May 1996*)

Summary

A 1·5 kilobasepair repeated DNA sequence is duplicated in direct orientation so as to flank the *suppressor of forked* gene in the euchromatin–heterochromatin transition region on the X chromosome of *Drosophila melanogaster*. These two copies are almost identical, but DNA blotting, analysis of cloned sequences and database searches show that elsewhere in the genome, homologous sequences are poorly conserved. They are often associated with other repeats, suggesting that they may belong to a scrambled and clustered middle repetitive DNA family. The sequences do not appear to be related to transposable elements and their location in different strains is conserved. *In situ* hybridization to metaphase chromosomes shows that homologous sequences are concentrated in the pericentric regions of the autosomes and the X chromosome. The sequences are not significantly under-represented in DNA from polytene tissue and must lie in the replicated regions of polytene chromosomes. The almost perfect conservation of the two repeats around *suppressor of forked* in *D. melanogaster* suggests they arose by duplication or gene conversion. Suppression of recombination in this chromosomal region presumably allows this unusual organization to be stably maintained. In the X-ray induced allele, *suppressor of forked-L26*, the sequence between the repeats, including the gene, and one copy of the repeat have been deleted.

1. Introduction

Heterochromatin is a major component of the genome of higher eukaryotes. Cytogenetic and molecular studies have revealed much about the structure, DNA sequence, content and location of heterochromatin in *Drosophila melanogaster* (for reviews see Gatti & Pimpinelli, 1992; Pardue & Henning, 1994). *In situ* hybridization of satellite DNA to mitotic chromosomes has shown that these highly repeated sequences are mainly located in the constitutive heterochromatin (for reviews see Verma, 1988; Miklos & Cotsell, 1990; Gatti & Pimpinelli, 1992). Recent studies using transposable element probes showed that they are often found in constitutive heterochromatin (Pimpinelli *et al.*, 1995; Carmena & Gonzalez, 1995). Blocks of satellite sequences may therefore be interspersed with middle repeated sequences, or even with more complex sequences (Le *et al.*, 1995). Genetic elements in heterochromatin are often very large, such as the Y chromosome fertility factors, or composed of tandem arrays, such as the *Responder* and *suppressor of Stellate* loci. The typical association of heterochromatin with the centromere and telomeres may indicate that its role is primarily structural.

In *Drosophila* larval salivary glands, the euchromatic portions of the chromosomes are replicated and remain associated to give the characteristically banded giant polytene chromosomes. The centromeric heterochromatin of the polytenized chromosomes associates to form a chromocentre. This heterochromatin can be divided into two cytologically distinct types: α-heterochromatin, which forms the compact central region of the chromocentre, and β-heterochromatin corresponding to the morphologically diffuse network around it (Heitz, 1934). The relationship between these and the constitutive heterochromatin seen in mitotic chromosomes has been the subject of much discussion (for review see Miklos & Cotsell, 1990). Satellite DNA is under-represented in polytene chromosomes, presumably because of under-

* Corresponding author. Fax: +44 (171) 225 0960. e-mail: kohare@ic.ac.uk.
[1] Current address: Dipartimento di Genetica e Biologia Moleculare, Città Universitaria di Roma, Piazzale Aldo Moro 5, Rome I-00185, Italy.

replication or replication followed by elimination, and is present in the compact chromocentre (Gall, 1973; Lamb & Laird, 1987; Spradling & Orr-Weaver, 1987; Spradling *et al.*, 1993). *In situ* hybridization of transposable elements to polytene chromosomes typically results in hybridization to the chromocentre and telomeres, as well as to dispersed sites in the chromosome arms (Young, 1979).

The best-characterized region of β-heterochromatin in *Drosophila melanogaster* is at the base of the X chromosome, corresponding to cytological division 20 (Schalet & Lefevre, 1976). In polytene chromosomes, this region is morphologically indistinct with poorly formed bands, and is often separated from the chromocentre by a constriction. Mutagenesis studies have shown that the density of genes in division 20 is similar to that in the adjacent euchromatic divisions. Analysis of sequences cloned from divisions 19 and 20 by micro-dissection has shown that this region is rich in middle repeats, many of which are homologous to transposable elements (Miklos *et al.*, 1988). Analysis of 60 kb around the *unco-ordinated* locus in 19E on the X chromosome showed that unique sequences were interspersed with repeats (Miklos *et al.*, 1984; Healy *et al.*, 1988). The repeated sequences were organized as tandem arrays and contained sequences homologous to type I insertion sequences found in ribosomal DNA. Around the locus *lethal(1)B214* in 19F, 40% of the sequences were repeated, and included regions with homology to tRNA genes and the *copia* transposable element (Russell *et al.*, 1992). Similar results have been found for autosomal heterochromatic genes such as *light* (Devlin *et al.*, 1990) and *rolled* (L. Berghella & P. Dimitri, pers. comm.).

The *suppressor of forked* locus, *su(f)*, is the most proximal single copy gene on the X chromosome. It is in the β-heterochromatic region 20EF, and is presumably close to the mitotic heterochromatin. We have reported the cloning of the gene by P element transposon tagging (Mitchelson *et al.*, 1993). In polytene chromosomes, *su(f)* hybridizes to the most proximal part of division 20 and is not significantly under-represented (Yamamoto *et al.*, 1990). In mitotic chromosomes, *su(f)* hybridizes to the euchromatin–heterochromatin boundary (C. Gonzalez, M. Tudor & K. O'Hare, unpublished results). Analysis of the cloned interval around *su(f)* has shown that in the distal direction unique sequences are interspersed with repeated sequences, while all of the 30 kb isolated from the proximal side of the gene are repeated (M. Tudor & K. O'Hare, unpublished observations).

We are interested in the nature and organization of the repeats around *su(f)* and other heterochromatic genes, and in the possibility that they have a role in the expression of these genes. We report here that *su(f)* is closely flanked by 1·5 kilobasepair (kb) repeated sequences in the same orientation. These sequences do not appear to be mobile. Homologous sequences elsewhere in the genome are not closely related in overall organization to this 1·5 kb repeat and are often arranged in scrambled clusters. Database searches have revealed that homologous sequences are often in the intervals between genes. We also describe the cloning of the allele $su(f)^{L26}$, which was generated by X-irradiation. The gene and one copy of the 1·5 kb repeat have been deleted resulting in a structure that resembles a recombination between the two copies of the repeat.

2. Materials and methods

(i) Drosophila *procedures*

The wild-type *Drosophila melanogaster* strains Canton S and Oregon R were from the Bowling Green stock centre. The wild-type strain Hikone J was from M. G. Kidwell (University of Arizona, Tucson). The wild-type strain Amherst M56i and the mutant strain $su(f)^{L26}$ (renamed $su(f)^{19}$, in Lindsley & Zimm, 1992) were from A. Schalet (Yale University, Connecticut). All flies were kept on standard cornmeal/yeast/agar medium at 25 °C. Diploid (brain and imaginal disks) and polytene (salivary gland) tissues were hand-dissected from third instar larvae of the strain Canton S and DNA isolated as described by Yamamoto *et al.* (1990).

