Genet. Res., Camb. (2000), **75**, *pp.* 25–35. Printed in the United Kingdom © 2000 Cambridge University Press

Microsatellite variation in populations of *Drosophila pseudoobscura* and *Drosophila persimilis*

MOHAMED A. F. NOOR*, MALCOLM D. SCHUG AND CHARLES F. AQUADRO
Section of Genetics and Development, Cornell University, Ithaca, New York, USA

(*Received 3 December 1998 and in revised form 3 January 1999*)

Summary

We have isolated, characterized and mapped 33 dinucleotide, three trinucleotide and one tetranucleotide repeat loci from the four major chromosomes of *Drosophila pseudoobscura*. Average inferred repeat unit length of the dinucleotide repeats is 12 repeat units, similar to *D. melanogaster*. Assays of *D. pseudoobscura* and populations of its sibling species, *D. persimilis*, using 10 of these loci show extremely high levels of variation compared with similar studies of dinucleotide repeat variation in *D. melanogaster* populations. The high levels of variation are consistent with an average mutation rate of approximately 10^{-6} per locus per generation and an effective population size of *D. pseudoobscura* approximately four times larger than that of *D. melanogaster*. Consistent with allozymes and nucleotide sequence polymorphism, the dinucleotide repeat loci reveal minimal structure across four populations of *D. pseudoobscura*. Finally, our preliminary recombinational mapping of 24 of these microsatellites suggests that the total recombinational genome size may be larger than previously inferred using morphological mutant markers.

1. Introduction

Microsatellites have only recently become popular genetic markers in *Drosophila* despite their widespread use in other organisms for nearly a decade. These di-, tri- and tetranucleotide repeat arrays are co-dominant and often highly variable in repeat number, making them relatively simple and economical genetic markers for polymorphism, differentiation or quantitative trait locus (QTL) mapping studies. An archive of these markers has been assembled in *Drosophila melanogaster* and the closely related species within its clade, and these markers have been used successfully to study population differentiation, to survey genome-wide patterns of variation, and to identify evidence of the direct action of natural selection (e.g. Schlotterer *et al*., 1997; Irvin *et al*., 1998; Schug *et al*., 1998*a*). Nonetheless, few microsatellites have been identified and characterized in more distantly related *Drosophila* species.

Drosophila pseudoobscura has been used extensively in evolutionary genetic investigations and mapping of various phenotypic traits, particularly those associated with reproductive isolation (e.g. Tan, 1946; Orr, 1987; Noor, 1997). This species is ideal for evolutionary studies because of its large effective population size and relative lack of structure (e.g. Prakash *et al*., 1972; Schaeffer & Miller, 1992) as well as its close relationship to a sibling species (*D. persimilis*) and evidence of gene exchange between them (e.g. Wang *et al*., 1997). Although crosses can be made between the two sibling species, some of these studies have encountered significant problems due to the low availability of morphological mutant markers and, especially, the high levels of inviability associated with most of the available morphological markers (Noor, 1997). Further, as *D. pseudoobscura* is rather distantly related to *D. melanogaster*, understanding properties of microsatellites in this species can lead to a broader understanding of microsatellite evolution in general.

In this study, we present a database of 37 microsatellites in *Drosophila pseudoobscura* identified by three methods: (1) our genomic library screenings for dinucleotide repeat arrays, (2) searches of GenBank, the international sequence database, for di-, tri- or tetranucleotide repeat arrays, and (3) direct sequencing studies by other investigators. These microsatellites have been developed to aid our planned QTL mapping studies in *D. pseudoobscura* and its

* Corresponding author. Present address: Department of Biological Sciences, 138 Life Sciences Building, Louisiana State University, Baton Rouge, LA 70803, USA.

sibling species, *D. persimilis*. We have recombinationally mapped 24 of these microsatellites relative to each other and relative to available morphological mutant markers. The resulting map demonstrates that microsatellites are distributed across the genome of this species and suggests that the total recombinational length of the *D. pseudoobscura* genome may be larger than previously inferred from morphological mutant marker recombinational distances. Our data show that microsatellites are generally more variable in *D. pseudoobscura* than in *D. melanogaster*, possibly due to the larger effective population size of the former. We also present a preliminary study of population structure for four *D. pseudoobscura* populations using 10 of the microsatellites characterized in this study.

2. Materials and methods

(i) *Isolation of microsatellite sequences*

Preparation and screening of a subgenomic DNA library followed the same general procedure as Schug *et al*. (1998*b*). Briefly, CsCl-prepared genomic DNA from several California lines of *D. pseudoobscura* was generously provided by Stephen Schaeffer (The Pennsylvania State University). DNA was partially digested with *Sau3*AI. DNA fragments between 250 and 700 bp in length were extracted from the gel using glass beads (Qiagen), treated with calf alkaline phosphatase, and cloned into the *Bam*HI site of pUC18. Ligations were purified and transformed into DH5α maximum efficiency competent *E. coli* cells (GIBCO BRL). The cells were plated on LB in 150 mm Petri dishes and lifted onto nylon membranes. Denaturation of plasmid DNA and hybridizations were carried out as described by Schug *et al*. (1998*b*), with 75 ng each of $(AC)_{15}$, $(AG)_{15}$, $(AT)_{15}$ and $(CG)_{15}$ oligonucleotides end-labelled with [^{32}P]γ-dATP. Hybridization to colony lifts was performed with TEMAC hybridization solution overnight at 65 °C. Washes were as described previously (Schug *et al*., 1998*b*) with a single high-temperature wash at 65 °C. Nylon membranes were exposed to X-Omat AR film (Kodak) overnight at −70 °C. Positive colonies were picked and plated at low density for secondary screening. Inserts were amplified using universal pUC18 forward and reverse primers and cycle sequenced using a Thermo-Sequenase kit (Amersham).

(ii) *Fly stocks*

Fifty-one inbred *D. pseudoobscura* lines were used in our population survey. We collected flies in 1996 and established 15 isofemale lines from Flagstaff, Arizona; seven from the Abajo Mountains, immediately west of Monticello, Utah; 14 lines from Goldendale, Washington; and 15 lines from Cheney, Washington, Goldendale and Cheney are approximately 300 km apart, as are the Abajo Mountains and Flagstaff. Hence, these populations represent two southeastern samples and two northwestern samples from within the geographic range of *D. pseudoobscura*. Flagstaff and Abajo natural populations are nearly fixed for the third chromosome inversion 'Arrowhead' (Anderson *et al*., 1991). In contrast, west coast populations have a large number of different inversions at intermediate frequencies. Since this inversion polymorphism is very old and the products of recombination are inviable in inversion heterozygotes, large differences are expected at all third chromosome microsatellites between the western and eastern populations.

