Chromosoma (1993) 102:253–266

CHROMOSOMA
© Springer-Verlag 1993

A combined molecular and cytogenetic approach to genome evolution in *Drosophila* using large-fragment DNA cloning

Elena R. Lozovskaya, Dmitri A. Petrov, Daniel L. Hartl

Department of Organismic and Evolutionary Biology, The Biological Laboratories, Harvard University, 16 Divinity Avenue, Cambridge, MA 02138, USA

Received: 3 November 1992; in revised form: 4 December 1992 / Accepted: 5 December 1992

Abstract. Methods of genome analysis, including the cloning and manipulation of large fragments of DNA, have opened new strategies for uniting molecular evolutionary genetics with chromosome evolution. We have begun the development of a physical map of the genome of *Drosophila virilis* based on large DNA fragments cloned in bacteriophage P1. A library of 10,080 P1 clones with average insert sizes of 65.8 kb, containing approximately 3.7 copies of the haploid genome of *D. virilis*, has been constructed and characterized. Approximately 75% of the clones have inserts exceeding 50 kb, and approximately 25% have inserts exceeding 80 kb. A sample of 186 randomly selected clones was mapped by in situ hybridization with the salivary gland chromosomes. A method for identifying *D. virilis* clones containing homologs of *D. melanogaster* genes has also been developed using hybridization with specific probes obtained from *D. melanogaster* by means of the polymerase chain reaction. This method proved successful for nine of ten genes and resulted in the recovery of 14 clones. The hybridization patterns of a sample of P1 clones containing repetitive DNA were also determined. A significant fraction of these clones hybridizes to multiple euchromatic sites but not to the chromocenter, which is a pattern of hybridization that is very rare among clones derived from *D. melanogaster*. The materials and methods described will make it possible to carry out a direct study of molecular evolution at the level of chromosome structure and organization as well as at the level of individual genes.

Introduction

One of the major unsolved problems in molecular evolutionary genetics is the extent to which the organization of genes in the genome is itself a product of natural selection. Chromosome rearrangements are frequently found in the genomes of related species. For example, the genomes of *Drosophila virilis* and *D. melanogaster* are quite different in chromosome number and structure. Although the gene content of the chromosome arms is generally conserved in the two species, there is considerable scrambling of the physical locations of the genes along the arms (Alexander 1976; Loukas and Kafatos 1986; Neufeld et al. 1988; Whiting et al. 1989). The genome of *D. virilis* – 313 million base pairs (Mb) – is approximately double the 165 Mb genome of *D. melanogaster* (Laird 1973; John and Miklos 1988). Do such differences in gene order and genome size have any functional significance? At present, there is no body of data indicating whether differences in gene organization or genome size in distantly related species result from chance or necessity. We use the terms "chance" and "necessity" in the sense of Monod (1970): by chance, we mean the result of random genetic drift or other nonadaptive contingencies; by necessity, we mean the adaptive result primarily of mutation and natural selection.

In addressing these issues experimentally, a strong argument can be made for choosing *Drosophila* as a model organism. Most species of *Drosophila* carry a record of their evolutionary history in the arrangement of bands along the giant polytene chromosomes of the larval salivary glands. In many groups of species, the ancestor-descendant relationships can be inferred from careful comparisons of the chromosome banding patterns. Indeed, the interpretation of chromosome banding patterns in order to infer the phylogenetic relationships among species is one of the legacies of *Drosophila* evolutionary genetics (see reviews in Ashburner et al. 1981, 1982, 1986). Strong inferences about evolutionary history are possible because the chromosome banding patterns are unique. As emphasized by Stone (1962, p 528): "The polytene chromosomes are the largest and most complex ordered living system with a determined and consistent inherited pattern that is changed in geometrically simple and limited ways through two-break rearrangements. No other cellular organelle in natural populations of living organisms has an equivalent ordered

Communicated by: W. Hennig
Correspondence to: D.L. Hartl

complexity in which changes can so easily be determined. Therefore chromosome evolution in the genus *Drosophila* and in other Diptera with polytene chromosomes is most readily and accurately measurable."

Among the species whose polytene chromosomes are exceptionally favorable for analysis is *D. virilis* and its close relatives (collectively called the *virilis* group). The species *D. virilis* itself was first described by Sturtevant (1916). At the present time, the *virilis* group of species is subdivided into two phylads – the *virilis* phylad and the *montana* phylad. The *virilis* phylad includes five closely related species: *D. virilis*, *D. lummei*, *D. novamexicana*, *D. americana*, and *D. texana*; the *montana* phylad includes seven closely related species: *D. montana*, *D. lacicola*, *D. flavomontana*, *D. borealis*, *D. littoralis*, *D. ezoana*, and *D. kanekoi* (Throckmorton 1982). In pairwise comparisons, the species differ in more than 100 chromosomal inversions, which are either fixed or polymorphic, as well as in several chromosomal fusions and other structural variations (Patterson and Stone 1952; Throckmorton 1982). The phylogenetic relationships among the species are quite well established from chromosomal comparisons, with only minor ambiguities (Throckmorton 1982).

The species in the *virilis* group have been among the most important for studies of evolution in the subgenus *Drosophila* for several reasons. First, the species illustrate the various intergradations of genetic relationships, species differences, and reproductive isolation that exist in closely related species derived from a common ancestor (Patterson and Stone 1952; Orr and Coyne 1989): among 45 interspecific crosses in the species group that have been examined, 25 yield viable and fertile or semifertile hybrids of one or both sexes, and *D. virilis* itself is interfertile with all species in the *virilis* phylad and with four of the seven species in the *montana* phylad (Throckmorton 1982). Secondly, the relationships among species in the *virilis* group provide strong justification for defining a superspecific category – in *Drosophila* systematics designated the species group (equivalent to superspecies) – in which member species are more closely related to each other than to members of other species groups (Stone 1962; Throckmorton 1982). Third, *D. virilis* has been the object of extensive classical genetic analysis using mutants: approximately 300 loci have been identified by mutation (Alexander 1976; Gubenko and Evgen'ev 1984), many of the gene homologies with *D. melanogaster* have been identified (Alexander 1976), and an efficient system of hybrid dysgenesis for transposon tagging has been described (Lozovskaya et al. 1990; Scheinker et al. 1990). Fourth, *D. virilis* is closely related to *D. pinicola*, a rare, difficult-to-culture species that Sturtevant regarded as the most phenotypically representative of the ancestral form of the genus (Patterson and Stone 1952). The *virilis* species group, which is in the subgenus *Drosophila*, is only distantly related to the *melanogaster* species group, which is in the subgenus *Sophophora*. The time of divergence of *Drosophila* from *Sophophora* is not known with precision, but it is estimated at approximately 60 million years (Beverley and Wilson 1984).

The principal strength of cytogenetic analysis is that it is more genealogical than morphological. For example, within many species groups of *Drosophila*, the banding patterns of the polytene chromosomes are sufficiently similar that phylogenetic relationships can be inferred, even among species that are morphologically nearly indistinguishable. The principal weakness of cytogenetic analysis is that it lacks the definitiveness and generality of molecular biology. For example, between species groups of *Drosophila*, the differences in banding pattern are usually too great for homologous chromosomal regions to be identified reliably, even between species groups that, judged by morphological criteria, are virtually certain to be closely related (Stone 1962). During the past 20–30 years, perhaps because of its disconnection from molecular biology, the cytogenetic approach to *Drosophila* evolution gradually went into eclipse, and it appeared as if the bridge between chromosomal studies and molecular biology would be built very slowly – and also somewhat haphazardly – as molecular data on genes of interest to particular research groups were accumulated from multiple species. There was no question that this would happen eventually, because comparative DNA sequence data are among the most easily obtained and thus provide an efficient way to identify potentially important functional domains because these regions are conserved in evolution. With a divergence time of approximately 60 million years (Beverley and Wilson 1984), the nucleotide sequences that are conserved between *D. virilis* and *D. melanogaster* are likely to be functionally important (or else exceptionally lucky in avoiding chance fixation of nearly neutral mutations). For this, and the other reasons cited above, *D. virilis* has become the organism of choice for molecular evolutionary comparisons with *D. melanogaster*, and a significant number of comparative studies of this type have already been carried out. For example, GenBank release 73.0 (September 1992) contains 36 coding sequences from *D. virilis* that are homologous to sequences in *D. melanogaster*. To date, the most ambitious comparative study between *D. virilis* and *D. melanogaster* comprises 160 kb of DNA sequence in each species in the region of the Antennapedia complex (Hooper et al. 1992). There have also been studies to determine chromosomal homologies between *D. virilis* and *D. melanogaster* by in situ hybridization with the salivary gland chromosomes (Loukas and Kafatos 1986; Whiting et al. 1989).

More recently, methods of genome analysis have opened new strategies for uniting molecular evolutionary genetics with chromosome evolution. Genome analysis includes the cloning and manipulation of large fragments of DNA in the size range 100–1,000 kb, the assembly of cloned fragments into continuous physical maps of the genome, and ultimately large-scale DNA sequencing (Watson 1990). Physical mapping of the genome of *D. melanogaster* is well underway by several groups using complementary strategies (Merriam et al. 1991). Our laboratory has adopted a mapping strategy based on in situ hybridization of large fragments of *Drosophila* DNA cloned into yeast artificial chromosomes (YACs) (Ajioka et al. 1991; Hartl et al. 1992) and bacteriophage

P1 vectors (Smoller et al. 1991; Hartl and Lozovskaya 1992). At present, the YAC-based map of the euchromatic genome comprises 149 discrete regions of continuous coverage (contigs) ranging in size from 200–4,200 kb and averaging approximately 650 kb (Hartl 1992). (Euchromatin refers to the portions of the chromosomes that become polytene in the salivary gland chromosomes, approximately the distal two-thirds of all chromosome arms except for the *Y* chromosome.) In addition to the YACs, more than 1,250 P1 clones from the euchromatin have been mapped by in situ hybridization (Lozovskaya et al., unpublished). These materials provide access to molecular probes throughout the genome and make it possible to examine the genome on a scale with lower resolution than nucleotide sequences but higher resolution than cytological observation.