(ii) *Recombinant DNA procedures*

DNA preparation from adult flies, DNA blotting, screening of libraries, subcloning and DNA sequencing using the dideoxy chain termination technique on M13 subclones were by standard procedures. The hybridization and washing conditions were of moderate stringency, and detected a match of 87% over 116 bases, but not one of 72% over 141 bases (see text). The cloning of *su(f)* from *D. melanogaster* and *D. simulans* is described by Mitchelson *et al.* (1993) and Langley *et al.* (1993), respectively. The DNA sequences of the 1·5 kb repeats are included in our sequence for *su(f)* from *D. melanogaster*, accession number X62679. The DNA sequences for the intervals flanking *su(f)* in *D. simulans* have been given the accession numbers L16771 and L16772. The 1·5 kb *Xho* I–*Hind* III fragment containing the proximal repeat was subcloned between the *Sal* I and *Hind* III sites of the vector pBluescribe M13 minus to make the plasmid pRP1·5 (Fig. 1). It was also cloned as two fragments into M13 vectors: X5L is a clone of the 550 basepair (bp) *Xba* I fragment; and XBH is a clone of the 800 bp *Xba* I–*Hind* III fragment. A library of *Sal* I fragments was made in the vector λ EMBL4 using DNA from heterozygous $su(f)^{L26}/su(f)^{+}$ females. To distinguish phage containing the mutant-specific fragment from phage with inserts of wild-type fragments, they were first screened with a probe from −5·1 to −6·6 and

https://doi.org/10.1017/S0016672300034169 Published online by Cambridge University Press

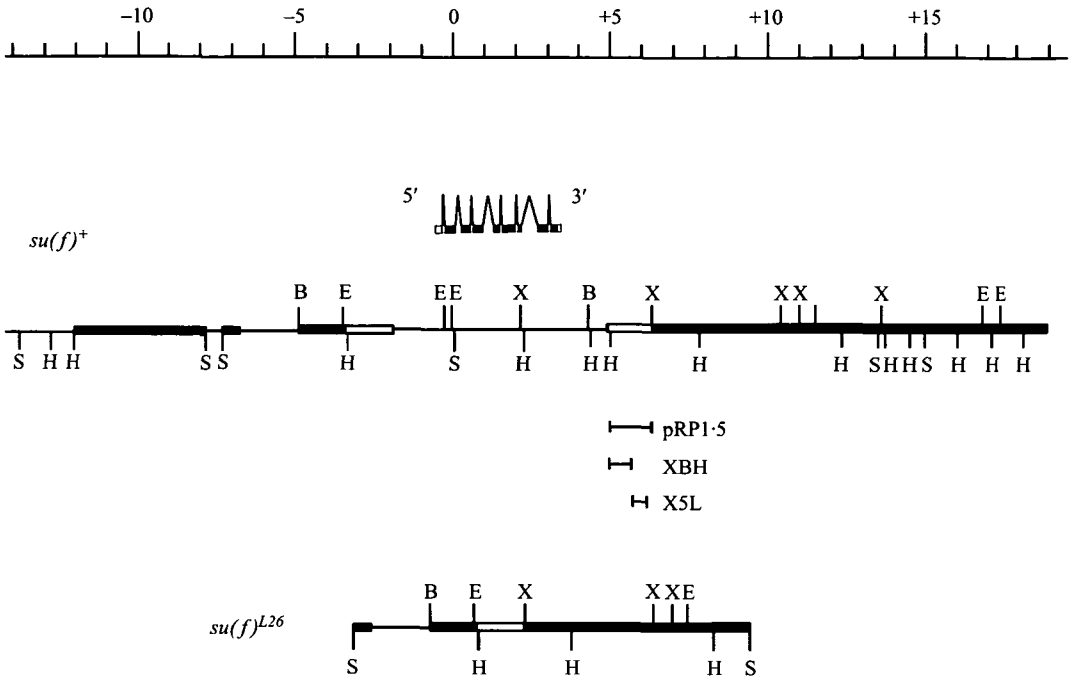

Fig. 1. Map of the *su(f)* region. The co-ordinate system (in kb) uses the *Sa* I site at the 5′ end of the gene as the origin. The proximal direction (with respect to the centromere) is to the right. Unique intervals are shown as thin lines, intervals containing repeated DNA sequences as filled boxes and the 1·5 kb repeats as open boxes. Subclones of the proximal copy of the 1·5 kb repeat are shown. The map of $su(f)^{L26}$ is shown below. X, *Xho* I; E, *Eco*R I; B, *Bam*H I; H, *Hin*d III; S, *Sal* I.

then counterscreened with a probe from 0·0 to −2·1 (see Fig. 1). In this way phage containing the *Sal* I fragment from the wild-type chromosome (0·0 to −7·5) were eliminated, leaving a single candidate phage for $su(f)^{L26}$ which was purified and analysed.

(iii) In situ *hybridization to metaphase chromosomes*

Biotin-labelled pRP1·5 was hybridized to metaphase chromosomes prepared from Oregon R third instar larval brains as described in Abad *et al.* (1992). Signal detection was by immunofluorescence and photographs were taken using a BioRad MRC-600 confocal microscope.

(iv) *Analysis of DNA sequences*

DNA sequences were assembled from gel readings using the Microgenie package from Beckman. Sequences were compared with databases using BLAST and FASTA. Diagonal comparisons of *su(f)* from *D. melanogaster* and *D. simulans* were made using the MacVector package from IBI.

3. Results

(i) *The* su(f) *gene is flanked by 1·5 kb direct repeats in* Drosophila melanogaster

Analysis of a 33 kb chromosomal walk around the *su(f)* locus revealed that two 1·5 kb repeated sequence intervals from either side of the gene had similar restriction enzyme maps and cross-hybridized (Fig. 1). These repeats are in the same orientation and flank the 7 kb unique region in which the *su(f)* gene lies (Mitchelson *et al.*, 1993). Although most of this interval consists of repetitive sequences, these 1·5 kb repeats show no homology with the rest of the cloned region. The restriction enzyme map of the *su(f)* region, including these repeats, is conserved between wild-type strains from different geographic origins (Langley *et al.*, 1993; our unpublished results), indicating that the repeats are not polymorphic. Furthermore, the restriction enzyme map of the 1·5 kb repeat shows no obvious similarity to those of known transposable elements (Lindsley & Zimm, 1992).

The sequences of the two repeats were determined and are shown in Fig. 2. As judged by divergence between the sequences, the repeat is 1437 bases long. They are almost identical with the following differences in the proximal copy compared with the distal copy: (i) substitution of G for C at position 8; (ii) a deletion of TAG at position 346–348; (iii) deletion of one T from a run of Ts at position 615–624; (iv) substitution of CTT for ATC at position 694–696; (v) substitution of C for T at position 1404. The ends of the repeats show none of the features characteristic of transposable elements. There are no flanking duplications, nor any direct or inverse terminal repeats. The overall AT content is 64%, which is similar to that of the *D. melanogaster* genome. There are no long open reading frames on either strand.

https://doi.org/10.1017/S0016672300034169 Published online by Cambridge University Press