We collected *D. persimilis* from two populations in California in 1997: 15 lines from Mather and seven lines from Mount St Helena. First-generation offspring were frozen and used for microsatellite analyses. Thus, 44 wild chromosomes were analysed.

(iii) *Microsatellite assay conditions*

Primers were identified in the sequences flanking the microsatellite repeat unit using Primer version 3·0 (URL http://www.genome.wi.mit.edu//cgi-bin/primer/primer3.cgi) such that PCR fragments were approximately 100–300 bp in length. Fly DNA was extracted from single-fly squish preparations (Gloor *et al*., 1993) of each line. After optimization, one primer was end-labelled, and PCR was performed in a 10 μl reaction volume with 0·5 μM of each primer, 200 μM dNTPs, 1 μl 10 × buffer (100 mM Tris pH 8·3, 500 mM KCl, 15 mM $MgCl_2$), 1 U *Taq* polymerase, and 1 μl from a 50 μl single fly squish preparation. A single soak at 95 °C for 5 min was followed by 35–40 cycles of 1 min at 94 °C, 1 min at the annealing temperature (Table 1), 1 min at 72 °C. Following PCR, 10 μl of formamide loading dye was added to each reaction, and the fragments were separated by electrophoresis on a 0·8 × glycerol tolerant sequencing gel at 70 W for approximately 2 h. A pUC18 DNA sequencing reaction was included adjacent to the PCR products as a size standard. Dried gels were exposed to BIOMAX film (Kodak) for 1–2 days.

(iv) *Chromosomal localization*

Microsatellites were localized in *D. pseudoobscura* by crossing a male from one of the Goldendale isofemale lines to several females from an inbred marked stock possessing morphological mutations on three of the four major chromosomes: *yellow* on the X chromosome, *glass* on the second and *orange* on the third. The F1 males were then backcrossed to females from the mutant stock. Since there is no recombination in male *Drosophila*, the microsatellite allele of the mutant stock will always segregate with the visible mutation on that chromosome. Fly DNA was extracted by single-fly squish preparations from the original male

https://doi.org/10.1017/S0016672399004024 Published online by Cambridge University Press

and female, the F1 males and females ($n = 6$), and the backcross offspring ($n = 12$), and their alleles at the microsatellite loci were identified as above.

(v) *Localization within chromosomes*

Most microsatellites were recombinationally mapped within chromosomes after being localized to a particular chromosome. F1 females of the crosses between the marker stock and Goldendale lines 41 (third chromosome inversion Treeline: cross no. 1) and 98 (third chromosome inversion Standard: cross no. 2) were then backcrossed to males from the marker stock (third chromosome inversion Standard). Between 200 and 300 offspring were then frozen and prepared using single-fly squish preparations as above. We evaluated the recombinational distance between the markers on a particular chromosome using the proportion of crossovers between each pair of two markers and Haldane's (1919) mapping function:

$$x = -1/2 \ln (1 - 2 * \theta),$$

where x is the calculated distance in centrimorgans (cM) and θ is the observed fraction of crossovers. This measure of recombinational distance was superior to the crossover fraction itself (Morgan's mapping function) in this study because of the rather large distances between some of the markers we investigated. Additionally, we present the map distances using Kosambi's (1944) mapping function, which takes chiasmal interference into account:

$$x = 1/2 \arctanh (2 * \theta).$$

Along the second chromosome, we were unable to map all the available microsatellites from a given cross, so we have combined data from two crosses in the results presented. Each recombinational fraction and map distance is marked with respect to the cross that was used to evaluate it.

We did not map microsatellites on the third chromosome because of complications resulting from the inversion polymorphism in this species. A later study may present the recombinational fractions between the third chromosome microsatellites in a variety of gene arrangements. In addition, we have been unable to map these microsatellites to salivary gland chromosome bands because of the very small sizes of the DNA fragments extracted, but efforts are continuing in that regard.

(vi) *Analysis of population variation*

The PCR product lengths were translated into numbers of repeats by comparison of total fragment size with the total fragment size of the original sequenced region. The assayed microsatellite allele was assumed to differ from the original sequence in the number of repeats within the longest perfect repeat array, and this estimated number of repeats within the alleles is presented in the tables. At some loci, some allele sizes were too small to be possible solely as a result of variation in the dinucleotide repeat. For these microsatellites, we assumed that the smallest size observed possessed an allele completely lacking the dinucleotide repeat, and we estimated the number of repeats for the remaining alleles relative to this size.

For the *D. pseudoobscura* population variation survey, one allele was chosen at random from each heterozygous individual to control for inbreeding within lines. Since no inbreeding occurred before the *D. persimilis* lines were assayed, two alleles were scored for each individual. Expected heterozygosity was calculated as $H = (n/n-1)(1-\Sigma p_i^2)$, where n is the number of alleles scored and p_i is the frequency of the ith allele. Mean repeat number and variance in repeat number were calculated using StatView (Abacus Concepts). F_{ST} for *D. pseudoobscura* was calculated and tested for significance using FSTAT (Goudet, 1995). We do not report such statistics as R_{ST} and $(\delta\mu)^2$ because they are not appropriate for closely related populations (Goldstein *et al.*, 1995; Slatkin, 1995) such as the populations of *D. pseudoobscura* included in our study (Prakash *et al.*, 1969; Schaeffer & Miller, 1992; Jenkins *et al.*, 1996).