With the long-term objective of resolving issues of chance versus necessity in the evolution of the *Drosophila* genome, we have initiated the development of a physical map of the genome of *D. virilis*. Because they have a number of technical advantages over YACs (Sternberg 1992), bacteriophage P1 clones have been chosen as the primary vector. A library of P1 clones containing multiple copies of the genome of *D. virilis* has been constructed and characterized. A sufficient sample of clones has been mapped by in situ hybridization with the salivary gland chromosomes in order to yield an average density of clones throughout the euchromatic genome of greater than one P1 clone per million base pairs. The hybridization patterns of a sample of clones containing inserts derived from heterochromatin have also been determined. Heterochromatin includes the pericentromeric regions of the chromosomes and the *Y* chromosome (alpha heterochromatin), plus a diffuse, meshlike region at the base of each chromosome arm (beta heterochromatin); the alpha heterochromatin is grossly underreplicated in salivary gland nuclei, and the beta heterochromatin remains more or less underreplicated (Miklos and Cotsell 1990). Approximately 50% of the genome of *D. virilis* consists of heterochromatin, as compared with approximately 25% in *D. melanogaster* (Gall 1973; Laird 1973; Miklos 1985; Lohe and Brutlag 1986). Finally, a sample of sequenced genes from *D. melanogaster* was used to derive probes for screening the *D. virilis* P1 library in order to provide a general method for isolating homologous *D. virilis* genes and for correlating the physical maps.

Materials and methods

Drosophila strains. The P1 library was constructed with DNA isolated from *D. virilis* strain 9, a wild-type strain collected in 1970 in Batumi, Georgia (former USSR), and maintained in laboratory culture since that time (Lozovskaya et al. 1990). In situ hybridizations were carried out with larvae from strain 9 and, in some cases, also with *D. virilis* strain 160. Strain 160 is a laboratory strain containing mutations in each of the large autosomes as follows: chromosome *2, b* (broken crossveins); chromosome *3, tb* (tiny bristles) and *gp-L2* (gap in longitudinal wing vein 2); chromosome *4, cd* (cardinal eye color); chromosome 5, *pe* (peach eye color); and chromosome *6, gl* (glossy eye surface) (Lozovskaya et al. 1990).

Origin of cloned DNA. High molecular weight DNA was prepared from nuclei of 0.5 g of adult fly tissue by homogenization in NIB buffer (10 mM Tris-HCl, 10 mM NaCl, 10 mM EDTA, 0.15 mM spermine, 0.15 mM spermidine, 0.5% Triton X-100, pH 8.5). After centrifugation, the nuclear pellet was washed twice with NIB with 3% sarkosyl. The DNA was banded in a CsCl gradient (Bingham et al. 1981) and dialyzed three times against TE (10 mM Tris-HCl, 1 mM EDTA, pH 8.0). High molecular weight DNA recovered from the CsCl gradient was partially digested with Sau3A to produce a high fraction of molecules in the size range 60–150 kb, using the methods outlined in Smoller et al. (1991). Aliquots of 400 µl of this material were layered onto 18 ml sucrose gradients (10%–40% sucrose) and centrifuged in an AH627 rotor for 20 h at 29,100 *g*. Fractions of 300 µl were collected from the gradients and aliquots analyzed by electrophoresis in a contour-clamped homogeneous electric field (CHEF) gel (Vollrath and Davis 1987). Fractions containing DNA predominantly in the size range 70–100 kb were dialyzed for 1.5 h at 25° C on filters (Millipore type VSWP 02500) floated on TE. Fractions were concentrated with *sec*-butanol and precipitated in ethanol.

Library construction. Construction of the P1 library followed procedures essentially as outlined in Smoller et al. (1991) but using a different P1 vector designated pAd10sacBII diagrammed in Fig. 1 (Pierce et al. 1992). The cloning site of the vector is a BamHI site flanked progressively on the left by a T7 promoter, a NotI site, and the promoter of the *sacB*II gene of *Bacillus amyloliquefaciens*, and flanked progressively on the right by an SP6 promoter, an SfiI site, and the *sacB*II structural gene for levansucrase. Large fragments of DNA inserted into the cloning site disrupt expression of the *sacB*II gene and thereby allow cells of *Escherichia coli* to survive in medium containing 5% sucrose. Vector arms resulting from digestion of pAd10sacBII with BamHI and ScaI were treated with calf intestinal alkaline phosphatase and ligated with Sau3A fragments of *D. virilis* as described in Smoller et al. (1991). The resulting linear molecules were cleaved at the *pac* site in the vector (Fig. 1) and packaged in vitro as described in Sternberg (1990). Molecules are packaged stepwise in the counterclockwise direction from the *pac* site (as drawn in Fig. 1) until the phage head has been filled (approximately 100–115 kb), after which cleavage occurs. Packaged phage were used to infect *E. coli* strain NS3529, which carries the mutations *recA mcrA Δ(mcrB, mrr, hsdR, hsdM)* along with two λ prophages (λimm434nin5X1-cre and λimmlLP1). This strain produces a site-specific recombinase that targets the *loxP* sites and circularizes the vector by recombination (Sternberg 1990). The infected NS3529 cells were plated on LB agar containing 25 µg/ml kanamycin and 5% sucrose. The resulting colonies, each representing an independent P1 clone, were picked individually and arrayed in the wells of 105 microtiter plates, each containing 96 wells, for a total of 10,080 clones.

DNA preparation from P1 clones. Bacterial clones containing single P1 clones were inoculated into LB medium containing 25 µg/ml kanamycin and 1 mM IPTG (isopropyl 1-thio-β-D-galactopyranoside) and grown overnight at 37° C. Plasmid DNA was extracted using the alkaline lysis method (Birnboim and Doly 1979).

Gel electrophoresis. DNA from P1 clones was digested with EcoRI, BamHI, or BglII and the resulting fragments separated in conventional agarose gels (Sambrook et al. 1989). Large fragments produced by NotI or SfiI digestion were separated in 20 cm × 20 cm, 1% agarose gels in 0.5 TBE buffer (0.045 Tris-borate, 0.001 M EDTA, pH 8.3) in a BioRad CHEF-DR II apparatus for CHEF gel electrophoresis (Vollrath and Davis 1987).

Cytological analysis. DNA from P1 clones was labeled with biotin-dCTP or biotin-dUTP by the random hexamer method (Feinberg and Vogelstein 1984). Hybridization of labeled DNA to polytene chromosome squashes in situ was carried overnight at 37° C as described by Langer-Safer et al. (1982) in 1.4 × SSC (1 × SSC is 0.15 M NaCl, 0.015 M sodium citrate), 7% dextran sulfate, 35%

Table 1. Buffer and hybridization conditions[a]

Reagent	1	2	3	4	5	6
SSC	4×	1.4×	4×	4×	2×	4×
Dextran sulfate (%)	2	7	2	10	7.5	10
Denhardt's solution	1×	–	10×	–	–	1×
Formamide (%)	–	35	–	–	65	–
Salmon sperm DNA (mg/ml)	–	0.6	0.6	–	–	–
Temperature (°C)	58.5	37	33	65	4, 10, or 16	58.5
Time (h)	16	16	16	16	32–48	16

[a] Conditions used in attempting direct in situ hybridization of P1 clones from *Drosophila melanogaster* with salivary gland chromosomes of *D. virilis* and vice versa. Hybridization was detected using the horseradish peroxidase method (Enzo Biochemicals) on chromosomes stained with Giemsa

N,N-dimethyl formamide, and 0.6 mg/ml sonicated denatured salmon sperm DNA. A variety of conditions were tried in an attempt to hybridize P1 clones from *D. melanogaster* (Smoller et al. 1991) directly with salivary gland chromosomes from *D. virilis* (or the other way around). These included the conditions listed in Table 1. Denhardt's solution (1×) is 0.02% bovine serum albumin, 0.02% Ficoll, and 0.02% polyvinylpyrrolidone of average molecular weight 360,000. The set of conditions listed in column 1 is that used by Whiting et al. (1989) in their hybridizations with conserved genes cloned in plasmids. The set of conditions in column 2 is that used routinely in our laboratory for in situ hybridization of P1 clones with salivary gland chromosomes of the same species. None of the sets of conditions in Table 1 gave satisfactory results in the cross-specific hybridizations. Tentatively, we believe that the problem results from the long evolutionary divergence time between *D. virilis* and *D. melanogaster*. Comparisons of coding and regulatory sequences with noncoding sequences between *D. melanogaster* and *D. virilis* indicate that relatively short stretches of conserved sequence are interrupted by relatively long stretches of nonconserved sequence (Bray and Hirsh 1986; Kassis et al. 1986; Colot et al. 1988; Heberlein and Rubin 1990). The relative length and interspersion of conserved and nonconserved sequences implies that any region of a P1 insert that is sufficiently homologous for hybridization will have a relatively low specific activity because of the large amount of nonhybridizing material present in the same clone.

Screening the D. virilis P1 library. Nylon membranes, each containing 384 geometrically arrayed P1 clones, were inoculated with liquid cultures stored in 96-well microtiter dishes using a custom-fabricated multiprong replicator (Brownstein et al. 1989). Filters were placed onto LB plates containing 1 mM IPTG and 25 μg/ml kanamycin for 6 h at 37° C. Cells were lysed on chromatography (Whatman 3 MM) paper saturated with 3% sodium dodecyl sulfate (SDS), denatured on 3 MM paper saturated with 0.5 M NaOH and 1.5 M NaCl, neutralized on 3 MM paper saturated with 0.5 M Tris-HCl, 1.5 M NaCl, pH 7.5, and washed on 3 MM paper containing 2×SSC. The filters were then baked at 80° C for 40 min. Standard hybridization conditions were high phosphate buffer (0.5 M NaCl, 0.1 M Na_2HPO_4, 0.005 M EDTA, 1% SDS) at 65° C for 16–18 h (Smoller et al. 1991). Filters were washed twice in 1% Sarkosyl, 1 mM Tris-HCl, pH 8.0, and twice in 1 mM Tris, pH 8.0, at 25° C. Autoradiography was for 16–18 h.

Hybridizations at reduced stringency, using probes obtained by the polymerase chain reaction (PCR) from sequenced genes of *D. melanogaster*, were carried out overnight at 42° C in 5×SSPE (1×SSPE is 0.15 M NaCl, 0.01 M NaH_2PO_4 and 0.001 M EDTA), 0.1% SDS, 1×Denhardt's solution, 0.1 mg/ml sonicated denatured salmon sperm DNA, and 29% formamide (O'Neil and Belote 1992). After hybridization, filters were washed at 50° C in three changes of 0.5×SSPE containing 0.1% SDS.