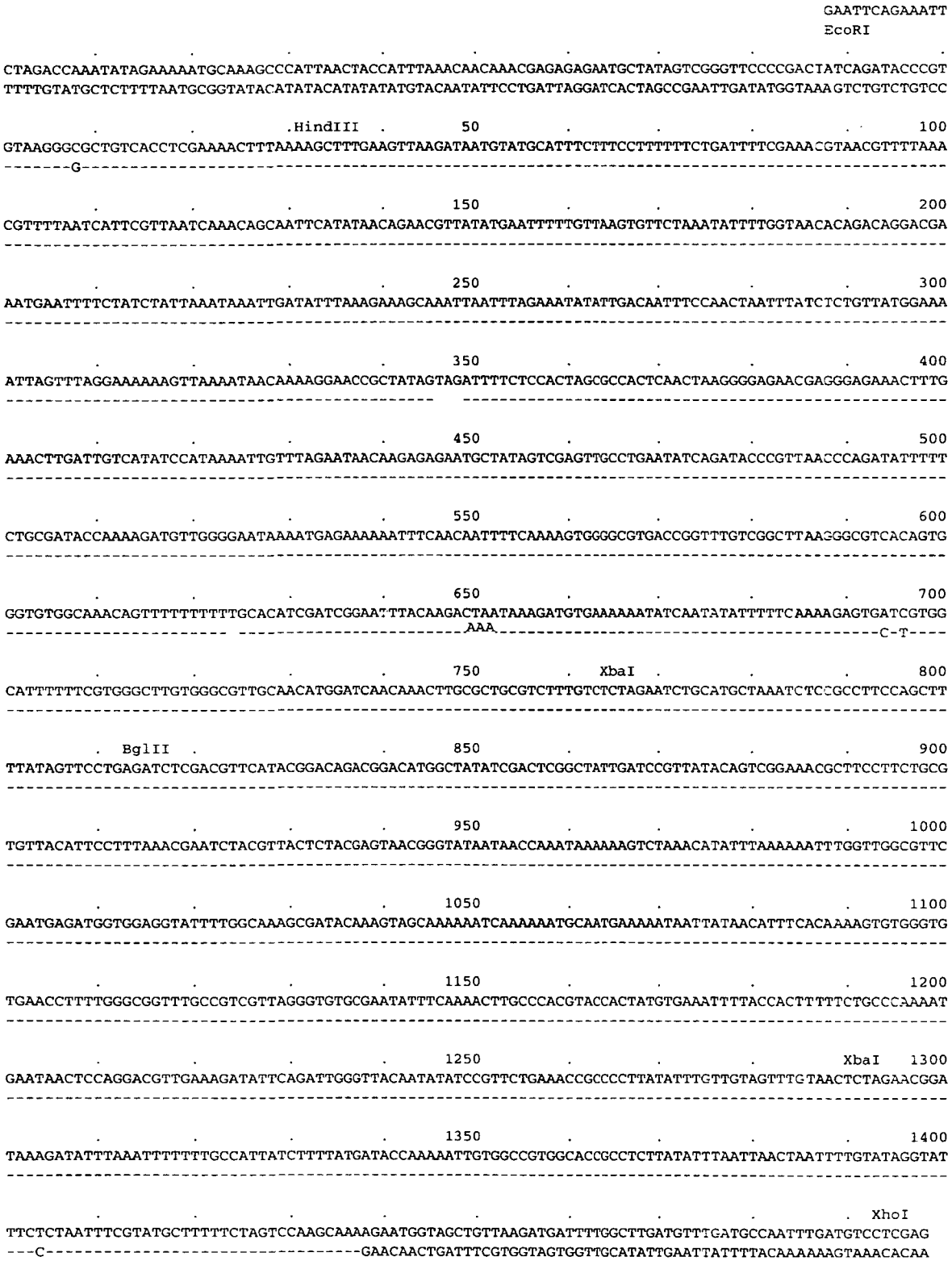

Fig. 2. DNA sequence of the 1·5 kb repeats. The upper line shows the complete DNA sequence of the proximal 1·5 kb repeat. The lower line shows the DNA sequence of the distal copy, where identical nucleotides are shown as (–), base substitutions are shown and deletions are gaps.

(ii) *Sequences homologous to the 1·5 kb repeat are clustered with other repeats*

DNA blotting experiments using the 1·5 kb repeat as probe (plasmid pRP1·5) have shown that homologous sequences are present elsewhere in the genome, and have generated information on their organization and conservation. The most strongly hybridizing bands do correspond to those from the 1·5 kb repeats from *su(f)* (see below, Figs 4 and 6). On the basis of the signal strength and other data from cloned sequences (see below), we believe that this precise DNA sequence

https://doi.org/10.1017/S0016672300034169 Published online by Cambridge University Press

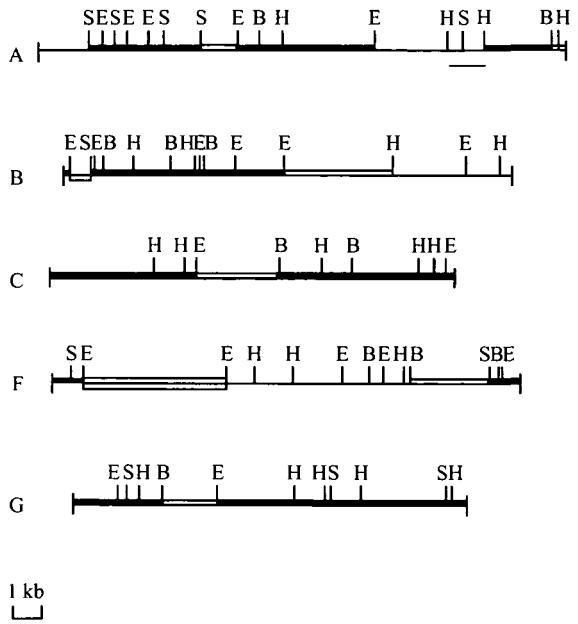

Fig. 3. Restriction map of cloned regions with homology to the 1·5 kb repeat. Low copy number intervals are shown as lines. Intervals homologous to X5L are open boxes, intervals homologous to XBH are grey boxes, intervals with other repeats are filled boxes. E, *Eco*R I; B, *Bam*H I; H, *Hin*d III; S, *Sal* I. The fragment from A used as a probe in Fig. 6 is indicated.

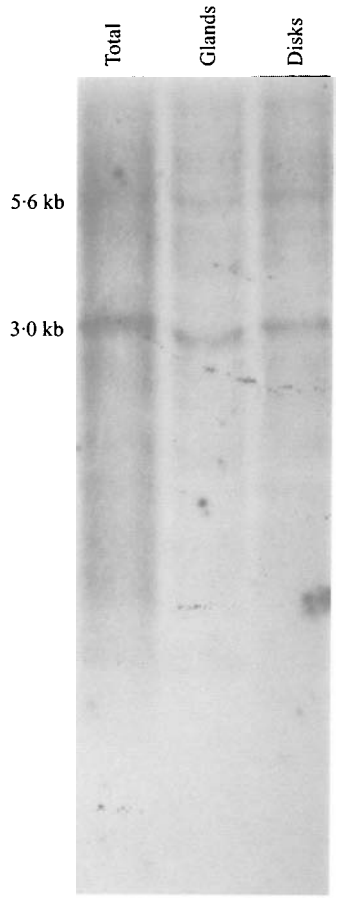

Fig. 4. Analysis of DNA from diploid and polytene tissue. DNA from diploid and polytene tissue was digested with *Hin*d III and probed with pRP1·5. The position of bands expected from *su(f)* are indicated. The loading was controlled by probing with the *white* locus and the efficacy of the dissection was verified by probing with a type I rDNA repeat and a 359 bp satellite DNA (not shown).

is present in low copy number, possibly only the two copies at *su(f)*, and that many homologous but less well conserved sequences exist elsewhere in the genome. Homologous sequences were also detected in other members of the *melanogaster* subgroup by DNA blotting (data not shown).

Part of the 1·5 kb repeat, X5L (Fig. 1), was used to screen a lambda library and a gridded array of cosmids (Hoheisel *et al.*, 1991). Under the conditions used about 3% of the recombinants hybridized, but with considerable variation in signal strength. This suggested that there were many related sequences in the genome, but that there was significant variation in the extent and/or degree of homology.

Five phages which showed a range of signal strength when hybridized with X5L were analysed. The inserts were mapped and the homologous intervals defined (Fig. 3). None of the clones had a map similar to the *su(f)* region in general or the 1·5 kb repeat in particular. The insert in phage A contained two regions which hybridized to X5L, neither of which was homologous to the remainder of the 1·5 kb repeat present in clone XBH (Fig. 1). In phage B, the insert had one interval that hybridized to both X5L and XBH and a second interval which hybridized to XBH only. One region in phage F hybridized to X5L alone and a second region hybridized to both X5L and XBH. Phage C and G each contained a single region with homology to X5L. This shows that the homologous sequences are often clustered, but poorly conserved in organization compared with the sequences flanking *su(f)*. Regions with homology to one part of the 1·5 kb repeat (X5L) are not always found associated with the rest of the repeat (XBH). It seems likely that a similar result would have been found if XBH had been used as the initial probe to select the lambda clones. Labelled genomic DNA was used to probe DNA blots of these clones in order to define the regions within the cloned intervals which are sufficiently repeated in the genome to give a signal (a 'reverse Southern blot'). This showed that the entire inserts in C and G are repeated. In A, B and F there are other repeated regions, in addition to those which hybridize to the 1·5 kb repeat, but also some low copy number regions which gave no or very little signal. This co-localization with other repeated sequences suggests that these inserts may be derived from heterochromatic regions of the genome.