3. Results

(i) *Characterization of microsatellites*

Our library screen for dinucleotide repeats yielded 35 positive clones, of which 28 had dinucleotide repeat arrays far enough from the fragment ends to allow for usable primers. All positive clones contained repeated units ranging from 4 to 18 uninterrupted repeat units (average 11·7 repeat units). Twenty-six of the repeat units were $(AC)_n$ and two were $(AG)_n$ (Table 1). This length is comparable to that observed in *D. melanogaster* microsatellites derived using similar DNA library screening techniques (Schug *et al.*, 1998*b*). We do not present the frequency of microsatellites in the *D. pseudoobscura* genome because our partial genomic DNA digest may have been closer to a complete digest than desired. Microsatellites were named 'DPS', followed by the chromosome and a three-digit number. Many of the clones also contained tri-, tetra and pentanucleotide repeat units near the dinucleotide repeat. This association may confound comparisons that rely on estimated numbers of repeats from fragment lengths (Schug *et al.*, 1998*a*; but see Section 4). In addition to these dinucleotide repeats that we isolated, seven microsatellites were identified by searching GenBank, one was provided by Martha Hamblin and one was provided by Steve Schaeffer.

The 37 microsatellites characterized in this study were distributed across four chromosomes (Table 1).

https://doi.org/10.1017/S0016672399004024 Published online by Cambridge University Press

Table 1. *Microsatellites identified in* Drosophila pseudoobscura

Name	Chromosome	Size	Annealing temperature (°C)	Primers	Repeat motif	Genbank Accession
DPSX001	X	200	61·0	gaatctctctcctgttgcgg ccacactcgctttcccata	$(ac)_7ag(ac)_7$	AF157568
DPSX002	X	150	61·0	attcttgtcgctctgttggc tcagctgcgtaacaatctgg	$(gt)_{15}$	AF157569
DPSX003	X	200	61·0	gcctacagtgagagctgcct tggggagtggacttatctcg	$(gt)_5at(gt)_7$	AF157570
DPSX004	X	100	58·0	aagtacttcattttgtcttgg cgtgcgcgcttataattctt	$(gt)_{13}$	AF157571
DPSX005	X	250	56·0	acggcaacggtacttgaatc gttttgattccaggcgtgat	$(ca)_{16}$	AF157572
DPSX006	X	200	56·0	agccagctctgtggtctgtt aaaacggtttcattgttgcc	$(gt)_{13}$	AF157573
DPSX007	X	100	56·5	cactcgaggttattgaacgg aatctatggcgggttctaag	$(ca)_5cg(ca)_5$	AF157574
DPSX008	X	150	56·0	ccacagcgtagtgagcagat tttccttctgtgtgttggca	$(ac)_{18}$	AF157575
DPSX009	X	250	56·0	tcaggaaaagaacagcagca cgccacagcaaatcaactta	$(tg)_3g(tg)_3g(tg)_3$	AF157576
DPSX010	X	100	56·0	aaaaggccttattgtagttg agagattctcacccaccatg	$(tg)_8$	AF157577
runt	XL	250	53·8	ccctgccacaagtaacaagc agacaaaaggggcaggtatc	$(gt)_4gc(gt)_9$	U22357
E74A	XR	200	55·5	agagacagctcctgctcctg actcgggccgattttagttt	$(aac)_7$	S73515
DPS2001	2	200	61·0	caaagacagagccaaagcct tgggcattaaagtgcaatca	$(ac)_{15}$	AF157578
DPS2002	2	150	56·0	acatccgcatccacatacg cgtcctgccaaagtgtttct	$(tg)_3tt(tg)_9$	AF157579
DPS2003	2	200	55·0	catttcaagcagaagacgca cctcgggtattatttcgggt	$(ca)_{13}$	AF157580
DPS2004	2	200	56·0	ggtacccaaagccaatctca acgtcctgttgaaagccact	$(ca)_{11}$	AF157581
DPS2005	2	200	56·0	attgattggggctacgtgtc gctaacccaatgatgaggga	$(ct)_5cc(ct)_{14}$	AF157582
DPS2006	2	200	56·0	tttatcatgtgcccgagtga tcgctttaactcgtttcgct	$(ac)_{12}$	AF157583
DPS2007	2	200	56·0	tgcggagagagtttgtgaga gaactacagccagcgagagg	$(tg)_{15}$	AF157584
Bcd	2	200	59·0	ccaggctcagggccagcgcc gcatctgatgcggcacgtgg	$(cag)_4$	X55735
Gld	2	150	57·0	ttcacacccctgagcacaag gtcttcattgctgccgttgc	$(ctga)_4$	M29299, X07359
Mlc	2	100	58·0	caacagaatgttgccaacagc acgcctttgggatctcgaac	$(gt)_{12}$	L08052
Rh1	2	200	56·0	ggcaaccaccagcgaggccg gcttttagatattggaggcaag	$(caa)_5$	X65877
trop1	2	200	58·0	gattaccttgttcttatgtggc cgagattgatgatatttggcag	$(ct)_3cg(ct)_6$	AF039274
DPS3001	3	150	59·0	gggaaaccataagaaaatgcc gtacatgaatcggctacggg	$(ca)_{13}$	AF157585
DPS3002	3	300	61·0	gagtccccaaaatccgaaac cccacaacggacagaaaaat	$(gt)_{17}$	AF157586, AF157587
DPS3003	3	250	58·0	ggcccgaaaataaaacaaca ctgcactctctttccccctt	$(gt)_7g(gt)_5$	AF157588
DPS3004	3	250	56·0	tgaacgtggtgggtgtagaa gtgacaaagaggaggtccca	$(ga)_{13}$	AF157589
DPS3005	3	250	56·0	ggtccagaataaatgcccaa aactgcattgccaaacacaa	$(gt)_4ct(gt)_9$	AF157590
DPS3006	3	150	56·0	caagtacggcaaggatttgg tgttgcctacacatttccca	$(tg)_{12}$	AF157591

https://doi.org/10.1017/S0016672399004024 Published online by Cambridge University Press

Table 1. (*cont.*)

Name	Chromosome	Size	Annealing temperature (°C)	Primers	Repeat motif	Genbank Accession
DPS3007	3	200	56·0	ttaagcagatgggggatgag tttgcaagggcactaaaagc	$(tg)_{12}$	AF157592
DPS3008	3	200	58·0	ggatgattgaagggctgaca ttgataaattgccccacaca	$(gt)_{14}$	AF157593
engrailed	3	150	56·0	ccttctccagcgagcaat tgtaaatattttggtgcaaatatga	$(ac)_{13}$	–
DPS4001	4	300	61·0	gtctgctgcgattaaaagcc cggcaggcggtataaaaata	$(ca)_{12}$	AF157594
DPS4002	4	100	61·0	taccgtatgcaacccagctt cggaatgcactctgctgata	$(ac)_{16}$	–
DPS4003	4	250	61·0	ttctgtccgctgcagccctc tatcaagccatcttctgcac	$(gt)_{11}$	–
dpp	4	150	54·0	ctgatgttgcagagcacgat tctttctttttcctcgtcgc	$(ac)_{8}$	U63856