PCR and probes. Probes present in plasmids were labeled with ^{32}P by the random hexamer method (Feinberg and Vogelstein 1984). However, most probes were labeled with ^{32}P by means of incorporation during amplification of genomic DNA of *D. melanogaster* using the PCR (Saiki et al. 1985). The labeled PCR products were separated from unincorporated nucleotides by centrifugation through G50 Sephadex columns and the probes used directly for subsequent hybridization with P1-containing filters. PCR amplification was carried out in 20 μl of reaction buffer containing 25 μM each of dCTP, dGTP, and dTTP; 1 μl of high-specific-activity [^{32}P]-ATP; 1 unit of Taq polymerase; and 0.2 μM of each primer. PCR reaction buffer consists of 10 mM Tris-HCl, pH 8.4, 50 mM KCl, 2.5 mM $MgCl_2$, and 0.01% gelatin. Standard amplification conditions were 25 cycles of denaturation at 95° C for 1 min, primer annealing (typically at 55° C, but adjusted for individual primer pairs) for 1 min, and primer extension at 72° C for 1 min. The oligonucleotides used for PCR amplification of *D. melanogaster* genes were as follows: *y* (yellow): 3999–4019 (5′-GATTCCGGCCACTCTGACCTA-3′) and 5033–5013 (5′-ACGGCTTGTTTTGGTATTGAA-3′), PCR product 1,035 bp, annealing temperature 55° C; *w* (white): 71–91 (5′-GGGCAGTTCCGGTGCCGGAA-3′) and 448–428 (5′-GTCTGGCATTCGCCTCGCT-3′), PCR product 378 bp, annealing temperature 55° C; *sn* (singed): 101–121 (5′-CGATCTCATCTGAACAAGTACC-3′) and 500–479 (5′-GCACGTATGCAGAGCATAGCGA-3′), PCR product 400 bp, annealing temperature 60° C; *Pcp* (Pupal cuticle protein): 261–282 (5′-CTAGTGATGACGGGGGTGCTC-3′) and 606–587 (5′-AACCCCCCGGTGCCCCATTC-3′), PCR product 325 bp, annealing temperature 65° C; *mam* (master mind): 520–540 (5′-CAGCAGAGTTTGGGCATTCAC-3′) and 1100–1080 (5′-TGTGTCTGTGCCTGCTGTTGG-3′), PCR product 581 bp, annealing temperature 59° C; *Hsp82* (Heat shock protein 82): 2160–2189 (5′-CCAGAAGAAGCAGAGACCTT-3′) and 3034–3018 (5′-TCCTCCTGGGAGATATC-3′), PCR product 875 bp, annealing temperature 58° C; *Sgs3* (Salivary gland secretion protein 3): 1–23 (5′-GAATTCGCAACGCTTTCTGTGTG-3′) and 840–820 (5′-CTCATAACTGATACTGATTG-3′), PCR product 840 bp, annealing temperature 53° C; *hb* (hunchback): 5917–5936 (5′-CAGCCAGTCGCCACATCTCA-3′) and 6623–6604 (5′-GCCGTCTCCTACGACTCATC-3′), PCR product 707 bp, annealing temperature 62° C; *Rp49* (Ribosomal protein 49, large subunit): 441–460 (5′-AGGCCCAAGATCGTGAAGAA-3′) and 876–856 (5′-CTGCGTCTCAAGAAGAACGA-3′), PCR product 436 bp, annealing temperature 58° C; *v* (vermilion): 1690–1705 (5′-ACCAAGTGCCCATTC-3′) and 2058–2073 (5′-TGAGGCGGTAGTTTTT-3′), PCR product 384 bp, annealing temperature 54° C. The histone genes were probed with a mixture of PCR products that hybridizes with all the histone genes, generously provided by L.D. Strausbaugh. The *norpA* (no receptor potential) gene was obtained by hybridization using plasmid pSHIBI39 containing cDNA for the entire *norpA* coding region from *D. melanogaster* (Bloomquist et al. 1988), generously provided by W.L. Pak. The clones containing ribosomal DNA (rDNA) were obtained by hybridization using plasmid p28SSP containing a 1.3 kb BglII – HindIII fragment from the 28 S ribosomal RNA (rRNA) gene of *D. melanogaster* contained in pDM238 (Lohe and Roberts 1990), generously provided by A.R. Lohe. Clones containing the *su(s)* suppressor of sable) gene were obtained by screening with a 900 bp fragment of the *su(s)* gene from *D. virilis*, generously provided by R.A. Voelker.

Results

Molecular characterization of the P1 library

The *D. virilis* P1 library consists of 10,080 individual clones arrayed in the wells of standard 96-well microtiter

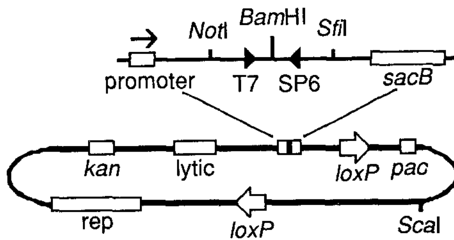

Fig. 1. Diagram of the P1 cloning vector pAd10sacBII (Pierce and Sternberg 1991; Pierce et al. 1992). The *kan* gene codes for kanamycin resistance. The plasmid replicon (*rep*) maintains the plasmid at approximately one copy per cell. The lytic replicon (*lytic*), which is under the inducible control of the *Escherichia coli* β-galactosidase promoter, increases the copy number to approximately 20

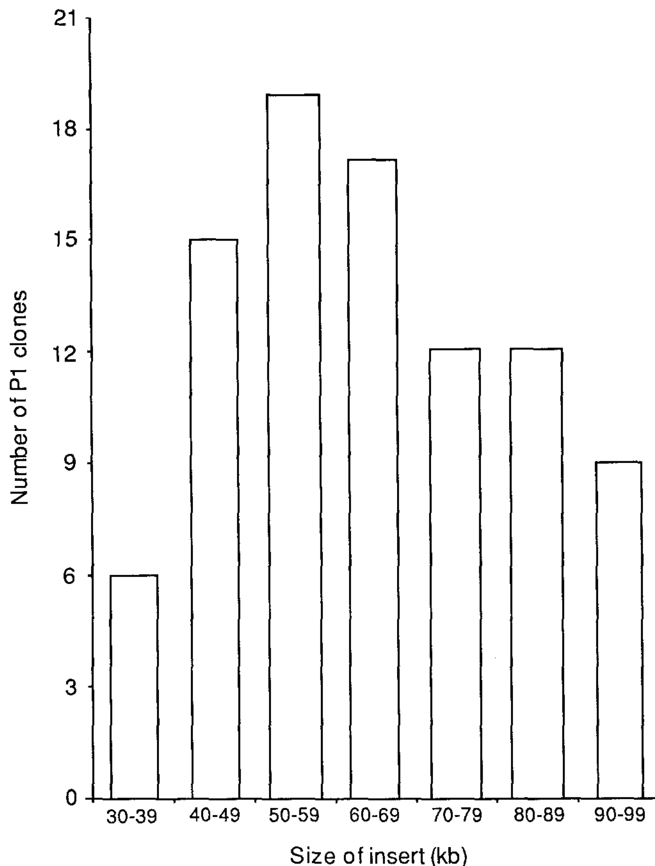

Fig. 2. Distribution of insert sizes among 90 randomly selected P1 clones containing DNA fragments from *Drosophila virilis*

plates. The library was constructed using the P1 cloning vector pAd10sacBII (Fig. 1) because this vector enables positive selection for recombinant molecules containing inserted DNA fragments (Pierce and Sternberg 1991; Pierce et al. 1992). The sizes of the inserts were estimated from 90 randomly selected clones. The size of each P1 clone was determined by summing the sizes of the restriction fragments produced from complete digestion with EcoRI, BamHI, or BglII. The size of each insert was obtained from the total by subtracting the 16.6 kb attributable to vector sequences (Pierce and Sternberg 1991). A subset of estimated sizes was checked independently by digesting the P1 clones to completion with either NotI or SfiI and separating the fragments in CHEF gels (Vollrath and Davis 1987). The distribution of insert sizes among the P1 clones is summarized in Fig. 2. Although some of the clones contain relatively small inserts (< 40 kb), most of the inserts are larger. The average insert size was 65.8 with an SEM (standard error of the mean) of ± 2.1 kb. Approximately 75% of the clones have inserts exceeding 50 kb, and approximately 25% have inserts exceeding 80 kb.

Among the clones in Fig. 2, a sample of 58 was mapped by in situ hybridization to the salivary gland chromosomes (see below). The hybridization patterns included 43 clones with major euchromatic sites of hybridization, frequently with additional weak hybridization at multiple euchromatic sites, probably resulting from repetitive DNA present in the inserts, and often with the chromocenter as well. An additional 10 clones showed strong to moderate hybridization with the chromocenter, which consists mainly of underreplicated pericentromeric heterochromatin and the *Y* chromosome, and frequently showed multiple weak euchromatic sites of hybridization as well. Five clones showed only multiple weak euchromatic sites of hybridization without hybridization to the chromocenter. Among these three sets of clones, the average size of insert in the P1 clones and the SEM was 67.5 ± 2.9 kb, 64.6 ± 5.8 kb, and 75.4 ± 11.5 kb. Hence, P1 clones with primary sites of hybridization in the chromocenter – or clones containing abundant repetitive DNA with no single major euchromatic site of hybridization – do not generally have smaller inserts than P1 clones that hybridize to single major euchromatic sites.

Cytological distribution of P1 clones

To assess the coverage of the genome qualitatively, as well as to provide a set of clones from diverse regions of the euchromatin, in situ hybridizations of 186 randomly selected P1 clones were carried out with the polytene salivary gland chromosomes. The cytological distribution of 164 clones mapping to the euchromatin is shown by the solid dots in Fig. 3. The open circles depict 23 clones identified by screening the *D. virilis* P1 library with specific probes for the various genes indicated (see next section). A detailed description of these and other mapped clones is presented in the Appendix. A few of the clones are apparent duplicates picked twice from the original plates containing the P1 transformants. Excluding the duplicates, as well as those identified by screening the library, the distribution of randomly selected P1 clones was 25 on the *X* chromosome, 44 each on chromosomes *2* and *3*, and 24 each on chromosomes *4* and *5*.