The DNA sequence of the region from phage G homologous to X5L (Fig. 3) was determined (accession number U42215) and found to match for 65% over 348 bases including a core of 87% over 116 bases. Although no homology was detected between G and XBH in DNA blotting experiments, there was

https://doi.org/10.1017/S0016672300034169 Published online by Cambridge University Press

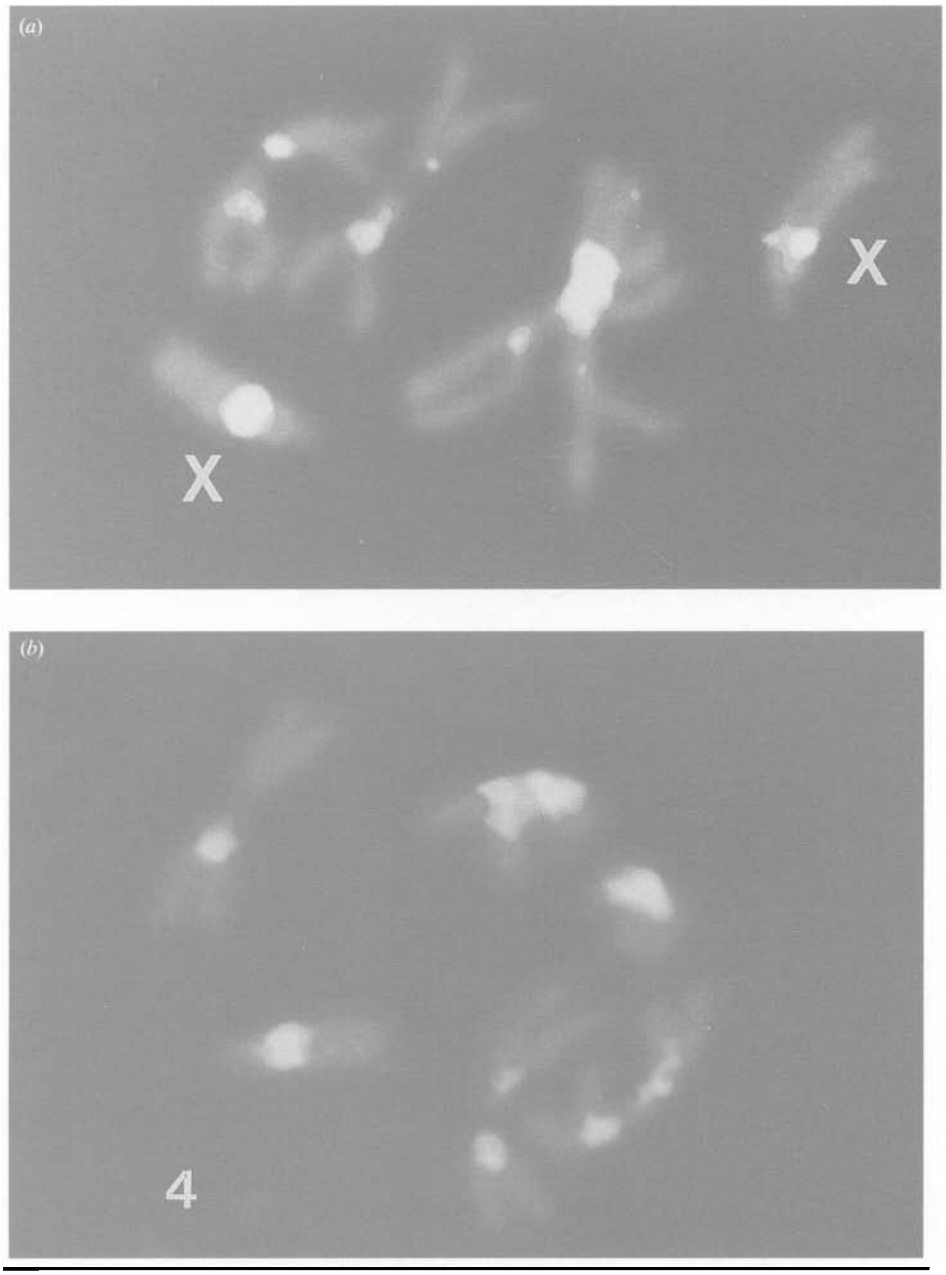

Fig. 5. *In situ* hybridization of 1·5 kb repeat to metaphase chromosomes. Metaphase chromosomes were dissected from Oregon R third instar flies and probed with biotinylated pRP1·5. The hybridization appears to mark out β-heterochromatin. In (*a*), strong hybridization can be seen at the euchromatin–heterochromatin junction of the X chromosome. The autosomes cannot be identified with confidence, but there is clearly good hybridization to the pericentric β-heterochromatic regions on both arms of one of the autosomes, with the other autosome showing strong hybridization to one arm and weak hybridization to the other arm. This is consistent with the distribution of β-heterochromatin and suggests that the chromosome showing asymmetric hybridization is the third chromosome. The absence of hybridization to the fourth chromosome is evident in (*b*).

significant homology at the DNA sequence level (59% over 345 bases including a core of 72% over 141 bases). This gives a measure of the stringency of our hybridization experiments.

(iii) *Sequences homologous to the 1·5 kb repeat are not under-represented in DNA from polytene tissue*

The representation in polytene DNA of sequences homologous to the 1·5 kb repeat was examined by probing DNA from diploid and polytene tissues (Fig. 4). The loading control for this blot can be seen in figure 3 of Yamamoto *et al.* (1990), where a probe from euchromatic gene (the *white* locus) was used. Separation of tissues was verified by probing with two heterochromatic sequences: the type I rDNA insertion sequence (Glover, 1981) and a 359 bp satellite (Peacock *et al.*, 1978) (data not shown). The two major bands of 5·6 and 3·0 kb in Fig. 4 correspond to the *Hind* III fragments from the repeats flanking *su(f)* (Fig. 1). The fainter bands are from homologous sequences elsewhere in the genome. A comparison of

https://doi.org/10.1017/S0016672300034169 Published online by Cambridge University Press