Table 2. *Drosophila pseudoobscura microsatellite and morphological marker recombinational distances calculated using Haldane and Kosambi mapping functions*

Markers	Haldane	Kosambi	Recombination	Cross
X chromosome				
X009–X006	19·7 cM	16·8 cM	41/252	1
X006–[*y*]	34·5 cM	27·3 cM	63/253	1
[*y*]*–X003*	4·4 cM	4·2 cM	10/240	1
X003–runt	23·2 cM	19·5 cM	23/124	1
runt–X002	16·7 cM	14·6 cM	18/127	1
X002–X001	0·5 cM	0·5 cM	1/216	1
X001–X004	10·4 cM	9·5 cM	19/202	1
X004–X008	10·8 cM	9·9 cM	23/236	1
E74A–X005	2·8 cM	2·7 cM	2/73	1
Unlinked: X007, X010				
Chromosome 2				
2005–2001	23·7 cM	19·9 cM	31/164	1
2001–2007	15·3 cM	13·5 cM	26/197	1
2007–2002	24·9 cM	20·7 cM	48/245	1
2002–[*gl*]	10·6 cM	9·6 cM	24/252	1
[*gl*]*–bcd*	57·8 cM	41·9 cM	74/216	2
bcd–gld	51·3 cM	38·0 cM	42/131	2
gld–rh1	0·0 cM	0·0 cM	0/130	2
rh1–2003	20·3 cM	17·3 cM	40/240	1
Chromosome 4				
dpp–4001	14·9 cM	13·2 cM	26/202	2
4001–4003	49·4 cM	36·9 cM	65/207	2
Unlinked: 4002				

The four microsatellites denoted as chromosome 4 were autosomal loci that did not segregate with either the second or third chromosome morphological marker but did segregate with each other. Since the fifth chromosome is a dot chromosome with probably little or no recombination, and since the total recombinational span of these from autosomal loci covers nearly 100 cM (see below), it is likely that these microsatellites are distributed along the fourth chromosome. Additionally, *dpp* is expected to be on chromosome 4 by Muller's chromosome arm homology with *D. melanogaster* (Richter *et al*., 1997).

The number of dinucleotide repeats appears to be homogeneously distributed among the *D. pseudoobscura* chromosomes. In our library screening, we found 1·4 times as many dinucleotide repeats on the X

https://doi.org/10.1017/S0016672399004024 Published online by Cambridge University Press

Table 3. *Comparison of microsatellites in* Drosophila pseudoobscura *and* D. persimilis

	D. pseudoobscura				*D. persimilis*		
Microsatellite	H^a	No. of alleles[b]	Mean[c]	F_{ST}	H^a	No. of alleles[b]	Mean[c]
DPSX001	0·920	16/44	5·3	0·01	0·895	11/44	10·5
DPSX002	0·718	8/50	12·5	0·09*	0	1/44	7·0
DPSX003	0·864	15/47	10·7	0·02*	0·896	15/44	6·9
DPSX004	0·859	11/43	13·4	0·03	0·874	9/33	14·3
DPS2001	0·771	10/45	20·4	−0·02	0·365	6/44	16·5
DPS3001	0·600	10/45	9·8	0·13**	0·890	12/44	10·0
DPS3002	0·924	18/40	9·3	0·01	0·958	20/44	7·2
DPS3003	0·939	23/42	9·9	0·04*	0·922	16/44	10·0
DPS4001	0·963	23/49	12·5	0·01	0·896	11/44	12·9
DPS4002	0·834	8/47	7·4	0·08*	0·280	3/44	3·3
Mean	0·84		10·3	0·04***	0·70		8·1

* $P < 0·05$; ** $P < 0·01$; *** $P < 0·001$.
[a] Expected heterozygosity.
[b] Number of length alleles observed per number of chromosomes assayed.
[c] Mean inferred number of repeats.

chromosome ($n = 10$) as on the second chromosome ($n = 7$), 1·25 times as many on the third ($n = 8$) and 3·3 times as many microsatellites on the fourth ($n = 3$). These comparisons are necessarily very rough due to our small sample size, but are consistent with the relative sizes of the *D. pseudoobscura* chromosomes (Anderson, 1993).

Previous genetic maps of *Drosophila pseudoobscura* gave total recombination lengths of 228 cM, 101 cM and 69 cM for the X (Orr, 1995), second (Anderson, 1993) and fourth chromosomes (Anderson, 1993), respectively. Orr's (1995) map used the Kosambi mapping function for calculating recombinational distances, while Anderson's (1993) map used uncorrected recombination fractions (Orr, 1995). The total map lengths derived from our current study suggest that the genetic maps may be larger (Table 2). If we conservatively assume that the unlinked loci are only 50 cM from the mapped loci, then we obtain minimum recombination lengths of 255 cM, 160·9 cM (all microsatellites linked) and 100·1 cM for the X, second and fourth chromosomes using Kosambi's mapping function. These estimates are even larger if determined using Haldane's mapping function and are only marginally lower using an uncorrected recombination fraction.

(ii) *Genetic variation within and between species*

(*a*) *D. pseudoobscura*. The *D. pseudoobscura* average microsatellite heterozygosities ranged from 0·60 to 0·96 per locus (mean 0·84; Table 3). Levels of heterozygosity and variances in repeat number were high across all individual populations of *D. pseudoobscura* at all loci except *DPS3001* (Table 4). Eastern populations are nearly fixed for the Arrowhead inversion type, whereas western populations are highly polymorphic (Anderson *et al*., 1991). Since levels of variation at *DPS3001* are reduced in the eastern Utah and Arizona populations (Table 4), it is very likely that this locus is within the Arrowhead inversion, and the reduced levels of variation reflect segregating polymorphism at Arrowhead, rather than a local selective sweep near the *DPS3001* locus.

