Genet. Res., Camb. (2004), **83**, *pp.* 19–29. © 2004 Cambridge University Press
DOI: 10.1017/S0016672303006542 Printed in the United Kingdom

Isolation and characterization of dinucleotide repeat microsatellites in *Drosophila ananassae*

M. D. SCHUG*, E. E. REGULSKI†, A. PEARCE AND S. G. SMITH
Department of Biology, 301 Eberhart Building, University of North Carolina Greensboro, Greensboro, NC 27455, USA

(*Received 22 August 2003 and in revised form 28 October 2003*)

Summary

Drosophila ananassae is a cosmopolitan species with a geographic range throughout most of the tropical and subtropical regions of the world. Previous studies of DNA sequence polymorphism in three genes has shown evidence of selection affecting broad expanses of the genome in regions with low rates of recombination in geographically local populations in and around India. The studies suggest that extensive physical and genetic maps based on molecular markers, and detailed studies of population structure may provide insight into the degree to which natural selection affects DNA sequence polymorphism across broad regions of chromosomes. We have isolated 85 dinucleotide repeat microsatellite sequences and developed assay conditions for genotyping using PCR. The dinucleotide repeats we isolated are shorter, on average, than those isolated in many other *Drosophila* species. Levels of genetic variation are high, comparable to *Drosophila melanogaster*. The levels of variation indicate the effective population size of an Indonesian population of *D. ananassae* is 58 692 (infinite allele model) and 217 284 (stepwise mutation model), similar to estimates of effective population size for *D. melanogaster* calculated using dinucleotide repeat microsatellites. The data also show that the Indonesian population is in a rapid expansion phase. Cross-species amplification of the microsatellites in 11 species from the Ananassae, Elegans, Eugracilis and Ficusphila subgroups indicates that the loci may be useful for studies of the sister species, *D. pallidosa*, but will have limited use for more distantly related species.

1. Introduction

D. ananassae is the most abundant *Drosophila* species in much of the tropical and subtropical regions of the world (Tobari, 1993) and has even been observed in the milder American climatic regions (Dobzhansky & Dreyfus, 1943). Its centre of geographic origin is thought to be in Southeast Asia and it has most probably colonized much of the world very recently, invading a variety of climatic zones. It currently exists in many semi-isolated populations in the geographic regions where it has been studied (Johnson, 1971; Stephan & Langley, 1989; Stephan, 1989; Lynch & Crease, 1990; Tomimura *et al*., 1993).

Population structure is evident along clines in India (Prakash *et al*., 1994; Singh, 1998), and is particularly strong among the island populations in the South Pacific Ocean (Johnson, 1971; Tomimura *et al*., 1993). Detailed cytological and genetic maps based on polytene chromosomes and visible mutants have been constructed for *D. ananassae* (reviewed in Tobari, 1993). These maps have been useful for determining whether regions of chromosomes have low or normal rates of recombination. Based on the high levels of population structure and ability to classify regions of chromosomes into normal and low rates of recombination, Stephan *et al*. (1998) and Chen *et al*. (2000) demonstrated that DNA sequence variation in two single-copy nuclear genes on the X chromosome (*vermilion* and *furrowed*) with low rates of recombination show strong evidence of selective sweeps in populations from Sri Lanka, India, Nepal and Myanmar. One gene in a region of normal recombination (*Om1D*), on the other hand, shows no evidence of selection, but rather significant population structure represented by a pattern of isolation-by-distance.