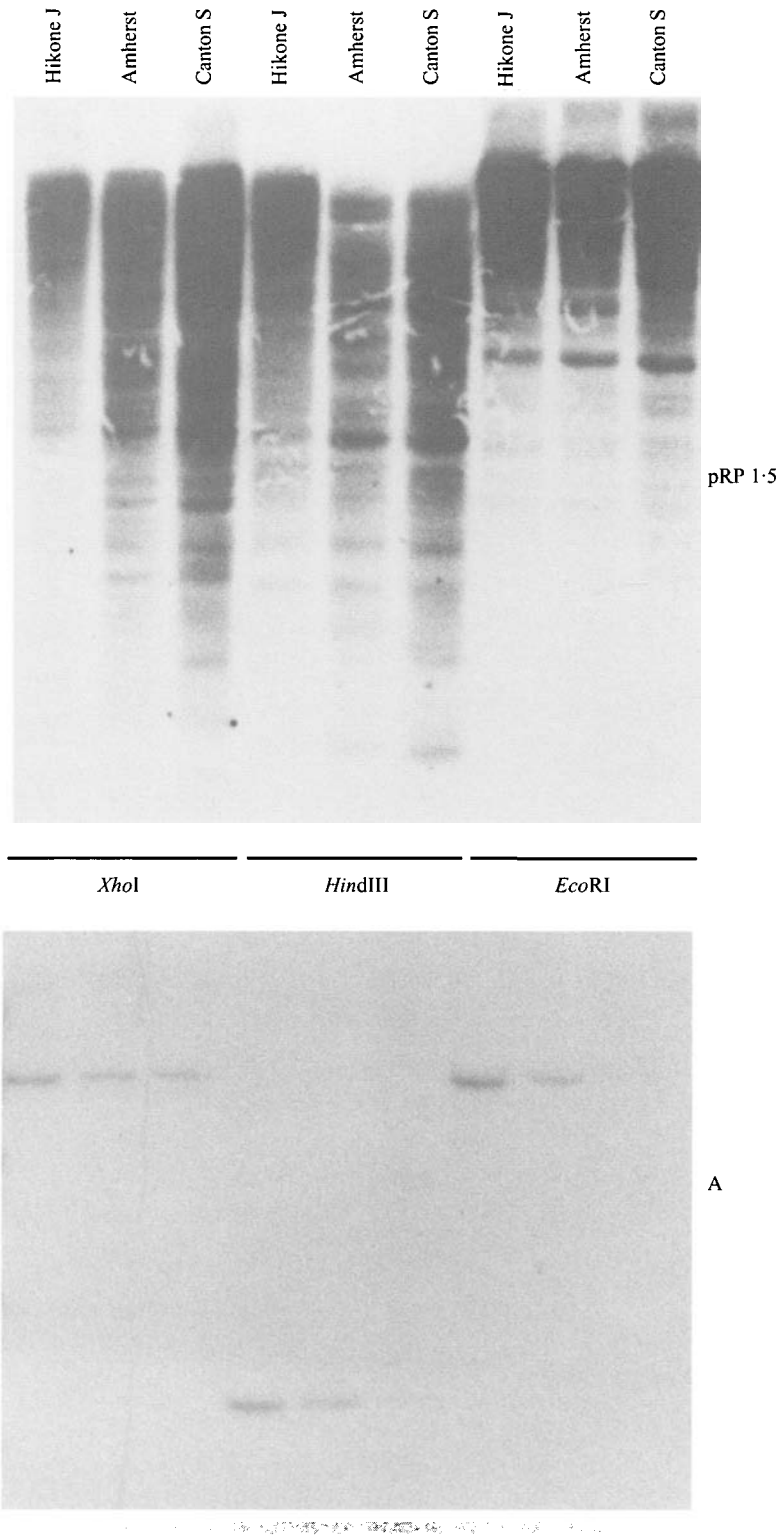

Fig. 6. Sequences homologous to the 1·5 kb repeats are not mobile. DNA from Canton S (Ohio), Amherst M56i (Massachusetts) and Hikone J (Japan) was digested with *Eco*R I, *Hin*d III and *Xho* I and probed with pRP1·5. The blot was stripped and reprobed with a low copy number 1·5 kb *Hin*d III fragment from phage A (see Fig. 3).

the signal strength suggests that none of these sequences is obviously under-represented in polytene chromosomes. If they are heterochromatic, then they are likely to be in β- rather than α-heterochromatin.

The chromosomal location of sequences homologous to the 1·5 kb repeat was determined by *in situ* hybridization to metaphase chromosomes from Oregon R third instar larval brains. Strong hybridization can be seen where *su(f)* maps, at the euchromatin–heterochromatin boundary on the X chromosome (Fig. 5*a*). The extent of the hybridization is greater than that seen when single-copy sequences from around *su(f)* are used (C. Gonzalez, pers. comm.), indicating that this signal is not simply

https://doi.org/10.1017/S0016672300034169 Published online by Cambridge University Press

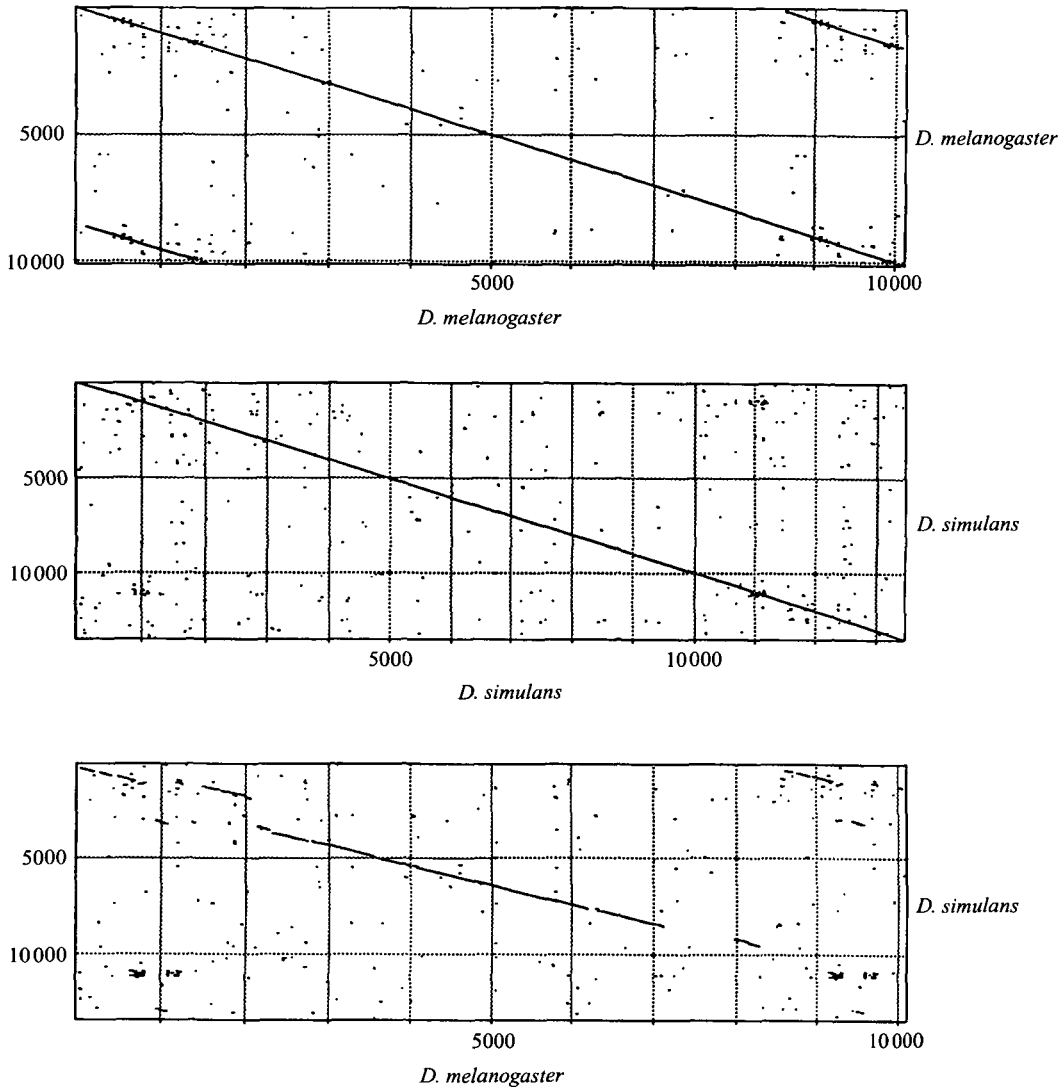

Fig. 7. Tandem repeats around *su(f)* in *D. melanogaster* but not in *D. simulans*. Diagonal plots of comparisons of *D. melanogaster* with itself (top), *D. simulans* with itself (middle) and *D. melanogaster* with *D. simulans* (bottom). A match of 65% over 30 bases scored. The location of the *su(f)* transcription unit in *D. melanogaster* is 3150 to 7350. This region is well conserved in *D. simulans* and corresponds to approximately 4450 to 8750. The 1·5 kb repeats correspond to 114 to 1546 and 8636 to 10072 in *D. melanogaster*.

arising from the two flanking repeats. Less strong hybridization is seen at similar positions on the autosomes. Hybridization is seen to both arms of one of the autosomes, and mainly to one arm of the other autosome. This is consistent with the distribution of β-heterochromatin on the autosome arms, where there is a similar amount on both arms of chromosome 2 but less on 3R than on 3L (Miklos & Cotsell, 1990). No hybridization was detected to the fourth chromosome (Fig. 5*b*).