Several studies of flies and other organisms have shown that levels of heterozygosity are closely associated with the repeat unit length of a microsatellite locus (e.g. Weber, 1990; Goldstein & Clark, 1995; Schug *et al*., 1998*c*). Within *D. pseudoobscura*, we observed no significant association between microsatellite mean or maximum repeat number and heterozygosity or variance in repeat number. Excluding the third chromosome loci that may be within inversion polymorphisms does not change this result. However, such an association may have been obscured by the presence of variable tri-, tetra- or pentanucleotide repeat arrays within the same PCR product as our dinucleotide repeats. Further, we have examined only 10 loci. It may be premature to conclude that dinucleotide array length is not associated with variability in *D. pseudoobscura*.

The combined data from all the loci shows that structure among the four populations of *D. pseudoobscura* is significant, but low ($F_{ST} = 0·041$, $P < 0·001$; Table 3). Five of the individual loci showed significant population structure. Four of these five loci either possess 10 or fewer alleles or were on the third chromosome. The third chromosome inversion poly-

https://doi.org/10.1017/S0016672399004024 Published online by Cambridge University Press

Table 4. *Microsatellite length diversity in* D. pseudoobscura *and D. persimilis*

	DPSX001	*DPSX002*	*DPSX003*	*DPSX004*	*DPS2001*	*DPS3001*	*DPS3002*	*DPS3003*	*DPS4001*	*DPS4002*
					D. pseudoobscura					
Cheney, *Washington* (*NW*)										
Heterozygosity	0·927	0·538	0·891	0·909	0.849	0.834	0·895	0·794	0·956	0·858
No. of alleles/no. chr.[a]	8/11	3/14	8/14	7/11	6/12	5/9	8/12	7/13	11/14	6/13
Mean repeat no.[b]	5·4	11·6	10·4	13·8	20·3	9·7	9·3	10·9	10·9	6·5
Variance repeat no.[c]	9·8	0·6	9·6	3·1	5·7	4·2	7·7	20·5	9·1	5·8
Goldendale, *Washington* (*NW*)										
Heterozygosity	0·857	0·670	0·846	0·845	0·794	0·857	0·934	0·939	0·990	0·712
No. of alleles/no. chr.	11/14	4/14	7/14	6/14	6/13	7/14	5/6	9/12	13/14	4/12
Mean repeat no.	5·6	12·4	10·4	12·8	20·3	9·2	9·8	7·4	12·1	7·1
Variance repeat no.	7·3	1·2	8·6	6·7	2·9	2·4	14·0	24·0	24·2	1·7
Abajo, *Utah* (*SE*)										
Heterozygosity	1·000	0·523	0·900	0·833	0·866	0·000	1·000	1·000	0·952	0·808
No. of alleles/no. chr.	5/5	3/7	4/5	3/4	4/6	1/7	7/7	6/6	6/7	4/7
Mean repeat no.	6·1	12·2	10	15·5	20·7	10·5	5·6	11·2	12·7	8·6
Variance repeat no.	9·9	5·6	3·0	8·0	1·5	0	20·1	48·9	14·6	4·0
Flagstaff, *Arizona* (*SE*)										
Heterozygosity	0·879	0·861	0·813	0·880	0·692	0·256	0·933	0·945	0·935	0·781
No. of alleles/no. chr.	7/14	7/14	6/14	6/14	4/14	3/15	10/15	9/11	9/14	5/15
Mean repeat no.	4·7	13·5	11·8	13·1	20·4	10·0	10·8	10·7	14·5	7·9
Variance repeat no.	10·9	4·9	9·1	6·9	1·2	2·4	20·6	37·9	9·6	1·4
					D. persimilis					
Mather, *California*										
Heterozygosity	0·903	0·000	0·920	0·912	0·457	0·899	0·954	0·890	0·906	0·239
No. of alleles/no. chr.	10/30	1/30	13/30	9/19	5/30	11/30	16/30	11/30	11/30	2/30
Mean repeat no.	10·6	7·0	6·6	13·7	16·6	10·0	7·5	10·4	12·9	3·1
Variance repeat no.	12·5	0	4·6	6·9	3·1	4·4	16·8	4·4	7·0	0·1
Mount St Helena, *California*										
Heterozygosity	0·873	0·000	0·851	0·808	0·140	0·754	0·948	0·884	0·905	0·377
No. of alleles/no. chr.	8/14	1/14	8/14	5/14	2/14	6/14	11/14	9/14	9/14	3/14
Mean repeat no.	10·1	7·0	7·4	15·2	16·1	10·0	6·6	9·2	12·8	3·6
Variance repeat no.	12·6	0	8·6	2·4	0·3	4·6	29·5	22·8	7·7	2·6

[a] Number of alleles observed per number of chromosomes assayed.
[b] Mean inferred number of repeats.
[c] Variance in mean inferred number of repeats.

morphism probably affects the level of structure between eastern and western populations at *DPS3001*, but the level of population structure as shown by *DPSX002* and *DPS4002* is more likely a function of genetic drift or natural selection. The large number of alleles at some of these microsatellite loci could have obscured the evidence for population structure in this species. At some loci, almost half the individuals surveyed possessed unique alleles, making tests for allelic frequency differences problematic. Hence, given a relatively small total sample size, the most informative loci for population structure within the species are those with the smallest number of alleles. We therefore conclude that the mild though statistically significant structure we observed in *DPSX002* and *DPS4002* is most representative of real structure among *D. pseudoobscura* populations.

(*b*) *D. persimilis*. All the loci amplified by PCR in both *D. pseudoobscura* and *D. persimilis*. For *D. persimilis*, three of the loci exhibited relatively low heterozygosity ($H < 0{\cdot}40$; *DPSX002*, *DPS2001* and *DPS4002*), and the remaining loci had heterozygosities higher than 0·80 (Table 3). We discuss possible reasons for this difference between *D. pseudoobscura* and *D. persimilis* below. Nonetheless, variability at microsatellite loci in both these species appears to be much greater than observed in comparable studies of *D. melanogaster* (Wetterstrand, 1997; Schug *et al.*, 1998*b*).