The numbers and letters in Fig. 3 refer to the terminology for the polytene chromosomes and bands, which is taken from the cytological maps of Gubenko and Evgen'ev (1984). The large chromosomes are designated *X* (the sex chromosome) and *2–5* (the large autosomes). Chromosome *6*, a tiny "dot" chromosome corresponding to *D. melanogaster* chromosome *4*, is not shown be-

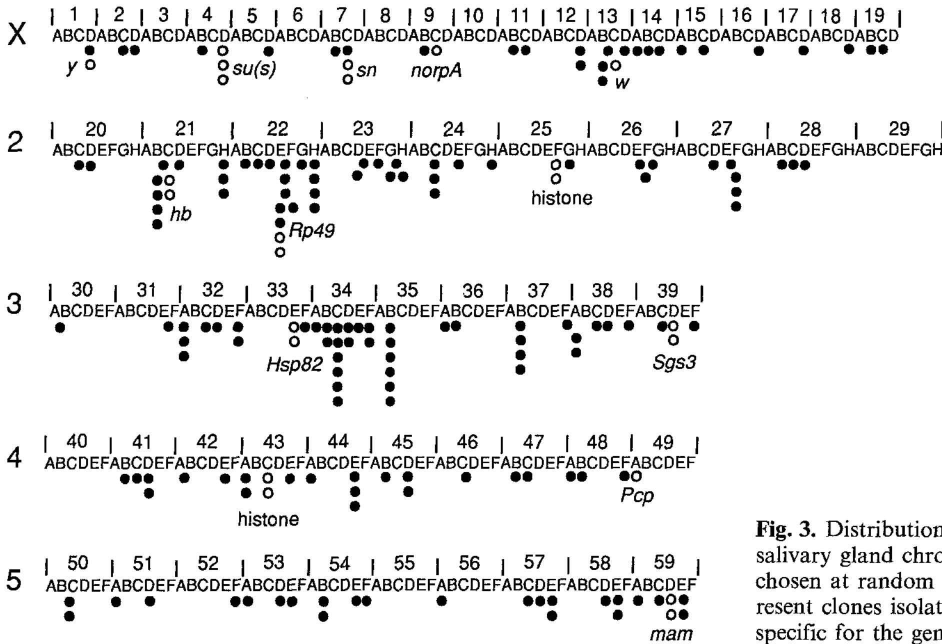

Fig. 3. Distribution of P1 clones in the euchromatin of the salivary gland chromosomes. The *solid dots* represent clones chosen at random from the P1 library. The *open circles* represent clones isolated by screening the library with probes specific for the genes indicated

cause no P1 clones from *6* have been recovered to date. The *X* chromosome is divided into 19 numbered sections starting at the tip (sections 1–19); each of the four large autosomes is similarly divided into 10 numbered sections (chromosome *2*, sections 20–29; chromosome *3*, sections 30–39; chromosome *4*, sections 40–49; chromosome *5*, sections 50–59); and chromosome *6* comprises section 60. Each numbered section, in turn, is subdivided into lettered subdivisions from distal to proximal. Sections 1–19 (*X*) and 60 (chromosome *6*) each have four subdivisions A–D; sections 20–29 (chromosome *2*) each have eight subdivisions A–H; and sections 30–59 (chromosomes *3*, *4*, and *5*) each have six subdivisions A–F. The bands in each subdivision are numbered consecutively from left to right.

The number of clones localized by in situ hybridization is still too limited to support a quantitative statistical analysis of the randomness and completeness of coverage. However, the density of mapped clones is sufficient to provide molecular probes of the euchromatic genome of *D. virilis* with an average spacing between probes of approximately 1 Mb.

Isolation of P1 clones by library screening

Initially we had hoped that P1 clones from *D. melanogaster* could be used directly for in situ hybridization with chromosomes of *D. virilis*, but we were unable to identify conditions making this generally possible (see Materials and methods). The main reason for the difficulty may be that there is too much nucleotide divergence in noncoding regions and at synonymous sites in coding regions between *D. virilis* and *D. melanogaster*. These species are in different subgenera (*Drosophila* and *Sophophora*, respectively) and shared a last common ancestor approximately 60 million years ago (Beverley and Wilson 1984). The divergence time among species of *So-*

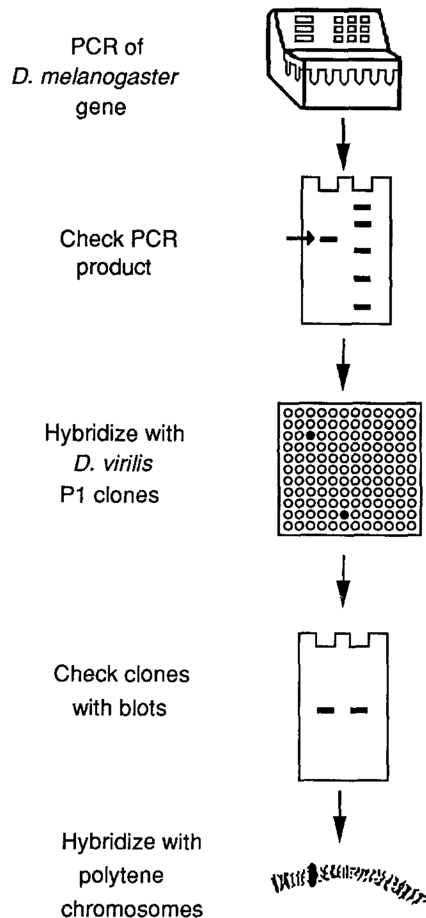

Fig. 4. Method for isolating *Drosophila virilis* P1 clones containing homologs of *D. melanogaster* genes. Probe DNA is produced by polymerase chain reaction (PCR) amplification from genomic DNA of *D. melanogaster* and checked by electrophoresis for size and spurious amplification products. The probe is then used to screen the *D. virilis* P1 library. Positive clones are verified by DNA hybridization and localized in the genome by in situ hybridization with the polytene chromosomes

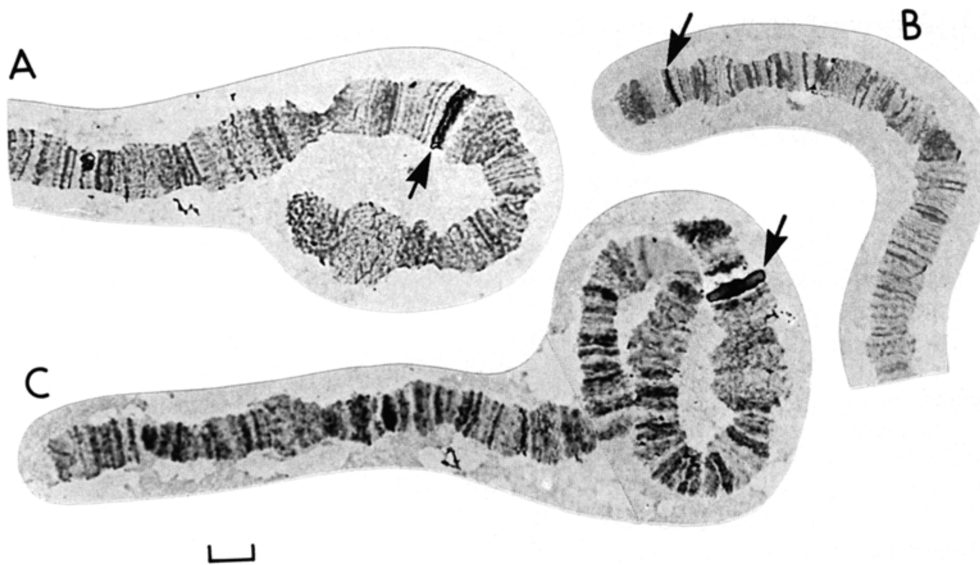

Fig. 5A–C. Representative in situ hybridizations of P1 clones of *Drosophila virilis*. **A** Clone v95–39 containing the *hb* gene maps to *2*: 21C. **B** Clone v10–39 containing the *y* gene maps to *X*: 1D. **C** Clone v33–94 containing the *mam* gene maps to **5**: 59D. Bar represents 10 μm

phophora is considerably shorter, and in these cases in situ hybridizations across species work reasonably well (Segarra and Aguadé 1992).

As an alternative, we used the method outlined in Fig. 4. Probes were produced by means of PCR from genomic DNA of *D. melanogaster* using appropriate oligonucleotide primers. The PCR products were verified for size and absence of spurious amplification products using gel electrophoresis and then used to screen the *D. virilis* P1 library by hybridization with DNA from P1-containing *E. coli* colonies immobilized on filters. Clones with positive hybridization signals were identified and verified by means of conventional Southern (1975) blots using the *D. melanogaster* PCR product as probe. Finally, each clone was assigned a position on the *D. virilis* polytene chromosome map by in situ hybridization. Except for *su(s)* (suppressor of sable), the open circles in Fig. 3 represent clones identified by this hybridization procedure. The *su(s)* clones were identified directly by hybridization with a fragment of the *su(s)* gene from *D. virilis*. More detailed information on the clones and their locations in *D. virilis* and *D. melanogaster* is given in Table 2. Among ten *D. melanogaster* genes examined using the procedure in Fig. 4, only *v* (vermilion) failed to yield P1 clones containing the *D. virilis* homolog.

A number of lines of evidence indicate that the P1 clones identified by the procedure in Fig. 4 do contain genes that are homologous with the genes from *D. melanogaster* from which the probes were obtained. First, the P1 clones putatively containg *y* (yellow) (Fig. 5B), *sn* (singed), *w* (white), and the histone genes map by in situ hybridization to the known positions of these genes in the salivary gland chromosomes (Alexander 1976; Gubenko and Evgen'ev 1984); the cytological position of *norpA* (no receptor potential) also agrees with the predicted position of this gene based on its relation to other genes on the *X* chromosome that have been mapped in both *D. virilis* and *D. melanogaster*. Secondly, most probes – for example, *hb* (hunchback) (Fig. 5A) and *mam* (mastermind) (Fig. 5C) – yielded multiple in-

Table 2. P1 clones from *Drosophila virilis* identified by homology with probes from *D. melanogaster*

Gene[a]	Location (*D. melanogaster*)	P1 clones	Location (*D. virilis*)
y	*X* at 1B1	v10–34	*X* at 1D
w	*X* at 3C2	v26–79	*X* at 13C
norpA	*X* at 4B6–C1	v33–95	*X* at 9C
sn	*X* at 7D1–2	v29–64	*X* at 7C
		v37–19	*X* at 7C
Pcp	*2L* at 27C	v93–93	*4* at 49A2
Histone genes[b]	*2L* at 39D–E	v40–45	*4* at 43C (also *2* at 25F)
		v21–67	*4* at 43C (also *2* at 25F)
mam	*2R* at 50C20–23	v33–94	*5* at 59D
		v02–12	*5* at 59D
Hsp82	*3L* at 63B–C	v40–74	*3* at 33E
		v50–16	*3* at 33E
Sgs3	*3L* at 68C	v35–7	*3* at 39D
		v36–27	*3* at 39D
hb	*3R* at 85A–B	v11–07	*2* at 21C
		v95–39	*2* at 21C
Rp49	*3R* at 99D	v88–16	*2* at 22E
		v81–70	*2* at 22E
Ribosomal DNA[c]	Chromocenter	v68–38	Chromocenter
		v71–36	Chromocenter

[a] Gene symbols: *y*, yellow; *w*, white; *norpA*, no receptor potential A; *sn*, singed; *Pcp*, Pupal cuticle protein; *mam*, master mind; *Hsp82*, heat shock protein 82; *Sgs3*, Salivary gland secretion protein 3; *hb*, hunchback; *Rp49*, large-subunit ribosomal protein 49 [synonym: Minute(3)99D]

[b] A total of eight P1 clones containing histone genes were recovered, two of which were examined by in situ hybridization. Designations: v40–45, v12–21, v33–78, v21–67, v22–46, v40–61, v1–4(?), v29–32(?). The *D. melanogaster* histone clones are: 12–62, 12–92, 14–51, 17–11, 17–34, 20–46, 30–32 and (either 1–52 or 3–52) and (either 39–27 or 40–27)

[c] A total of 16 P1 clones containing rDNA were recovered, 2 of which were examined by in situ hybridization

dependent clones that hybridized to the same location in the salivary gland chromosomes. Third, the clones containing the putative homologous genes map to the predicted chromosome based on the known chromosome homologies with *D. melanogaster* (Alexander 1976; Gubenko and Evgen'ev 1984).