(iv) *Sequences homologous to the 1·5 kb repeats are not mobile*

To investigate whether the repeats were polymorphic, three wild-type strains of different geographic origin (Canton S from Ohio, Amherst M56i from Massachusetts and Hikone J from Japan) were analysed. When *Eco*R I, *Hind* III and *Xho* I digests of DNA of these strains were probed with the 1·5 kb repeat, very similar patterns were detected (Fig. 6). Low copy number fragments purified from clones A, B and F described above were also used as probes. The results (Fig. 6; data not shown for B and F) show that the DNA maps of the three regions corresponding to these probes are the same. This indicates that sequences homologous to the 1·5 kb repeat are in the same chromosomal positions (as defined by the flanking restriction enzyme sites) in these three strains. Sequences homologous to the 1·5 kb repeat are therefore not likely to be members of an active family of transposable elements.

(v) *Sequences homologous to the 1·5 kb repeats around* su(f) *in* D. simulans

We used DNA blotting to examine the *su(f)* gene cloned from *D. simulans* by Langley *et al.* (1993) for

https://doi.org/10.1017/S0016672300034169 Published online by Cambridge University Press

Table 1. *Homologies with 1·5 kb repeat*

Accession no.	Strand	Sequence	Score[a]	Match[b]	Position in 1·5 kb repeat
L16772	+	*D. simulans su(f)* flanking sequence (distal)	1253	538/760	1–725
X53541	–	*D. melanogaster* U1-82·1 snRNA gene	777	256/324	552–873
X53542	–	*D. melanogaster* U1-82·3 snRNA gene	666	264/363	478–828
V00228	+	*D. melanogaster* tRNA gene cluster	510	171/217	490–705
U42215	+	*D. melanogaster* phage X5L-G	484	424/687	434–1086
L42553	–	*D. melanogaster msl-2* gene	473	225/333	472–802
L11345	+	*D. melanogaster erect wing* gene	357	120/154	711–864
U04853	+	*D. melanogaster* trypsin gene cluster	350	174/264	748–1002
L16771	+	*D. simulans su(f)* flanking sequence (proximal)	342	150/214	522–724
L16771	–	*D. simulans su(f)* flanking sequence (proximal)	309	118/163	784–942
L39676	–	*D. melanogaster* D21 subclone (ADH region)	301	167/276	596–861
X86015	–	*D. melanogaster* H^+-ATPase beta subunit	273	172/278	559–826
J01068	–	*D. melanogaster* ARS sequence	264	148/225	709–925
X16802	–	*D. melanogaster Ddc* gene 5′ enhancer	250	98/130	513–639

[a] The optimized score from FASTA.
[b] Expressed as number of bases identical over length of match.

homology to the 1·5 kb repeat. The region analysed included the gene and about 9 kb of flanking sequences on each side. We presumed that the orientation of the gene with respect to the centromere is the same in *D. simulans* as it is in *D. melanogaster*. A region with significant homology to the 1·5 kb repeat was detected on the distal side of the gene, but there was less homology on the proximal side. The regions corresponding to these intervals were subcloned and sequenced. These sequences overlap and extend that determined by Langley *et al.* (1993) for *su(f)* from *D. simulans* to make a 13·5 kb contiguous sequence. From the alignment with the *D. melanogaster* sequence, the *su(f)* transcription unit in *D. simulans* corresponds approximately to nucleotides 4450 to 8750 of this 13·5 kb sequence. In Fig. 7 we have compared the two sequences to illustrate the location of the 1·5 kb repeats in *D. melanogaster*, the absence of flanking repeats in the sequenced interval from *D. simulans*, and the homologies between the sequences flanking *su(f)* in *D. simulans* and the 1·5 kb repeat. The match between the distal 1·5 kb repeat from *D. melanogaster* and the corresponding region from *D. simulans* is 71% over 750 bp (Table 1). The best match with sequences on the proximal side of the gene in *D. simulans* was only 70% over 214 bp. There was a second match in this region of 72% over 163 bp between a different part of the 1·5 kb repeat and the other strand of the *D. simulans* sequence.

(vi) *Sequences homologous to the 1·5 kb repeat in databases*

The sequence of the 1·5 kb repeat was compared with the EMBL and Genbank databases using FASTA and BLAST. The very best matches were with the sequences flanking *su(f)* from *D. melanogaster* and *D. simulans*. However, the best 40 matches found using BLAST were all with other *Drosophila* sequences, usually with multiple intervals of homology. Data from the output of the FASTA search are shown in Table 1, which also includes the match with the sequence from phage G (see above). Many of the matches were with sequences which have homology to the Lefka repeat defined from comparison of sequences from *D. melanogaster* sequence tagged sites (Madueno *et al.*, 1995). The interval from position 951 to 990 in the 1·5 kb repeat is 87·5% identical to Lefka. However, the matches between the 1·5 kb repeat and the database entries shown in Table 1 are more extensive than those between the 1·5 kb repeat and Lefka, and include regions of the 1·5 kb repeat which are not homologous to Lefka.

(vii) *The structure of* $su(f)^{L26}$ *resembles that of a recombination event between the 1·5 kb repeats*

As part of our characterization of *su(f)* we have analysed mutant alleles by DNA blotting (Mitchelson *et al.*, 1993). In some digests of DNA from heterozygous $su(f)^{L26}/su(f)^{+}$, bands from within the *su(f)* transcription unit were submolar whilst bands from outside this region were not. This information, together with several DNA fragments specific for the mutant allele, allowed us to deduce a restriction enzyme map for $su(f)^{L26}$ (Fig. 1). This map suggested that the gene and one copy of the 1·5 kb repeat had been deleted. To confirm this, we cloned a mutant-specific *Sal* I fragment from $su(f)^{L26}$. The restriction enzyme map of this confirmed that sequences between the two 1·5 kb repeats and one copy of the 1·5 kb repeat had been deleted. As the genetics of *su(f)* is complex, with complementation between lethal alleles, we routinely use $su(f)^{L26}$ as our reference null allele in genetic experiments (Simonelig *et al.*, 1996).

The 1·6 kb *Eco*R I–*Xho* I fragment containing the site of the deletion was subcloned into M13 vectors and a partial sequence determined. The single copy of

https://doi.org/10.1017/S0016672300034169 Published online by Cambridge University Press

the 1·5 kb sequence present in $su(f)^{L26}$ appears to be identical to the distal copy in $su(f)^{+}$. It has the characteristic polymorphisms of the distal repeat at positions 8, 346–348 and 1404 (see Fig. 2). The sequences for the two polymorphic positions in the middle of the repeat were not determined. The sequences flanking the distal side of the single copy of the 1·5 kb repeat in $su(f)^{L26}$ exactly match those on the distal side of the distal copy in $su(f)^{+}$, and on the proximal side they exactly match those from the proximal side of the proximal copy in $su(f)^{+}$. Although this structure resembles that of a recombination event between the two flanking repeats, recombination is very rare at the base of the X chromosome. As $su(f)^{L26}$ was generated using X-rays (Lefevre, 1981), perhaps a chromosome break was generated on the proximal side of the distal copy of the repeat, and this break was then healed by repair against the proximal copy. A similar deletion of one repeat and the intervening unique sequence has recently been recovered following mobilization of a P element insertion in *su(f)* (Williams & O'Hare, 1996). This is distinguishable from $su(f)^{L26}$ only in that the remaining 1·5 kb repeat has polymorphisms from both the distal repeat (at position 8) and the proximal repeats (at all the other positions).