(*c*) *Comparison with* D. melanogaster. The high levels of genetic variation at the dinucleotide repeat loci in *D. pseudoobscura* relative to dinucleotide repeats in *D. melanogaster* may be a function of a higher mutation rate, or larger effective population size (N_e) of *D. pseudoobscura*. Using the squared difference in average numbers of repeats ($(\delta\mu)^2$) between *D. pseudoobscura* and *D. persimilis*, Goldstein *et al.* (1995) showed that $E[(\delta\mu)^2 t] = 2\mu t$, where t is the number of generations since separation/speciation and μ is the mutation rate per generation. The time since species separation is estimated to be 500000 years (Aquadro *et al.*, 1991; Babcock & Anderson, 1996), and we assume five generations per year. Solving this equation for μ results in an estimated mutation rate of $2{\cdot}18 \times 10^{-6}$ mutations per generation, consistent with the empirically derived $9{\cdot}3 \times 10^{-6}$ mutations per generation in *D. melanogaster* dinucleotide repeats (Schug *et al.*, 1998*c*). Secondly, if we assume that the mutation rate of *D. pseudoobscura* dinucleotide repeats is similar to that of *D. melanogaster*, we can also estimate the effective population size (N_e) using two methods (for references see Schug *et al.*, 1998*c*). If the difference in variability in these species is a function of a higher mutation rate in *D. pseudoobscura*, we expect to overestimate the effective population size of this species using the *D. melanogaster* empirically derived mutation rate. The infinite alleles model predicts $N_e\mu = H/[4(1-H)]$, where H is the average heterozygosity and μ is the mutation rate. Solving this equation results in an N_e of 141000. The stepwise mutation model predicts that $1-H = 1/\mathrm{sqrt}(1+8\,N_e\mu)$. Solving this equation for N_e results in an estimated effective population size of 512000. These figures are about one order of magnitude smaller than the only other estimate of N_e using molecular genetic data in this species (Schaeffer, 1995). Excluding third chromosome loci from these calculations does not substantially change any of these figures. We discuss possible interpretations of these calculations below.

4. Discussion

We have isolated and characterized an array of 37 di-, tri- and tetranucleotide repeat arrays from the genome *D. pseudoobscura*. We used 24 of these microsatellites to develop a recombinational linkage map of three major chromosomes of *D. pseudoobscura*. The estimate of the total recombinational length of these chromosomes was larger than previously inferred from morphological mutant markers but may simply be due to strain differences in rates of recombination (see below). The 10 microsatellites with which we studied population variation are highly variable within *D. pseudoobscura*, and seven of these are highly variable in *D. persimilis*. Genetic structure was significant among the four *D. pseudoobscura* populations, but levels of structure are very low, corresponding with studies of allozyme and nucleotide sequence variation which indicate that levels of gene flow among these populations is high (e.g. Prakash *et al.*, 1969; Schaeffer & Miller, 1992). We also found no compelling evidence for local selective sweeps but did find low variation at a third chromosome locus within populations fixed for the Arrowhead third chromosome inversion.

One might argue that the recombinational distances we observed were inflated by the 'interchromosomal effect': an increase in recombinational rates in some parts of the genome when recombination is inhibited in other areas. Since the third chromosome of this species possesses a rich inversion polymorphism, and one of our crosses was done between strains of different inversion types, recombination rates on the other three major chromosomes may have been slightly inflated. A concurrent mapping study by Hamblin and Aquadro (in press) used some of the same markers in a single homosequential cross and found a smaller total map length for the second chromosome (128 cM). They documented a slightly but non-significantly smaller recombinational distance (11·8 vs 17·3 cM) between *Rh1* and *DPS2003* than we noted with our heterokaryotypic cross. In contrast,

https://doi.org/10.1017/S0016672399004024 Published online by Cambridge University Press

our estimates of the distance between *Rh1* and *bcd* were nearly identical (38·5 vs 38·0 cM), and both used homosequential lines. These results are consistent with the interchromosomal effect inflating our map distances in cross 1. Direct comparisons of our estimated overall map length with the morphological marker maps are difficult because the two morphological marker studies did not identify the third chromosomal arrangements used (Anderson, 1993; Orr, 1995). Although the interchromosomal effect may have inflated some of our map distances, our data suggest that the total recombinational map length of *D. pseudoobscura* is larger than previously inferred from morphological mutant markers in at least some pairings.

Three microsatellites in *D. persimilis* were both much smaller and less variable than corresponding microsatellites in *D. pseudoobscura*. This finding is consistent with the more general phenomenon that has been documented in a wide variety of species (see Ellegren *et al*., 1995): microsatellites amplified in species other than that in which they were isolated often tend to be shorter and less variable (e.g. Rubinsztein *et al*., 1995). The explanation for this phenomenon, however, is a subject of considerable debate (Ellegren *et al*., 1995; Amos & Rubinsztein, 1996; Ellegren *et al*., 1997; Crawford *et al*., 1998; Hutter *et al*., 1998). A bias may result from the selection of longer-than-average repeat arrays in one species. Homologues in congeners may thus not have evolved equally long and/or uninterrupted repeats at these loci (Ellegren *et al*., 1997; Hutter *et al*., 1998). Alternatively, the pattern may be a byproduct of the larger effective population size of the focal species (Rubinsztein *et al*., 1995; Amos & Rubinsztein, 1996). The results of Hutter *et al*. (1998) support the ascertainment bias hypothesis for the variability of *Drosophila melanogaster* microsatellites relative to *D. simulans*.

Several other processes can also explain the observation of differences between species in length or variability of their microsatellites. For example, the substitution of a base-pair in the middle of the microsatellite in one taxon, hence creating an imperfect repeat, can reduce the variability of such an array (Goldstein & Clark, 1995). The microsatellites may have become variable (due to unequal crossing over or replication slippage) after the split of the two taxa. Hence, one taxon would be invariant (or much less variable) for the ancestrally short microsatellite whereas the other might be quite variable for a longer microsatellite. This may be true for *DPS4002* in our study: the microsatellite may have only recently begun to evolve in *D. persimilis* while it may have done so long ago in *D. pseudoobscura*. Species-specific differences in regional rates of recombination (e.g. True *et al*., 1996) may affect the rate of mutation and variability at microsatellite loci. Indeed, this may be a composite phenomenon, with different explanations for different loci.