Stability of P1 clones containing rDNA

To examine the stability of P1 clones containing repetitive DNA sequences, clones containing rDNA from *D. virilis* were subcultured repeatedly by streaking the bacteria for single colonies and growing them overnight in liquid culture. This cycle was repeated four times. After each cycle an aliquot of culture was used to prepare the P1 clones for digestion with EcoRI and/or BglII and electrophoresis as described in Smoller et al. (1991). Among 14 rDNA clones tested in this manner, 13 yielded an invariant pattern of restriction fragments throughout the experiment (data not shown). These data are comparable to those of Smoller et al. (1991) in suggesting that P1 clones containing repetitive DNA with the sequence complexity of rDNA are not subject to frequent and large-scale rearrangement.

Hybridization patterns of clones containing repetitive DNA

Many of the P1 clones also contain repetitive DNA sequences that hybridize with multiple sites at scattered locations in the euchromatin. Most of these sequences probably represent transposable elements of various types (Lozovskaya et al. 1990). There is minor variation in the number of hybridization sites within strains, typically of the magnitude one to four sites per chromosome arm, and therefore it is not possible to determine whether two P1 clones contain the same transposable element based on number of hybridization sites alone (e.g., clone v71-3 versus v71–40). Furthermore, considering the relatively large size of the inserts in P1 clones, a particular

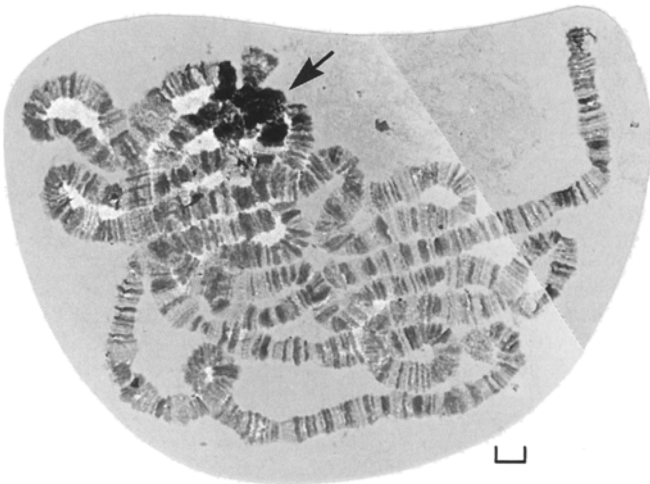

Fig. 6. In situ hybridization of clone v68–52, which hybridizes primarily with the chromocenter and the base of chromosome *6*. In some preparations, there are also approximately ten sites of weak hybridization in the euchromatin. Bar represents 10 μm

Table 3. P1 clones that hybridize with multiple euchromatic sites and the chromocenter

P1 clone	*X*	*2*	*3*	*4*	*5*	Total	*6*	Chromocenter
v68–52	3	3	2	1	2	11	+	+
v68–83	1	5	4	2	5	17	1	+/−
v71–91	3	10	2	2	2	19	+	+
v71–3	3	6	5	7	4	25	−	+
V71–40	4	5	5	5	6	25	+	+
v71–95	4	16	11	12	9	52	+	++
v71–13	20	18	1	13	+	>52	+	++
v68–70	5	4	25	15	11	60	+	+
v68–61	14	13	16	11	6	60	+	+
v71–2	14	13	16	13	7	63	+	++
v71–25	20	23	3	8	10	64	+	+
v68–8	8	20	15	14	9	66	+	++
v71–7	13	15	19	12	7	66	−	+/−
v68–96	18	18	11	15	14	76	+	++
v71–89	19	20	20	17	7	83	−	+/−
v71–81	30	20	8	20	13	91	+	+
v67–76	22	39	14	25	26	126	+	++
v71–15	40	35	25	16	27	143	+	++
v71–45	50	45	20	16	22	153	+	+/−
v68–82	52	46	15	17	27	157	+	+/−

clone may contain members of more than one family of transposable elements. Among 186 clones analyzed, 50 hybridized with multiple euchromatic sites. Many of these clones also showed hybridization with the chromocenter, in which the underreplicated mass of pericentromeric heterochromatin and the *Y* chromosome are located. A typical example is shown in Fig. 6. The clones with the multiple-sites-plus-chromocenter pattern of hybridization, sorted according to the total number of hybridization sites in chromosomes *X* through *5*, are listed in Table 3. The numbers shown are averages of the counts in several nuclei, rounded to integers. In addition to generally minor variation in euchromatic hybridization sites within a strain, there is sometimes very great variation between strains. For example, P1 clone v71–89 hybridizes with an average of 83 sites in the euchromatic arms of strain 9 but with just one euchromatic site in strain 160.

A second type of pattern with multiple hybridization sites in the euchromatin shows no hybridization in the chromocenter. Chromosome *6* usually also remains unlabeled. An example of this type of pattern is shown in Fig. 7. A list of 14 clones showing the pattern is presented in Table 4. Considering that the range and distribution in the number of euchromatic hybridization sites is very similar to those of the clones listed in Table 3, it seems highly unlikely that the multiple-sites-not-chromocenter pattern results from repeated isolation of P1 clones containing a particular transposable element with an unusual distribution in the genome. Rather, the genome of *D. virilis* must contain multiple transposable elements that are well represented in euchromatin but virtually absent from the heterochromatin.

The clones in Table 4 are of some interest in the context of genome evolution because they have a pattern of hybridization that, although quite common in *D. viri-*

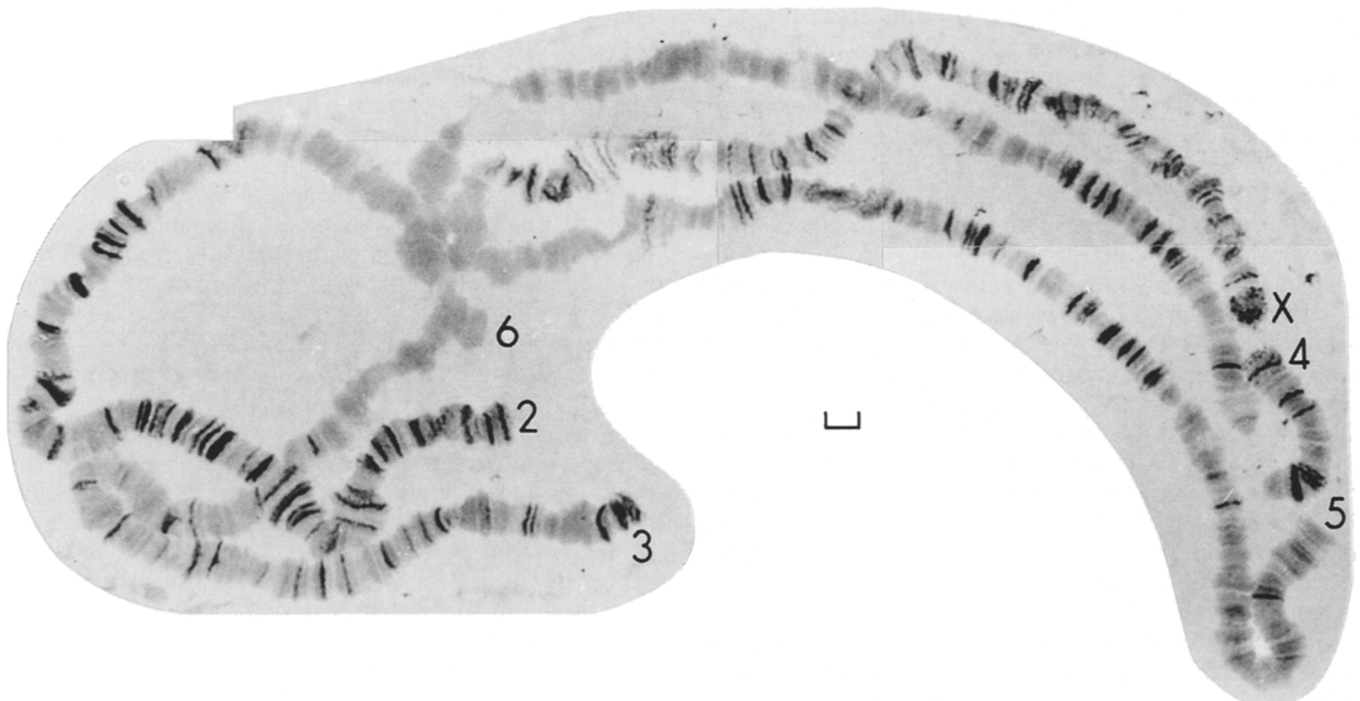

Fig. 7. In situ hybridization of clone v71–9, which hybridizes with multiple euchromatic sites but does not hybridize with the chromocenter. Bar represents 10 μm

Table 4. P1 clones that hybridize with multiple euchromatic sites and not the chromocenter