* Corresponding author. Tel: +1 336 2560086. e-mail: mdschug@uncg.edu
† Current address: Department of Molecular, Cellular, and Developmental Biology, Yale University, New Haven, CT 06520, USA.

https://doi.org/10.1017/S0016672303006542 Published online by Cambridge University Press

https://doi.org/10.1017/S0016672303006542 Published online by Cambridge University Press

Table 1. *$(AC)_n$ microsatellite repeat loci isolated and characterized from a* D. ananassae *subgenomic DNA library*

Locus	Accession no.	*D. melanogaster* chromosome physical location	Repeat type	PCR product size (bp)	*N*	Heterozygosity	Variance	No. of alleles	Mean RU	Max. RU	Forward/reverse primer 5′–3′
DAN183	AY445299	X – 2A	(CA)2TA(CA)5	227	8	0·75					CAAGCGGTAAACAATAAAATCA CACCTCCCCAATCTCCATAA
DAN30	AY445276	X – 6D3	(CA)2AC(CA)5	209							AGCAGAAGGCCACCCACTAT ATGCGTGTGAGTGTTGGTGT
DAN288b	AY445295	X – 6E	(TG)5	110							TGAGTGTGAGTGCGCAGATT CCCAATGTGGCGTATGAGTA
DAN184a	AY445300	X – 11A12	(CA)7	152							CATTGCATATGAAAGTCAATAAATAAT ACCAAATGGAAGGACAAGGA
DAN11	AY445263	X – 11B13	(TG)6	237	30	0	0	1	6	6	TGCCAACACAGTTACACAGGA ACACTCCAGCACACACTCCA
DAN12	AY445264	X – 11B14	(TG)6	197	30	0·26	2·2	4	5·5	6	TTTTGATGGAAGATGAAATGGA TGTTGCATCCTGCCATTACT
DAN83	AY445252	X – 11B14	(TG)6	273	26	0·31	1·08	3	11·35	22	TGCCAACACAGTTACACAGGA CATGTTGCATCCTGCCATTA
DAN172	AY445297	X – 11B14	(TG)6	286	8	0·71					TGCCAACACAGTTACACAGGA CACACGCCATTTTCATGTTG
DAN174	AY445298	X – 11B14	(TG)6	258	8	0·96					CACAATTACACAGGACTCATCACA CCTGCCATTACTGCACTCAC
DAN154	AY445296	X – 12B22	(TG)13	251	26	0·85	13	12	13·19	18	GGGAAAATGTGTCAGCAGGA AAAGGCTTAGTCGAGTGGTTTTT
DAN32	AY445254	X – 14C	(TG)15	277	30	0·87	25·91	14	10·5	20	GTCCTTATTCGCCCTGTGAG CTCCTCCTCCTGCTCCTTCT
DAN77a	AY445253	X – 17B	(TG)6	233	8	0·88					TGCTTCCACGATGTGTCTTC ACACCCCACACCCCTTACTC
DAN77b	AY445253	X – 17B	(TG)6	100	8	1					TGGATAAGGAGTAAGGGGTGTG AGCCCAATAGCCATACAAACA
DAN59	AY445313	X – 19A	(TG)8	260	28	0·78	9·11	8	7·96	15	AAGCTTTCCTTTGGTCTCTGC AAGCAAACAAAACGCTCACC
DAN160	AY445301	X – 19A	(TG)8	195	8	0·83					GGAAAAATCCGAATGGAACA AAAAACGCTCACCCACACAC
DAN185	AY445303	X – 20B	(TG)11	133							GTCCTGGCATGGGTCCTG TTTGTGCCAATAGTCGGTTG
DAN4	AY445258	2L – 32D3	(CA)10	194	24	0·79	3·32	6	9·79	11·5	GGTTCGGGTAAGACAGCAAA GCGGTGTGAGTTTGAGTGTG
DAN14	AY445266	2L – 32D3	(CA)6	146	8	1					TTTAAAACCGAACCCGACTC GTCTGCATGTGTGTGTGTGG
DAN17	AY445267	2L – 32D3	(TG)6	197	30	0·13	1·5	4	6	11	TTTTGATGGAAGATGAAATGGA TGTTGCATCCTGCCATTACT
DAN262a	AY445294	2L – 33A	(CA)5	161							TATAGGATCCCACGcACACA CCCAATTTCCCCAAGTcAAT
DAN73	AY445250	2L – 33B	(CA)8	145	26	0·53	0·65	4	5·69	7	TGACACATACCAATCTATTCACACC TATTGGCAGCACTGTGGAAA
DAN98a	AY445314	2L – 33B	(CA)2CT(CA)5	194	8	0·83					ACACCCCCACAAACAGGATA CTCCATTGTAACCCCCACAT
DAN24	AY445271	2L – 33F	(TG)8	203							CGGTTATCCTCGTTGGTGTC TAGGAGAGAGCCAGGACGAG
DAN166	AY445302	2L – 33B	(CA)5A(CA)4	167	8	0·29					TTACTGCCAGCTGCTGAAGA GGCGATAAAAGTCACGTCATC