The high levels of genetic variation at the dinucleotide repeat loci in *D. pseudoobscura* relative to dinucleotide repeats in *D. melanogaster* may be a function of a higher mutation rate, or larger effective population size of *D. pseudoobscura*. If this difference was a function of a higher mutation rate in *D. pseudoobscura*, we expect to overestimate the effective population size of this species using the *D. melanogaster* empirically derived mutation rate. We estimated effective population size given the empirically derived mutation rate documented in *D. melanogaster* (Schug *et al*., 1997, 1998*c*). Both estimates were similar, and somewhat smaller than that estimated from an independent study of nucleotide sequence polymorphism at *Adh* ($4{\cdot}5 \times 10^6$: Schaeffer, 1995). Our population size estimate may be lower than the *Adh* estimate for several reasons, such as the presence of additional repeats within the PCR product (see below). Studies of genetic variability in allozymes and other genetic markers also suggest that the effective population size of *D. pseudoobscura* is larger than the widely accepted size of *D. melanogaster* (e.g. Kreitman, 1983; Choudhary & Singh, 1987). Finally, the mutation rate estimated from $(\delta\mu)^2$ (Goldstein *et al*., 1995) is also consistent with the mutation rate determined in *D. melanogaster*. Based on these three analyses, it is likely that the mutation rate of dinucleotide repeats in *D. pseudoobscura* is similar to that in *D. melanogaster*. We also note that the presence of additional repeats within the PCR product does not confound our conclusions. If variability is increased by other repeat motifs within the same PCR products, then our calculated mutation rate would *overestimate* the true dinucleotide repeat mutation rate; hence our calculations are conservative.

Local selective sweeps cause reductions in linked neutral allelic variation in single populations while levels of variation remain high in other populations. We found no compelling evidence of local selective sweeps at any of the loci we surveyed. The levels of heterozygosity were very high in all *D. pseudoobscura* populations at all loci except *DPS3001*. However, since this locus is on the third chromosome, it may possess very low variability within each inversion type (perhaps due to a low recombination rate in the vicinity of the microsatellite) and only appear to be highly polymorphic in the northwestern populations because the inversions themselves are highly polymorphic. Utah and Flagstaff are virtually fixed for the Arrowhead inversion, whereas the remaining populations are highly polymorphic (Anderson *et al*., 1991), corresponding to the levels of variability we observed across populations at *DPS3001*. Thus, in our initial survey of microsatellite variation at 10 loci

in four populations we find no evidence consistent with local selective sweeps in regions of the chromosomes around these loci.

DPS3002 and *DPS3003* were not as structured across *D. pseudoobscura* populations as was *DPS3001*. At least two reasons can explain this difference. First, the large number of alleles relative to individuals sampled in those two loci may have obscured evidence of structure (though F_{ST} was significant in *DPS3003*). Secondly, *DPS3001* may be cytologically distant from *DPS3002* and *DPS3003*, hence not captured in the same inversions, or *DPS3001* may be closer to an inversion breakpoint hence preventing gene exchange between arrangements more effectively than at loci more centrally located within the inversion (e.g. Rozas *et al*., 1999).

Our finding of low but statistically significant structure across *D. pseudoobscura* populations corresponds with allozyme (Prakash *et al*., 1969), mtDNA sequence (Jenkins *et al*., 1996) and nuclear DNA sequence studies (Schaeffer & Miller, 1992) in this species. The higher levels of variability of microsatellites in this species relative to *D. melanogaster* appear to be a function of the larger effective population size rather than a higher mutation rate. These high levels of variability will make microsatellites ideal for a variety of genetic applications in *D. pseudoobscura* and *D. persimilis* (e.g. Noor & Aquadro, 1998).

P. Service and J. Farmer provided the isofemale lines from Flagstaff and Abajo, respectively. We also thank S. Schaeffer and M. Hamblin for sharing their results with us prepublication. This work was supported by a National Science Foundation/Alfred Sloan Foundation postdoctoral fellowship to M.N. in 1996, an NIH National Research Service Fellowship to M.S., and National Institutes of Health grant DM36431 to C.A.

References

Amos, W. & Rubinsztein, D. C. (1996). Microsatellites are subject to directional evolution. *Nature Genetics* **12**, 13–14.

Anderson, W. W. (1993). Linkage map of *Drosophila pseudoobscura*. In *Genetic Maps: Locus Maps of Complex Genomes* (ed. S. J. O'Brien), pp. 3.252–3.253. Plainview, NY: Cold Spring Harbor Laboratory Press.

Anderson, W. W., Arnold, J., Baldwin, D. G., *et al*. (1991). Four decades of inversion polymorphism in *Drosophila pseudoobscura*. *Proceedings of the National Academy of Sciences of the USA* **88**, 10367–10371.

Aquadro, C. F., Weaver, A. L., Schaeffer, S. W. & Anderson, W. W. (1991). Molecular evolution of inversions in *Drosophila pseudoobscura*: the amylase gene region. *Proceedings of the National Academy of Sciences of the USA* **88**, 305–309.

Babcock, C. S. & Anderson, W. W. (1996). Molecular evolution of the sex-ratio inversion complex in *Drosophila pseudoobscura*: analysis of the *Esterase-5* gene region. *Molecular Biology and Evolution* **13**, 297–308.

Choudhary, M. & Singh, R. S. (1987). Historical effective size and the level of genetic diversity in *Drosophila melanogaster* and *Drosophila pseudoobscura*. *Biochemical Genetics* **25**, 41–51.

Crawford, A. M., Kappes, S. K., Paterson, K. A., deGotari, M. J., Dodds, K. G., Freking, B. A., Stone, R. T. & Beattie, C. W. (1998). Microsatellite evolution: testing the ascertainment bias hypothesis. *Journal of Molecular Evolution* **46**, 256–260.

Ellegren, H., Primmer, C. R. & Sheldon, B. C. (1995). Microsatellite 'evolution': directionality or bias? *Nature Genetics* **11**, 360–362.

Ellegren, H., Moore, S., Robinson, N., Byrne, K., Ward, W. & Sheldon, B. C. (1997). Microsatellite evolution: a reciprocal study of repeat lengths at homologous loci in cattle and sheep. *Molecular Biology and Evolution* **14**, 854–860.

Gloor, G. B., Preston, C. R., Johnson-Schlitz, D. M., Nassif, N. A., Phillis, R. W., Benz, W. K., Robertson, H. M. & Engels, W. R. (1993). Type I repressors of P element mobility. *Genetics* **135**, 81–95.

Goldstein, D. B. & Clark, A. G. (1995). Microsatellite variation in North American populations of *Drosophila melanogaster*. *Nucleic Acids Research* **23**, 3882–3886.