4. Discussion

(i) *1·5 kb repeats flank the* D. melanogaster su(f) *gene*

We have found that the *su(f)* gene at the euchromatin–heterochromatin boundary on the X chromosome of *Drosophila melanogaster* is flanked by almost identical copies of a 1·5 kb sequence. There are many homologous sequences elsewhere in the genome, but probably no other precise copies. The homologous sequences are often clustered together with other repeated sequences. DNA blotting suggests that the most homologous of the related sequences are not under-represented in polytene DNA, while *in situ* hybridization to mitotic chromosomes shows that they are concentrated in β-heterochromatic regions.

We have shown by *in situ* hybridization to polytene chromosomes that *su(f)* is located at the base of the X chromosome in other species of the *melanogaster* group of *Drosophila* (data not shown). We have analysed the *su(f)* locus cloned from *D. simulans* and have found a sequence similar to the 1·5 kb repeat on the distal side of the gene, although the sequences on the proximal side are less homologous. Other repeats around *su(f)* in *D. melanogaster* are not present in the interval cloned from *D. simulans* (data not shown). In species of the *obscura* group of *Drosophila*, *su(f)* has a euchromatic location (Segarra & Aguade, 1992). It would be interesting to investigate the organization of the *su(f)* gene cloned from a member of the *obscura* group.

(ii) *Are the 1·5 kb repeats transposable elements?*

There is no evidence that the 1·5 kb sequence itself can transpose. It has none of the sequence features commonly found in transposable elements and appears to be non-coding DNA. It matches neither the maps nor the sequences of known *D. melanogaster* transposable elements. Both copies of the 1·5 kb repeat are present at the *su(f)* locus in wild-type strains from different geographic origin, so they are not polymorphic. However, this region may have undergone some form of genetic 'hitch-hiking' (Langley *et al.*, 1993), so the lack of polymorphism for these repeats might simply reflect the location of the copies in the chromosomal interval swept through the population.

All this indicates that the 1·5 kb repeats themselves do not transpose, but does not exclude them being related to active transposable elements. DNA blotting suggests that the poorly conserved, but related, sequences elsewhere in the genome are not mobile. If they were members of a previously active transposable element family whose functional members had been eliminated by mutation, or through recombination and segregation, then better conservation might have been expected for the residual homologous sequences. This is the case for the non-mobile elements related to the I factor that are found in β-heterochromatic regions of R strains (Vaury *et al.*, 1990). However, the degree of relatedness between copies of previously active transposable elements would clearly depend upon the evolutionary history of that particular transposable element family.

(iii) *Evolution of the flanking repeats*

The near identity of the repeats at *su(f)* suggests that the repeats arose by duplication or gene conversion. Duplication by transposition seems unlikely, so perhaps gene conversion is responsible for this sequence arrangement. Comparison of *D. melanogaster* with *D. simulans* suggests that a distal sequence may have been duplicated into the proximal side of the gene in *D. melanogaster* after divergence between progenitors of *D. melanogaster* and *D. simulans*. If duplication had occurred earlier, there is no obvious reason why the two copies would have been maintained so faithfully in the lineages leading to *D. melanogaster* but not in *D. simulans*.

The regions flanking euchromatic genes in *D. melanogaster* are, with the major exception of transposable elements, generally single copy sequence. Recombination between repeats flanking euchromatic genes would lead to chromosome rearrangements, and recombination between copies of transposable elements in different chromosomal positions has been found in some rearrangements (for review see Lim & Simmons, 1994). However, recombination is suppressed at the base of the X chromosome, so

https://doi.org/10.1017/S0016672300034169 Published online by Cambridge University Press

recombination between the repeats flanking *su(f)* would not be expected. Indeed, suppression of recombination and organization of sequences as tandem arrays, from rDNA to satellite DNA, are features of heterochromatin. We have found, and are characterizing, a tandem array of repeats proximal to *su(f)* in *D. melanogaster* (M. Tudor & K. O'Hare, unpublished results).

(iv) *Is there a role for these sequences?*

Examination of the precise location within the database entries where matches to the 1·5 kb repeat are found shows that they are often in intergenic regions. These non-coding sequences may not be subject to much selective pressure, and are usually more AT-rich than genes in *D. melanogaster*. However, the 1·5 kb repeat is not more AT-rich than average, and the homologies are not simply between AT-rich regions in the 1·5 kb repeat and other AT-rich sequences. If these regions are usually not transcribed, they may have evolved their particular sequence similarities for reasons to do with chromatin structure. They could simply be sequences where nucleosomes are stably maintained.

A more interesting possibility is that these sequences have a role in marking out regions to be transcribed during the organization of chromatin structures. A number of chromatin domain boundary sequences have been characterized in *D. melanogaster*. These include scaffold attachment sites (Gasser & Laemmli, 1986), specialized chromatin structures (Kellum & Schedl, 1991, 1992), and sequences which isolate *cis*-acting regulatory elements (Roseman *et al.*, 1993; Galloni *et al.*, 1993). Although all are AT-rich, they show no extensive DNA sequence homology; nor is there much homology between these sequences and the 1·5 kb repeat. From analysis of position effect variegation where euchromatic and heterochromatic sequences are brought together by chromosomal rearrangement, it has been suggested that there are sequences which act to stop spreading of the inactive chromatin structure found in heterochromatin into the neighbouring euchromatin (Tartof *et al.*, 1984). The repeats flanking *su(f)* are ideally positioned if their function was to limit the possibly inhibitory effect of neighbouring heterochromatin upon *su(f)* expression. The detection of sequences homologous to this repeat at the euchromatin–heterochromatin boundary of the X chromosome and the autosomes by *in situ* hybridization is also consistent with this role. The *su(f)* gene has been re-introduced into *D. melanogaster* without its flanking repeats (Mitchelson *et al.*, 1993) and it functions normally at these presumably euchromatic ectopic sites. However, it remains possible that the repeats do have a function in the normal chromosomal location of *su(f)*.

M.T. was supported by a studentship from the Science & Engineering Research Council and A.M. by a studentship from the Medical Research Council. We would like to thank A. Schalet and M. Kidwell for flies, J. Hoheisel for filters bearing gridded cosmid clones, C. Gonzalez and D. Glover for hospitality and help with the *in situ* hybridizations to metaphase chromosomes, F. Karch for DNA sequences of domain boundaries from the Bithorax Complex and C. Langley for the *su(f)* clones from *D. simulans* and for discussions on comparisons with the gene from *D. melanogaster*.

References

Abad, J., Carmena, M., Baars, S., Saunders, R., Glover, D., Ludena, P., Sentis, C., Tyler-Smith, C. & Villasante, A. (1992). Dodeca satellite: a conserved G+C-rich satellite from the centromeric heterochromatin of *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences of the USA* **89**, 4663–4667.

Carmena, M. & Gonzalez, C. (1995). Transposable elements map in a conserved pattern of distribution extending from beta-heterochromatin to centromeres in *Drosophila melanogaster*. *Chromosoma* **103**, 676–684.

Devlin, R. H., Bingham, B. & Wakimoto, B. T. (1990). The organization and expression of the *light* gene, a heterochromatic gene of *Drosophila melanogaster*. *Genetics* **125**, 129–140.

Gall, J. G. (1973). Repetitive DNA in *Drosophila*. In *Molecular Cytogenetics* (ed. B. A. Hamkalo & J. Papaconstantinou), pp. 59–74. New York: Plenum Press.

Galloni, M., Gyurkovics, H., Schedl, P. & Karch, F. (1993). The bluetail transposon: evidence for independent *cis*-regulatory domains and domain boundaries in the bithorax complex. *EMBO Journal* **12**, 1087–1097.

Gasser, S. M. & Laemmli, U. K. (1986). The organisation of chromatin loops – characterisation of a scaffold attachment site. *EMBO Journal* **5**, 511–518.