https://doi.org/10.1017/S0016672303006542 Published online by Cambridge University Press

DAN85	AY445251	2L – 37C	(TG)9	139	26	0·61	17·73	11	8·69	22	AGCTTTGTCATTGCGGTGTT CACCTCATTTGAAAAATTACCC
DAN13	AY445265	2L – 37C1	(TG)9	224	6	1					CAGAAGCACTTCCCCAAAAA TCACCCTAACACCTCATTTGAA
DAN26	AY445275	2R – 41C	(TG)6	196	30	0·26	1·26	3	5·6	6	TTTGATGGAAGATGAAATGGA TGTTGCATCCTGCCATTACT
DAN243a	AY445293	2R – 47A	(TG)7	151							TCAAGTGTCTCCCTGGTGTG ACGAGGGGTATCTTCGGGTA
DAN6	AY445259	2R – 47D	(CA)6	200	4	1					TCCTGCCATCCTACCAGACT TGCATTTTTCGTCTCGACAG
DAN65	AY445249	2R – 49A	(CA)10	213	30	0·67	1·12	4	7·97	9	ATGCTGGCGACAAGTTCAAT TGCCAATTTCAATTAGCCAAC
DAN178	AY445304	2R – 49A	(CA)2AC(CA)10	307	8	0·83					TGCTGGCGACAAGTTCAATA GCGGAAGCTCTTCTGACTTT
DAN51	AY445256	2R – 50A2	(TG)11	207							CACAAAAACGGGAAAGGACT CGCCAAAGAATGTTCTCCTT
DAN56	AY445255	2R – 50A2	(TG)9	294							TTGTCATTGCGGTGTTTTGT AGCTCCAAAGAAGCTCCCTA
DAN76	AY445246	2R – 53C	(TG)8	221	26	0·08	0·04	2	12·04	13	CGTGCGTATGTGAGTGTGTG CCCTTTATTCGCCATCATGT
DAN81	AY445247	2R – 53C	(TG)11	131	26	0·46	1·5	5	9	11	ATTAATAGCCCGGCAATGTC GCACACTCACATGGCTGAAT
DAN144	AY445279	2R – 53C	(TG)6	251	8	0·75					ATGCTCAAGCGTGCGTAAGT ATTTCTTCCAGTGCCCCTTC
DAN69	AY445248	2R – 57B	(CA)10	258	26	0·69	12·73	11	8·69	14	TGCCCTGCTGACCAATTTA GCTTGGACTGCATATCGTCA
DAN186	AY445305	2R – 57B	(TG)8	263							AAGCCAGGAAGGGCTAAGAG CATTTGGGGCGTTGTAGTTT
DAN21	AY445272	2R – 57B1	(TG)8	201	30	0·53	8·53	6	9·8	20	CGGTTATCCTCGTTGGTGTC TAGGAGAGAGCCAGGACGAG
DAN147	AY445280	2R – 57C	(TG)10	196							TCCAATCCCAATACCAATCC GCACACTCACATGGCTGAAT
DAN120	AY445278	2R – 57D	(TG)11	232	26	0·62	2·31	5	12·69	14	TTAGCTGGTGTCGCTTCGAT GCACACTCACATGGCTGAAT
DAN40	AY445257	2R – 57D3	(TG)11	304	26	0·69	3·19	8	11·23	16	CTCGGACTTGTCCTGGGTAG AGAAGCCAACCATCCATCC
DAN25	AY445273	2R – 57D3	(CA)10	196							GCACACTCACATGGCTGAAT TCCAATCCCAATACCAATCC
DAN140	AY445277	2R – 57D3	(TG)11	232	8	0·96					TTAGCTGGTGTCGCTTCGAT GCACACTCACATGGCTGAAT
DAN167	AY445306	2R – 58D	(CA)11	296	8	1					GCCACCGAAAGTTCAAATTC AAATTCTCAATACCAGCTTTGC
DAN9	AY445261	3L – 62E4	(TG)11	266	20	0·999	15·75	10	9·75	18	AGCATTCGAGCGACTGTTG CAGGACCTTGTAACCGGAAA
DAN82	AY445245	3L – 62E5	(CA)8	237	30	0·83	71·5	15	10·71	28	CAGGACCTTGTAACCGGAAA AAAGGACCTCTAGAAACGGATG
DAN246	AY445290	3L – 63E5	(TG)11	267							GTCCTGGCATGGGTCCTG CGGcAAAATATGCATACCG
DAN249	AY445291	3L – 63E5	(TG)6	175							TTTACcAGTTTCCGCAGTCC GAACCAATCTGCcACTCACA
DAN253	AY445292	3L – 76D1	(TG)10	242							GAGGCGAACAATTGGCTATC CCCtTGGGTGGGtTGTTAG
DAN27	AY445274	3R – 82D	(TG)10	197	28	0·5	1·8	5	9·39	12	TTTGCCTTTGCTGTCAAGTG CCCACCAACGATGTGTAAAA
DAN94	AY445283	3R – 84F2	(CA)5TA(CA)2	216	8	0·29					GAGCGTAGAGAAGGGGGTTT TCCAGTTTGGAGCTTTCCAT