P1 clone	*X*	*2*	*3*	*4*	*5*	Total	*6*	Chromo-center
v71–75	3	1	7	2	1	14	–	–
v68–21	2	1	4	4	6	17	–	–
v71–37	3	0	15	7	4	29	–	–
v68–64	2	3	15	15	4	39	–	–
v68–81	1	0	26	13	10	50	–	–
v68–37	6	15	15	11	6	53	–	–
v68–80	1	2	28	15	12	58	–	–
v68–84	15	14	15	12	7	63	–	–
v71–61	5	2	30	20	18	75	–	–
v71–4	7	2	30	30	12	81	–	–
v71–21	21	18	4	13	27	83	–	–
v68–93	30	23	10	12	16	91	–	–
v68–91	44	36	10	12	24	126	–	–
v71–9	55	45	25	27	24	176	–	–

Table 5. Hybridization patterns of P1 clones chosen at random from *Drosophila virilis* and *D. melanogaster* libraries[a]

Hybridization pattern	*D. virilis*	*D. melanogaster*
Single major euchromatic site	73.1%	85.7%
Chromocenter (including those with repetitive euchromatic sites)	21.0%	14.0%
Multiple euchromatic sites	5.9%	0.3%
Total number of clones	186	867

[a] Each P1 clone was hybridized in situ with salivary gland chromosomes of the species of origin. An additional 50 P1 clones obtained from screening the library were also localized by in situ hybridization but are not included in this table

lis, is very uncommon in *D. melanogaster*. The detailed comparison is given in Table 5. Among 867 random P1 clones from *D. melanogaster* hybridized to the salivary gland chromosomes of *D. melanogaster*, only 3 clones (0.3%) showed this pattern of hybridization (Lozovskaya, Jones and Hartl, unpublished). In contrast, among 186 random clones from *D. virilis* hybridized with *D. virilis* salivary gland chromosomes, 11 (5.9%) showed the multiple-site-not-chromocenter pattern. The difference between the results is highly significant ($P \ll 0.001$). Another noteworthy difference between the two species is that, among clones with the multiple-sites-plus-chromocenter pattern of hybridization, 80% of the *D. virilis* clones yield greater than 50 euchromatic hybridization sites, in contrast with only 38% of the *D. melanogaster* clones. This difference is also highly significant ($P < 0.01$).

Discussion

The average size of the insert of *D. virilis* DNA in the P1(pAd10sacBII) library is 65.8 kb. Most of the clones cluster in the size range 40–80 kb, whereas the range 70–100 kb might have been expected. Some of our estimates of insert size could well be conservative, because restriction fragments whose electrophoretic mobility is too similar cannot be distinguished. On the other hand, the average insert size of 65.8 kb is significantly smaller than the average of approximately 81 kb that we have estimated, using the same methods, in a P1(Ad10sacBII) library of *D. melanogaster* (Smoller and Lozovskaya, unpublished). The two libraries are similar in the proportion of clones with inserts smaller than 40 kb – approxi-

mately 7% in the *D. virilis* library and approximately 8% in the *D. melanogaster* library (Smoller and Lozovskaya, unpublished). Several reasons why a library of P1 clones might contain inserts smaller than the expected size range of 70–100 kb have been discussed by Pierce and Sternberg (1991). One reason is related to the efficiency of the size-fractionation of fragments of partially digested genomic DNA in the sucrose gradients. Any preparation that contains a high proportion of small fragments will bias the average insert size because small fragments are most likely to have ends that can be ligated to the vector DNA. A second reason for small inserts is that P1 sometimes encapsulates small heads containing about 45 kb of DNA. Third, if the preparation of the P1 vector contains a population of multimeric molecules, and if digestion with ScaI is incomplete, then this subpopulation of vectors will accept smaller fragments of DNA. Finally, ligation of the ScaI-generated vector ends during the cloning reaction will decrease the size of the insert that can be accepted by the headful packaging mechanism.

The P1(pAd10sacBII) library of *D. virilis* contains clones in which 75% of the inserts are greater than 50 kb and 25% are greater than 80 kb. In order to estimate the representation of the genome, we assume that long tracts of simple-sequence, highly repetitive "satellite" DNA are not present in the library. Direct screening of the *D. virilis* library with probes for such sequences suggests that they are grossly underrepresented, which is consistent with previous findings for both YAC and P1 libraries of *D. melanogaster* (Lohe and Hartl, unpublished). The genome size of *D. virilis* is estimated as 313 Mb, approximately 43% of which consists of satellite DNA (Gall 1973; John and Miklos 1988). Discounting the satellite DNA, the size of the *D. virilis* genome that is potentially clonable is approximately 180 Mb. In a library of 10,080 P1 clones with inserts averaging 65.8 kb, the total amount of DNA is approximately 660 Mb. Therefore, we estimate that the redundancy in coverage of the genome in the P1 library is $660/180 = 3.7$ genome equivalents; that is, each unique sequence should be represented, on average, 3.7 times. Hence, the probability that a particular single-copy sequence will be missing from the P1 library by chance is approximately $e^{-3.70} = 0.02$. In practice, this means that the P1 libraries are 98% complete except for sequences like satellite DNA that are missing for biological reasons. In screening the P1 library with probes obtained by PCR from *D. melanogaster* DNA, we typically screened a total of 3,840 colonies, which amounts to approximately 1.4 genome equivalents. Among the single-copy sequences represented at least once in this subset of clones, the average number of occurrences is expected to be $1.4/[1-\exp(-1.4)] = 1.9$. In the ten single-copy sequences in Fig. 3 for which we recovered at least one clone, the average number was 1.6, which is in reasonable agreement with the expected number in view of the small sample size.

Random in situ hybridizations of over 1,000 *D. melanogaster* P1 clones with *D. melanogaster* salivary gland chromosomes generally indicate that the P1 cloning system gives a broad, nearly comprehensive coverage of the euchromatic genome (Lozovskaya, Jones and Hartl, unpublished). The in situ hybridizations carried out with the *D. virilis* P1 clones are so far consistent with this generalization, but the number of *D. virilis* clones localized to date is too limited to support a statistical analysis for randomness and completeness of coverage. However, the data for specific genes do confirm the homologies between the major chromosome arms inferred from previous studies (Alexander 1976; Gubenko and Evgen'ev 1984; Loukas and Kafatos 1986; Whiting et al. 1989). In particular, the *D. virilis* chromosomes *X, 2, 3, 4,* and *5* are homologous with the *D. melanogaster* chromosome arms *X, 3R, 3L, 2L,* and *2R*, respectively. However, what is particularly apparent for the genes along the *X* chromosome in Fig. 3 is that the order and spacing of the genes is quite different in the two species. The presence of two loci on different chromosomes that hybridize with probes for the *D. melanogaster* histone genes was also expected (Anderson and Lengyel 1983). Although a structural and functional analysis of the histone loci in *D. virilis* is needed, the pair of histone loci in *D. virilis* suggests that there has been transposition of genes between chromosome arms as well as rearrangements within arms. The prevalence of such transpositions in *D. virilis*, relative to *D. melanogaster*, is not known.

In *Drosophila*, two kinds of heterochromatin are distinguished (Miklos and Cotsell 1990). The alpha heterochromatin consists of the pericentromeric regions of the chromosomes, as well as the entire *Y* chromosome; it fails to undergo significant replication during polytenization and forms a small, compact, densely staining mass in each nucleus. The beta heterochromatin spans the region between the alpha heterochromatin and the fully replicated euchromatin; it is generally thought to remain somewhat underreplicated and appears as a diffuse, poorly banded, meshlike mass in the polytene chromosomes. Together, the two types of heterochromatin constitute the chromocenter and the bases of the polytene chromosome arms. Not all chromosomes contain the same amount of beta heterochromatin, and some chromosomes appear to lack beta heterochromatin altogether, as judged from their crisp, clear banding pattern extending all the way to the base where they join with the chromocenter. Judged on this criterion, chromosome *3R* in *D. melanogaster* and chromosomes *3* and *6* in *D. virilis* appear to lack beta heterochromatin (Miklos and Cotsell 1990). The density of genetic complementation groups in alpha heterochromatin is very low; although genes have been identified in beta heterochromatin, the density of genes relative to euchromatin is unknown. These considerations are relevant to P1 clones that hybridize with the chromocenter in salivary gland nuclei. Very different patterns of hybridization are observed among *D. virilis* P1 clones that hybridize with the chromocenter and the bases of the chromosome arms, and different patterns are also observed among such P1 clones from *D. virilis* and those from *D. melanogaster*. These clones may provide an important source of material for reconciling cytogenetic observations with the molecular biology of heterochromatin.

Our results also indicate an unexpected difference in the distribution of putative transposable element in the genomes of *D. virilis* and *D. melanogaster*. In particular, 5.9% of the P1 clones from *D. virilis* show hybridization with multiple euchromatic sites but not with the chromocenter or (usually) chromosome *6*. This type of pattern is rare in *D. melanogaster*, in which most clones with multiple euchromatic hybridization sites also hybridize with the chromocenter. The multiple hybridization sites in the euchromatin, as well as differences in the number and distribution of the sites within and among strains, suggest that the hybridizing DNA sequences are transposable elements. However, direct molecular studies will be necessary to determine the nature of these putative transposable elements, the mechanisms that restrict their distribution to the euchromatin, whether they are similarly distributed in the genomes of other members of the *virilis* species group, and whether similar elements occur in the genome of *D. melanogaster*.

The differences between strains in the distribution and abundance of putative transposable elements is also of interest particularly regarding strain 9 and strain 160. Crosses between females of strain 9 and males of strain 160 produce hybrid progeny with a large number of dysgenic traits including high frequencies of female and male sterility, male recombination, chromosomal nondisjunction, segregation ration distortion, and the appearance of numerous visible mutations in the progeny (Lozovskaya et al. 1990). It seems likely that the identification of transposable elements with grossly different numbers of copies in the two strains will provide probes for determining the molecular basis of this form of hybrid dysgenesis and for investigating its possible role in interspecific hybrid sterility.

The method in Fig. 4 for isolating P1 clones containing genes homologous to those in *D. melanogaster* entails screening the *D. virilis* P1 library with probes obtained by PCR from sequenced genes of *D. melanogaster*. Although the method requires library screening, rather than direct in situ hybridization with P1 clones from *D. melanogaster*, it has a number of advantages. First, it is generally applicable to any gene that has been sequenced in *D. melanogaster*. Secondly, it uses a PCR product as the probe and so does not require cloned material in plasmids or phages. Third, the method identifies multiple P1 clones from *D. virilis* that contain any gene of interest from *D. melanogaster*. The latter is perhaps the most important, because it opens up the possibility of broad DNA sequence comparisons between these species for a direct study of molecular evolution at the level of chromosome structure and organization as well as at the level of individual genes.