Gatti, M. & Pimpinelli, S. (1992). Functional elements in *Drosophila melanogaster* heterochromatin. *Annual Review of Genetics* **26**, 239–275.

Glover, D. (1981). The rDNA of *Drosophila melanogaster*. *Cell* **26**, 297–298.

Healy, M., Russell, R. & Miklos, G. (1988). Molecular studies on interspersed repetitive and unique sequences in the region of the complementation group *uncoordinated* on the X chromosome of *Drosophila melanogaster*. *Molecular and General Genetics* **213**, 63–71.

Heitz, E. (1934). Über α- und β-Heterochromatin sowie Konstanz und Bau der Chromomeren bei *Drosophila*. *Biologisches Zentralblatt* **54**, 588–609.

Hoheisel, J. D., Lennon, G. G., Zehnetner, G. & Lehrach, H. (1991). Use of high coverage reference libraries of *Drosophila melanogaster* for relational data analysis: a step towards mapping and sequencing of the genome. *Journal of Molecular Biology* **220**, 903–914.

Kellum, R. & Schedl, P. (1991). A position-effect assay for boundaries of higher order chromosomal domains. *Cell* **64**, 941–950.

Kellum, R. & Schedl, P. (1992). A group of SCS elements function as domain boundaries in an enhancer blocking assay. *Molecular and Cellular Biology* **12**, 2424–2431.

Lamb, M. M. & Laird, C. D. (1987). Three euchromatic DNA sequences under-replicated in polytene chromosomes of *Drosophila* are localized in constrictions and ectopic fibers. *Chromosoma* **95**, 227–235.

Langley, C., MacDonald, J., Miyashita, N. & Aguade, M. (1993). Lack of correlation between interspecific divergence and intraspecific polymorphism at the *suppressor of forked* region in *Drosophila melanogaster* and *Drosophila simulans*. *Proceedings of the National Academy of Sciences of the USA* **90**, 1800–1803.

Le, M.-H., Duricka, D. & Karpen, G. H. (1995). Islands of complex DNA are widespread in *Drosophila* centric heterochromatin. *Genetics* **141**, 283–303.

Lefevre, G. (1981). The distribution of randomly recovered X-ray induced sex-linked genetic effects in *Drosophila melanogaster*. *Genetics* **99**, 461–480.

Lim, J. K. & Simmons, M. J. (1994). Gross chromosome rearrangements mediated by transposable elements in *Drosophila melanogaster*. *Bioessays* **16**, 269–275.

Lindsley, D. L. & Zimm, G. G. (1992). *The Genome of* Drosophila melanogaster. New York: Academic Press.

Madueno, E., Papagiannakis, G., Rimmington, G., Saunders, R. D. C., Savakis, C., Sidén-Kiamos, I., Skavdis, G., Spanos, L., Trenear, J., Adam, P., Ashburner, M., Benos, P., Bolshakov, V. N., Coulson, D., Glover, D. M., Hermann, S., Kafatos, F. C., Louis, C., Majerus, T. & Modolell, J. (1995). A physical map of the X chromosome of *Drosophila melanogaster*: cosmid contigs and sequence tagged sites. *Genetics* **139**, 1631–1647.

Miklos, G. L. G. & Cotsell, J. N. (1990). Chromosome structure at interfaces between major chromatin types: α- and β-heterochromatin. *BioEssays* **12**, 1–6.

Miklos, G. L. G., Healy, M. J., Pain, P., Howells, A. J. & Russell, R. J. (1984). Molecular and genetic studies on the euchromatin–heterochromatin transition region of the X chromosome of *Drosophila melanogaster*: a cloned entry point near to the *uncoordinated* (*unc*) locus. *Chromosoma* **89**, 218–227.

Miklos, G. L. G., Yamamoto, M. T., Davies, J. & Pirrotta, V. (1988). Microcloning reveals a high frequency of repetitive sequences characteristic of chromosome 4 and the β-heterochromatin of *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences of the USA* **85**, 2051–2055.

Mitchelson, A., Simonelig, M., Williams, C. & O'Hare, K. (1993). Homology with *Saccharomyces cerevisiae RNA14* suggests that phenotype suppression in *Drosophila melanogaster* by *suppressor of forked* occurs at the level of RNA stability. *Genes and Development* **7**, 241–249.

Pardue, M. L. & Hennig, W. (1992). Heterochromatin: junk or collector's item? *Chromosoma* **100**, 3–7.

Peacock, W., Lohe, A., Gerlach, W., Dunsmuir, P., Dennis, E. & Appels, R. (1978). Fine structure and evolution of DNA in heterochromatin. *Cold Spring Harbor Symposia on Quantitative Biology* **42**, 1121–1135.

Pimpinelli, S., Berloco, M., Fanti, L., Dimitri, P., Bonaccorsi, S., Marchetti, E., Caizzi, R., Caggese, C. & Gatti, M. (1995). Transposable elements are stable structural components of *Drosophila melanogaster* heterochromatin. *Proceedings of the National Academy of Sciences of the USA* **92**, 3804–3808.

Roseman, R., Pirrotta, V. & Geyer, P. K. (1993). The *su(Hw)* protein insulates expression of the *Drosophila melanogaster white* gene from chromosomal position effects. *EMBO Journal* **12**, 435–442.

Russell, R. J., Healy, M. J. & Oakeshott, J. G. (1992). Molecular analysis of the *lethal(1)B214* region at the base of the X chromosome of *Drosophila melanogaster*. *Chromosoma* **101**, 456–466.

Schalet, A. & Lefevre, G. (1976). The proximal region of the X chromosome. In *The Genetics and Biology of* Drosophila, vol. 1b (ed. M. A. Ashburner & E. Novitski), pp. 848–902. New York: Academic Press.

Segarra, C. & Aguade, M. (1992). Molecular organisation of the X chromosome in different species of the *obscura* group of *Drosophila*. *Genetics* **130**, 513–521.

Simonelig, M., Elliott, K., Mitchelson, A. & O'Hare, K. (1996). Interallelic complementation at the *suppressor of forked* locus of *Drosophila* reveals complementation between suppressor of forked proteins mutated in different regions. *Genetics* **142**, 1225–1235.

Spradling, A. C., Karpen, G., Glaser, R. & Zhang, P. (1993). DNA elimination in *Drosophila*. In *Evolutionary Conservation of Developmental Mechanisms*, pp. 39–53. New York: Wiley–Liss.

Spradling, A. C. & Orr-Weaver, T. (1987). Regulation of DNA replication during *Drosophila* development. *Annual Review of Genetics* **21**, 373–403.

Tartof, K. D., Hobbs, C. & Jones, M. (1984). A structural basis for variegating position effects. *Cell* **37**, 869–878.

Vaury, C., Abad, P., Pelisson, A., Lenoir, A. & Bucheton, A. (1990). Molecular characteristics of the heterochromatic I elements from a reactive strain of *Drosophila melanogaster*. *Journal of Molecular Evolution* **31**, 424–431.

Verma, R. S. (1988). *Heterochromatin: Molecular and Structural Aspects*. Cambridge: Cambridge University Press.

Williams, C. J. & O'Hare, K. (1996). Elimination of introns at the *Drosophila suppressor of forked* locus by *P* element-mediated gene conversion shows that an RNA lacking a stop codon is dispensable. *Genetics* **143**, 345–351.

Yamamoto, M. T., Mitchelson, A., Tudor, M., O'Hare, K., Davies, J. A. & Miklos, G. L. G. (1990). Molecular and cytogenetic analysis of the heterochromatin–euchromatin junction region of the *Drosophila melanogaster* X chromosome using cloned DNA. *Genetics* **125**, 821–832.

Young, M. W. (1979). Middle repetitive DNA: a fluid component of the *Drosophila* genome. *Proceedings of the National Academy of Sciences of the USA* **76**, 6274–6279.