https://doi.org/10.1017/S0016672303006542 Published online by Cambridge University Press

Table 1. (*Cont.*)

Locus	Accession no.	*D. melanogaster* chromosome physical location	Repeat type	PCR product size (bp)	*N*	Heterozygosity	Variance	No. of alleles	Mean RU	Max. RU	Forward/reverse primer 5′–3′
DAN88	AY445243	3R – 94E	(TG)11	167	6	0·75					TGGCTGATAAGGCGACTAGA TGGCCGCCTAATTAGATACG
DAN148	AY445282	3R – 96C	(TG)11	251	8	1					CTTTTGTGGACAGCGCATAA AAGCGTTACAGCTTCCTTGG
DAN8	AY445260	3R – 98C	(CA)6	173	6	0·42					TCCTCTAGCAGAAGCCCAAG GTGTACGTGCGGTGTAAGGA
DAN70a	AY445236	3R – 98C	(TG)7	136	26	0·08	0·16	2	6·08	8	AAGATGCTTAAGTGTGTGGGTGT TATGAGGCATGAGGCACTGA
DAN70b	AY445236	3R – 98C	(TG)9	225							CGGGCAAAAAGTTCCAGTTA CTATCGAGTGCCGCAGACAT
DAN71a	AY445237	3R – 98C	(TG)7	153							GTGTACGTGCGGTGTAAGGA AGGCGAAGTGAGATGGCATA
DAN71b	AY445237	3R – 98C	(TG)7	302	8	0·54					TCCTGCTACGGCACTAGGAT TAACACGCATACGCCATGAT
DAN79	AY445244	3R – 98C	(TG)10	194	26	0·62	2·04	6	8·96	11	CTAAACAGCGTCGGTCCTCT TCCAAAAGTATCTGTGGCTGTG
DAN146	AY445281	3R – 98C	(CA)5	159	8	0·54					TGCGGTAAACGATTTGTTGT ACACAAACTCAACGAATCCAA
DAN84	AY445241	3R – 100C	(CA)10	250							GTTGGAAAGGCAATCACAGG GCATTTCAAGAGCGTGAGTG
DAN7	AY445262		(CA)7A(CA)6	180	26	0·69	4·1	5	8·58	12	CTCTTCTGCTTCGGCTTCTG AATGTGTGTCTGTGCGTTGG
DAN16	AY445268		(TG)9TG(TG)5	223							TCATTTCTCCTCCTGCTTGG GAGAGATGCGAAAGGACAGG
DAN20	AY445269		(TG)2GT(TG)10	172	30	0·93	11·26	10	9·6	14	CAGGCAGTGGGTTTAAGAGC CTCCTTCACCCTAACACCTCA
DAN23	AY445270		(TG)9	288							GCCGCTCTTCCCTCCTCT TGTAGCTCCAAAGTAACTCCCTA