Acknowledgements. We express our appreciation to David Smoller for advice in preparation of the library and to Nat Sternberg and James Pierce of DuPont/Merck and to Philip Moen of New England Nuclear for generously providing materials, advice, and encouragement during the phase of library construction. We also thank Allan R. Lohe, W.L. Pak, Linda D. Strausbaugh, and R.A. Voelker for providing probes, and Mike Devine, Bob Jones and Margaret Lackey for technical assistance. This work was supported by grants HG00357 and HG00750 from the National Institutes of Health.

Appendix. Cytological localizations of P1 clones endomatic sites

P1 clone	Cytology	Comments
v68–80***	1D	Also 32A and 2C; ~60 ES; no sites in CC or Ch6
v10–34	1D	Contains yellow gene (*y*)
v68–80***	2C	Also 32A and 1D; ~60 ES; no sites in CC or Ch6
v71–76	2D	
v68–90	4B	
v20–36	4D	Contains suppressor of sable gene [*su(s)*]
v20–45	4D	Contains suppressor of sable gene [*su(s)*]
v32–30	4D	Contains suppressor of sable gene [*su(s)*]
v68–6	5D	
v20–93	7B	Many weak ES
v29–64	7C	Contains singed gene (*sn*)
v37–19	7C	Contains singed gene (*sn*)
v71–78	7C1	Contains singed gene (*sn*)
v71–69	9B	
v33–95	9C	Contains PI-PLC gene (*norpA*)
v71–17	11B1	SI 112 kb
v71–83	11C	
v68–72	12D	
v71–93	12D1–2	
v71–11	13B	SI 58 kb
v71–20	13B	SI 73 kb
v68–44	13BC	
v26–79	13C	Contains white gene (*w*)
v71–5	13D1	SI 113 kb
v71–23	14A2	SI 53 kb
v68–33	14B1	
v71–70	14C1	Many ES; no sites in CC; SI 68 kb
v68–7	15A2	
v45–18	15C	
v71–68	16D2	
v71–47	17D3	SI 43 kb
v14–23	18D1–2	
v71–28	19B	Also CC + Ch6 + weak sites in 23C and 9B; SI 78 kb
v1–33	19C	
v68–43	20C	Also several other weak ES
v71–33	20D1	SI 63 kb
v68–94	21B	Also weak site in 52A
v71–12	21B1–2	SI 68 kb
v71–65	21B2	SI 43 kb; PD v71–72
v71–72	21B2	PD v71–65)
v71–16	21B2–C1	SI 83 kb
v11–07	21C	Also CC + some weak ES; contains hunchback gene (*hb*)
v95–39	21C	Contains hunchback gene (*hb*)
v68–67	21D1–2	
v31–37	21H	
v43–64	21H	
v39–75	21H	
v68–68	22B	
v68–24***	22C	Also 22D and 22E
v68–24***	22D	Also 22C and 22E
v71–29	22E	SI 53 kb
v68–24***	22E	Also 22C and 22D
v88–16	22E	Contains Ribosomal protein 49 gene (*Rp49*)
v81–70	22E	Contains Ribosomal protein 49 gene (*Rp49*)
v71–56	22E3–F1	SI 43 kb; PD v71–63 and/or v71–67
v71–63	22E3–F1	PD v71–56 and/or v71–67

Appendix (continued)

P1 clone	Cytology	Comments
v71–67	22E3–F1	PD v71–56 and/or v71–63
v71–50	22F1	
v68–19	22G	Also many other ES
v71–19	22H3	SI 88 kb
v71–22	22H3–4	SI 43 kb
v71–52	22H3–4	SI 43 kb; PD v71–73
v71–73	22H3–4	PD v71–52
v25–7	23D	
v68–30	23DE	
v68–83	23F	Also ~15 other ES
v68–54	23G	
v12–22	23GH	
v68–15	23H2	
v68–25	24C	
v68–1	24C1–2	
v71–43	24C2	SI 58 kb
v68–49	24E	
v39–62	24H	
v40–45**	25F	Also 43C+many weak ES; contains histone genes
v21–67**	25F	Also 43C+many weak ES; contains histone genes
v71–42	25G	
v32–20	26EF	
v68–78	26F	Also very weak CC+Ch6
v68–71	26FG	
v71–55	27D	SI 43 kb
v71–14	27E2–F	SI 68 kb
v71–57	27F	
v71–62	27F	
v71–80	27F	
v9–25	29B2	
v68–85	29C2	Also light CC+Ch6+perhaps other very weak ES
v36–22	29D	
v71–34	30A2–B1	Also weak 53C3; SI 83 kb
v68–69	31EF	
v68–23	32A	
v68–80***	32A	Also 1D and 2C; ~60 ES; no sites in CC or Ch6
v71–90	32A2–3	
v71–60	32C1–2	SI 68 kb
v68–65	32D	
v29–81	32F	
v40–11	32F	Also ~4–5 other weak sites on chromosome *3*
v40–74	33E	Contains Heat shock protein 82 gene (*Hsp82*)
v50–16	33E	Contains Heat shock protein 82 gene (*Hsp82*)
v71–53	33F3	SI 68 kb
v5–4	34A3	
v68–63	34B	
v71–94	34B2	
v68–46	34C	
v71–39	34C1	SI 83 kb
v71–59	34C2	SI 33 kb
v68–41	34C2	
v71–10	34C3	SI 78 kb
v68–34	34C3	Also weak site at 34B2
v71–26	34D1	SI 53 kb
v71–31	34D1	SI 73 kb
v71–7	34E	Also very weak CC+~65 ES; SI 78 kb
v68–84	34F	Also ~60 ES; none in CC
v68–37	34F1	Also ~50 other weak ES; none in CC

Appendix (continued)

P1 clone	Cytology	Comments
v71–89	35B	Also ~80 ES; weak CC; no sites in Ch6 (Strain 160: one site at 35B)
v68–50	35B	
v71–82	35B1	
v68–3	35B1	
v68–14	35B1	
v68–27	35B1	
v71–20	36A3	SI 53 kb
v68–10	36B1–2	
v68–42	37B	
v47–69	37B	
v68–11	37B3	Some nuclei also show extra site at 38E
v71–85	37B5	
v68–73	37F–38A	
v68–16	38A	
v71–77	38A1–2	
v28–40	38C	
v68–95	38D	
v68–12	38F	Also CC+Ch6+several ES
v71–92	39C	
v35–7	39D	Contains Salivary gland secretion 3 gene (*Sgs3*)
v36–27	39D	Contains Salivary gland secretion 3 gene (*Sgs3*)
v68–75	39F1	Also weak CC+Ch6
v71–46	41B	SI 78 kb
v68–64	41C	Also ~40 other weak ES; none in CC
v68–74	41D	Also very weak CC
v71–71	41D1	SI 68 kb
v68–56	42AB	Also site at ~45D
v71–38	42E3	SI 66 kb
v68–28	43A	Also several weaker sites on chromosome *4*
v68–4	43A1	
v40–45**	43C	Also 25F+many weak ES; contains histone genes
v21–67**	43C	Also 25F+many weak ES; contains histone genes
v68–48	43E	
v71–86	44A	
v68–62	44E	
v68–77	44E	
v65–74	44E	
v68–22	45B	Also CC+Ch6+few other ES
v31–31	45D	
v39–56	45D	
v68–70	46CD	Also weak CC+Ch6+~60 other weak ES
v71–6	47B1–2	SI 73 kb
v71–3	47C	Also light CC+~25 ES; SI 73 kb (Strain 160: different ES)
v68–86	48A	
v68–45	48B	
v68–21	48F	Also ~20 ES; none in CC (Strain 160: different ES)
v93–93	49A2	Contains Pupal cuticular protein gene (*Pcp*)
v71–54	50B4–C1	SI 58 kb
v71–32	50BC	SI 68 kb
v68–47	51A	
v68–92	51D	
v68–58	52F1–2	
v68–29	53A	
v71–79	53D1–2	
v71–44	53E	SI 83 kb
v68–55	54B	

Appendix (continued)

P1 clone	Cytology	Comments
v71–41	54B1	SI 58 kb
v71–8	54E2	SI 108 kb
v68–93	54F	Also ~100 other ES; none in CC
v68–66	56A3	
v68–32	57C	
v71–24	57D3	SI 33 kb
v68–87	57E	
v52–4	57E2	
v29–34	58D	
v68–13	58E	
v68–79	58E	
v71–49	59A2	SI 88 kb
v68–51	59C	
v33–94	59D	Contains mastermind gene (*mam*)
v2–12	59D	Contains mastermind gene (*mam*)
v71–1	59E	SI 83 kb
v68–35	59E2	
v71–2		CC heavy+Ch6+~60 ES; SI 103 kb
v71–4		~80 ES; none in CC; SI 58 kb (Strain 160: one site at 41C)
v71–9		~175 ES; none in CC; SI 123 kb (Strain 160: different ES)
v71–13		CC heavy+~55 ES; SI 68 kb
v71–15		CC heavy+~145 ES; SI 93 kb
v71–21		~85 ES; none in CC; SI 73 kb
v71–25		CC+Ch6+~65 ES; SI 38 kb
v71–27		CC+Ch6+very few ES+possibly tips; SI 63 kb
v71–35		Also 22D1+22F1+22G1+26F (and 1D1 in some nuclei)
v71–36		CC weak+nucleolus (confirmed ribosomal DNA clone)
v71–37		~30 ES; none in CC; SI 75 kb (Strain 160: weak site at 30 AB)
v71–40		CC+Ch6+~25 weak ES; SI 58 kb (Strain 160: similar pattern)
v71–45		CC weak+Ch6+~150 ES (~20 sites in middle of chromosome 5); SI 48 kb
v71–48		CC+Ch6
v71–51		CC heavy+Ch6+many ES; SI 89 kb
v71–58		CC+Ch6+many ES; SI 53 kb
v71–61		~80 ES; none in CC
v71–64		CC+many ES; SI 33 kb
v71–66		CC+many ES; SI 48 kb
v71–75		CC weak+~14 ES
v71–81		CC+Ch6+~90 ES
v71–84		CC+Ch6+>25 ES
v71–87		CC+Ch6
v71–88		CC+Ch6
v71–91		CC+Ch6+~20 weak ES
v71–95		CC+Ch6+~50 ES
v68–2		CC heavy+Ch6+~200 ES
v68–5		Many ES; CC and Ch6 very weak
v68–8		CC heavy+Ch6+~55 ES (Strain 160: different ES)
v68–17		Several sites on chromosome *3*; none on *X*, *2*, *6*, or in CC
v68–18		CC+Ch6+several ES
v68–20		Many ES; no sites in CC or Ch6
v68–26		CC+Ch6
v68–31		CC+Ch6
v68–36		CC+Ch6
v68–38		Nucleolus; confirmed ribosomal DNA
v68–39		CC+Ch6
v68–40		CC+Ch6+single weak sites on *X* and *2*