Goldstein, D. B., Ruiz-Linares, A., Feldman, M. W. & Cavalli-Sforza, L. L. (1995). Genetic absolute dating based on microsatellites and the origin of modern humans. *Proceedings of the National Academy of Sciences of the USA* **92**, 6720–6727.

Goudet, J. (1995). FSTAT, a program for IBM PC compatibles to calculate Weir and Cockerham's estimators of F-statistics.

Haldane, J. B. S. (1919). The combination of linkage values and the calculation of distances between the loci of linked factors. *Journal of Genetics* **8**, 299–309.

Hamblin, M. T. & Aquadro, C. F. (in press). DNA sequence variation and the recombinational landscape in *Drosophila pseudoobscura*: a study of the second chromosome. *Genetics*, in press.

Hutter, C. M., Schug, M. D. & Aquadro, C. F. (1998). Microsatellite variation in *Drosophila melanogaster* and *D. simulans*: a reciprocal test of the ascertainment bias hypothesis. *Molecular Biology and Evolution* **15**, 1620–1636.

Irvin, S. D., Wetterstrand, K. A., Hutter, C. M. & Aquadro, C. F. (1998). Genetic variation and differentiation at microsatellite loci in *Drosophila simulans*: evidence for founder effects in New World populations. *Genetics* **150**, 777–790.

Jenkins, T. M., Basten, C. J. & Anderson, W. W. (1996). Mitochondrial gene divergence of Colombian *Drosophila pseudoobscura*. *Molecular Biology and Evolution* **13**, 1266–1275.

Kosambi, D. D. (1944). The estimation of map distances from recombination values. *Annals of Eugenics* **12**, 172–175.

Kreitman, M. (1983). Nucleotide polymorphism at the alcohol dehydrogenase locus of *Drosophila melanogaster*. *Nature* **304**, 412–417.

Noor, M. A. F. (1997). Genetics of sexual isolation and courtship dysfunction in male hybrids of *Drosophila pseudoobscura* and *D. persimilis*. *Evolution* **51**, 809–815.

Noor, M. A. F. & Aquadro, C. F. (1998). Courtship songs of *Drosophila pseudoobscura* and *D. persimilis*: analysis of variation. *Animal Behaviour* **56**, 115–125.

Orr, H. A. (1987). Genetics of male and female sterility in hybrids of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics* **116**, 555–563.

Orr, H. A. (1995). A new linkage map of the *D. pseudo-*

obscura X chromosome. *Drosophila Information Service* **76**, 127–128.

Prakash, S., Lewontin, R. C. & Hubby, J. L. (1969). A molecular approach to the study of genic heterozygosity in natural populations. IV. Patterns of genic variation in central, marginal and isolated populations of *Drosophila pseudoobscura*. *Genetics* **61**, 841–858.

Richter, B., Long, M., Lewontin, R. C. & Nitasaka, E. (1997). Nucleotide variation and conservation at the *dpp* locus, a gene controlling early development in *Drosophila*. *Genetics* **145**, 311–323.

Rozas, J., Segarra, C., Ribó, G. & Aguadé, M. (1999). Molecular population genetics of the *rp49* gene region in different chromosomal inversions of *Drosophila subobscura*. *Genetics* **151**, 189–202.

Rubinsztein, D. C., Amos, W., Leggo, J., Goodburn, S., Jain, S., Li, S.-H., Margolis, R. L., Ross, C. A. & Ferguson-Smith, M. A. (1995). Microsatellite evolution: evidence for directionality and variation in rate between species. *Nature Genetics* **10**, 337–343.

Schaeffer, S. (1995). Population genetics in *Drosophila pseudoobscura*: a synthesis based on nucleotide sequence data for the *Adh* gene. In *Genetics of Natural Populations: The Continuing Importance of Theodosius Dobzhansky* (ed. L. Levine), pp. 329–352. New York: Columbia University Press.

Schaeffer, S. W. & Miller, E. L. (1992). Estimates of gene flow in *Drosophila pseudoobscura* determined from nucleotide sequence analysis of the alcohol dehydrogenase region. *Genetics* **132**, 471–480.

Schlotterer, C., Vogl, C. & Tautz, D. (1997). Polymorphism and locus-specific effects on polymorphism at microsatellite loci in natural *Drosophila melanogaster* populations. *Genetics* **146**, 309–320.

Schug, M. D., Mackay, T. F. C. & Aquadro, C. F. (1997). Low mutation rates of microsatellite loci in *Drosophila melanogaster*. *Nature Genetics* **15**, 99–102.

Schug, M. D., Hutter, C. M., Noor, M. A. F. & Aquadro, C. F. (1998*a*). Mutation and evolution in microsatellites of *Drosophila melanogaster*. *Genetica* **102/103**, 359–367.

Schug, M. D., Wetterstrand, K. A., Gaudette, M. S., Lim, R. H., Hutter, C. M. & Aquadro, C. F. (1998*b*). The distribution and frequency of microsatellite loci in *Drosophila melanogaster*. *Molecular Ecology* **7**, 57–70.

Schug, M. D., Hutter, C. M., Wetterstrand, K. A., Gaudette, M. A., Mackay, T. F. C. & Aquadro, C. F. (1998*c*). The mutation rate of di-, tri-, and tetranucleotide repeats in *Drosophila melanogaster*. *Molecular Biology and Evolution* **15**, 1751–1760.

Slatkin, M. (1995). A measure of population subdivision based on microsatellite allele frequencies. *Genetics* **139**, 457–462.

Tan, C. C. (1946). Genetics of sexual isolation between *Drosophila pseudoobscura* and *Drosophila persimilis*. *Genetics* **31**, 558–573.

True, J. R., Mercer, J. M. & Laurie, C. C. (1996). Differences in crossover frequency and distribution among three sibling species of *Drosophila*. *Genetics* **142**, 507–523.

Wang, R. L., Wakeley, J. & Hey, J. (1997). Gene flow and natural selection in the origin of *Drosophila pseudoobscura* and close relatives. *Genetics* **147**, 1091–1106.

Weber, J. L. (1990). Informativeness of human (dC-dA)$_n$(dG-dT)$_n$ polymorphisms. *Genomics* **7**, 517–524.

Wetterstrand, K. A. (1997). Microsatellite polymorphism and divergence in worldwide populations of *Drosophila melanogaster* and *D. simulans*. Masters thesis. Ithaca, NY: Cornell University.