Appendix (continued)

P1 clone	Cytology	Comments
v68–52		CC+Ch6+~10 weak ES
v68–53		CC+Ch6
v68–59		Many ES; none in CC
v68–60		CC+Ch6
v68–61		CC+Ch6+~60 ES
v68–76		CC heavy+Ch6+~160 ES
v68–81		~50 ES (mainly chromosomes *3* and *4*); no sites in CC or Ch6
v68–82		CC light+Ch6+~160 ES
v68–88		CC heavy+Ch6+many ES
v68–89		CC heavy+Ch6+many weak ES
v68–91		Strong sites on chromosomes *X* and *2*; ~130 on others; none in CC
v68–96		CC heavy+Ch6+~80 other ES
v30–81		Many ES
v28–70		CC+heavy Ch6+some ES
v29–1		CC+Ch6+mainly heterochromatin of chromosomes *X* and *5*

A few P1 clones have major sites of hybridization at two, or occasionally three, enchromatic sites (ES); these are denoted with two, or three, asterisks and listed separately at each site of hybridization. The comments pertain to cytological observations in *Drosophila virilis* strain 9; in some cases additional information is provided relative to strain 160. Symbols used in the tabulation are as follows: CC, chromocenter; Ch6, chromosome *6*; SI, estimated size of insert in P1 clone; PD, possible duplicate of another clone(s)

References

Ajioka JW, Smoller DA, Jones RW, Carulli JP, Vellek AEC, Garza D, Link AJ, Duncan IW, Hartl DL (1991) Drosophila genome project: One-hit coverage in yeast artificial chromosomes. Chromosoma 100:495–509

Alexander ML (1976) The genetics of *Drosophila virilis*: In: Ashburner M, Novitski E (eds) The genetics and biology of *Drosophila*. Academic Press, New York, pp 1365–1427

Anderson KV, Lengyel JA (1983) Histone gene expression in *Drosophila* development: multiple levels of gene regulation. In: Stein GS, Stein JL, Marzluff WF (eds) Histone genes. John Wiley & Sons, New York, pp 135–159

Ashburner M, Carson HL, Thompson JN Jr (1981) The genetics and biology of *Drosophila*, vol 3a. Academic Press, New York

Ashburner M, Carson HL, Thompson JN Jr (1982) The genetics and biology of *Drosophila*, vol 3b. Academic Press, New York

Ashburner M, Carson HL, Thompson JN Jr (1986) The genetics and biology of *Drosophila*, vol 3e. Academic Press, New York

Beverley SM, Wilson AC (1984) Molecular evolution in *Drosophila* and the higher diptera II. Time scale for fly evolution. J Mol Evol 21:1–13

Bingham PM, Levis R, Rubin GM (1981) Cloning of DNA sequences from the white locus of *Drosophila melanogaster* by a novel and general method. Cell 25:693–704

Birnboim HC, Doly J (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res 7:1513–1523

Bloomquist BT, Shortridge RD, Schneuwly S, Perdew M, Montell C, Steller H, Rubin G, Pak WL (1988) Isolation of a putative phospholipase C gene of *Drosophila*, norpA, and its role in phototransduction. Cell 54:723–733

Bray SJ, Hirsh J (1986) The *Drosophila virilis* dopa decarboxylase gene is developmentally regulated when integrated into *Drosophila melanogaster*. EMBO J 5:2305–2311

Brownstein BH, Silverman GA, Little RD, Burke DT, Korsmeyer SJ, Schlessinger D, Olson MV (1989) Isolation of single-copy human genes from a library of yeast artificial chromosome clones. Science 244:1348–1351
Colot HV, Hall JC, Rosbash M (1988) Interspecific comparison of the period gene of *Drosophila* reveals large blocks of non-conserved coding DNA. EMBO J 7:3929–3927
Feinberg AP, Vogelstein B (1984) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity: Addendum. Anal Biochem 137:266–267
Gall JG (1973) Repetitive DNA in *Drosophila*. In: Hamkalo BA, Papaconstantinou J (eds) Molecular cytogenetics. Plenum, New York, pp 59–73
Gubenko IS, Evgen'ev MB (1984) Cytological and linkage maps of *Drosophila virilis* chromosomes. Genetica 65:127–139
Hartl DL (1992) Genome map of *Drosophila melanogaster* based on yeast artificial chromosomes. In: Davies KE, Tilghman SM (eds) Genome analysis, vol 4: Strategies for physical mapping. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp 39–69
Hartl DL, Lozovskaya ER (1992) The *Drosophila* genome project: Current status of the physical map. Comp Biochem Physiol [B] 103:1–8
Hartl DL, Ajioka JW, Cai H, Lohe AR, Lozovskaya ER, Smoller DA, Duncan IW (1992) Towards a *Drosophila* genome map. Trends Genet 8:70–75
Heberlein U, Rubin GM (1990) Structural and functional comparisons of the *Drosophila virilis* and *Drosophila melanogaster* rough genes. Proc Natl Acad Sci USA 87:5916–5920
Hooper JE, Pérez-Alonzo M, Bermingham JR, Prout M, Rocklein BA, Wagenbach M, Edström J-E, de Frutos R, Scott MP (1992) Comparative studies of *Drosophila* Antennapedia genes. Genetics 132:453–469
John B, Miklos GLG (1988) The eukaryotic genome in development and evolution. Allen and Unwin, London
Kassis JA, Poole SJ, Wright DK, O'Farrell PH (1986) Sequence conservation in the protein coding and intron regions of the engrailed transcription unit. EMBO J 5:3583–3589
Laird CD (1973) DNA of *Drosophila* chromosomes. Annu Rev Genet 7:177–204
Langer-Safer PR, Levine M, Ward DC (1982) Immunological method for mapping genes on *Drosophila* polytene chromosomes. Proc Natl Acad Sci USA 79:4381–4385
Lohe AR, Brutlag DL (1986) Multiplicity of satellite DNA sequences in *Drosophila melanogaster*. Proc Natl Acad Sci USA 83:696–700
Lohe AR, Roberts PA (1990) An unusual Y chromosome of *Drosophila simulans* carrying amplified rDNA spacer without rRNA genes. Genetics 125:399–406
Loukas M, Kafatos FC (1986) The actin loci in the genus *Drosophila*: Establishment of chromosomal homologies among distantly related species by in situ hybridization. Chromosoma 94:297–308
Lozovskaya ER, Scheinker VS, Evgen'ev MB (1990) A hybrid dysgenesis syndrome ein *Drosophila virilis*. Genetics 126:619–623
Merriam J, Ashburner M, Hartl DL, Kafatos FC (1991) Toward cloning and mapping the genome of *Drosophila*. Science 254:221–225
Miklos GLG (1985) Localized highly repetitive DNA sequences in vertebrate and invertebrate genomes. In: MacIntyre RJ (ed) Molecular evolutionary genetics. Plenum, New York, pp 241–321
Miklos GLG, Cotsell JN (1990) Chromosome structure at interfaces between major chromatin types: alpha- and beta-heterochromatin. BioEssays 12:1–6
Monod J (1970) Le hasard et la necessité. Seuil, Paris
Neufeld TP, Carthew RW, Rubin GM (1988) Evolution of gene position: Chromosomal arrangement and sequence comparison of the *Drosophila melanogaster* and *Drosophila virilis* sina and Rh4 genes. Proc Natl Acad Sci USA 88:10203–10207
O'Neil MT, Belote JM (1992) Interspecific comparison of the transformer gene of *Drosophila* reveals an unusually high degree of evolutionary divergence. Genetics 131:113–128
Orr HA, Coyne JA (1989) The genetics of postzygotic isolation in the *Drosophila virilis* group. Genetics 121:527–537
Patterson JT, Stone WS (1952) Evolution in the genus *Drosophila*. Macmillan, New York
Pierce JC, Sternberg NL (1993) Using the bacteriophage P1 system to clone high molecular weight (HMW) genomic DNA. Methods Enzymol (in press)
Pierce JC, Sauer B, Sternberg N (1992) A positive selection vector for cloning high molecular weight DNA by the bacteriophage P1 system: Improved cloning efficacy. Proc Natl Acad Sci USA 89:2056–2060
Saiki RK, Scharf SJ, Faloona F, Mullis KB, Horn GT, Erlich HA, Arnheim N (1985) Enzymatic amplification of β-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science 230:1350–1354
Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: A laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York
Scheinker V, Lozovskaya E, Bishop JG, Corces VG, Evgen'ev MB (1990) A long terminal repeat-containing retrotransposon is mobilized during hybrid dysgenesis in *D. virilis*. Proc Natl Acad Sci USA 87:9615–9619
Segarra C, Aguadé M (1992) Molecular organization of the X chromosome in different species of the obscura group of *Drosophila*. Genetics 130:513–521
Smoller DA, Petrov D, Hartl DL (1991) Characterization of bacteriophage P1 library containing inserts of *Drosophila* DNA of 75–100 kilobase pairs. Chromosoma 100:487–494
Southern EM (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. J Mol Biol 98:503–517
Sternberg N (1990) Bacteriophage P1 cloning system for the isolation, amplification, and recovery of DNA fragments as large as 100 kilobase pairs. Proc Natl Acad Sci USA 87:103–107
Sternberg NL (1992) Cloning high molecular weight DNA fragments by the bacteriophage P1 sylstem. Trends Genet 8:11–16
Stone WS (1962) The dominance of natural selection and the reality of superspecies (species groups) in the evolution of *Drosophila*. Univ Texas Publ 6205:507–537
Sturtevant AH (1916) Notes on North American Drosophilidae with descriptions of twenty-three new species. Ann Ent Soc Am 9:323–343
Throckmorton LH (1982) The *virilis* species group. In: Ashburner M, Carson HL, Thompson JN Jr (eds) The genetics and biology of *Drosophila*, vol 3b. Academic Press, New York, pp 227–296
Vollrath D, Davis RW (1987) Resolution of DNA molecules greater than 5 megabases by contour-clamped homogeneous electric fields. Nucleic Acids Res 15:7865–7876
Watson JD (1990) The human genome project: Past, present, and future. Science 248:44–49
Whiting JH Jr, Pliley MD, Farmer JL, Jeffery DE (1989) In situ hybridization analysis of chromosomal homologies in *Drosophila melanogaster* and *Drosophila virilis*. Genetics 122:99–109