DAN31	AY443009		(TG)9	299	26	0·62	1·14	4	9·15	10	CTCTGGCTCTGGCTCCTCT GCCAAGGAGAAAATAACTCGTC
DAN33	AY445233		(TG)6	197	28	0·21	3·73	4	5·96	14	TTTTGATGGAAGATGAAATGGA TGTTGCATCCTGCCATTACT
DAN42	AY445312		(TG)6T(TG)8	128	24	0·75	21·66	11	19·75	27	GTTGGCGTTTTGTGGGTTT TGCCTCTCTTCTGCTTCCTT
DAN45	AY445234		(TG)5	173	28	0·71	3·96	5	5·54	7	AGCGGGCCAGTGACAAAA TAGCAAACAAAACGCTCACC
DAN54	AY445235		(TG)9	253							TCGTGGTATCACACTTGGTTG CAAGCAAAGGAGAAAATAACTCG
DAN75	AY445238		(CA)4AA(CA)6	147	8	0·5					CAGAAGCCCTCAAGAAAGGA GACGCCTCCAGTAGGGAGTA
DAN78	AY445239		(CA)7	161	24	0·31	0·18	3	6	7	TATAGGATCCCACGCACACA CCCAATTTCCCCAAGTCAAT
DAN89a	AY445240		(CA)6	285							CCAGAGGAGAGATGGCACAG ATGCAGCTTAACCCGTGACT
DAN124	AY445284		(TG)6	146	6	0·42					TTTTTCTTTGATTTTCTGTATGTGTG ATGCTGGCATGTGTGTAGGA
DAN136	AY445286		(CA)6A(CA)3A(CA)4	247	28	0·43	9·53	5	19·07	32	AACAAGAGAAGCCCCCAAGT CCCTACCTCCTCCAGCAATTA

DAN121	AY445285	(TG)7	189	8	0·88	AACTTCGATGAGTGCCAGGT ATGCTGGCATGTGTGTAGGA
DAN86	AY445242	(TG)8	186	6	1	TCCCAATACCAATCCCAATG GGCATGGATGCGACTTATTT
DAN161	AY445307	(CA)7TA(CA)2	155	6	1	TAATTCCTTTGGCAGCCTCA TCCTTCCCTTTCTCTCTCTCTG
DAN142	AY445287	(TG)7	240	6	0	CTTTTGTGGACAGCGCATAA TTCCTTGCCACCCTGTTAAA
DAN145	AY445288	(CA)8	146	8	0	TAGCAGAAGCCCGATTGAGT TGCTACGTAGTGCTGCCTTC
DAN180	AY445308	(CA)9	264	6	1	GCCGCTTTTCAGTAAGTAGGC GTTTCATGGAGTGTGTGTGTG
DAN182a	AY445309	(TG)2TC(TG)9	200	8	0·75	GCAGAGCTTAGGTGATGGACA AAAGCCTTATGAAAGCCATCC
DAN173	AY445310	(CA)11	335	6	0·92	CACTGGCGTACTGATTGAATG CAGAATACAATTTCACACATGAAACA
DAN175	AY445311	(TG)10	323	8	0·83	TGAAGATTAACCACACCAAGACA GTGGGTTGTTAGCCTCCTGT

Note: All samples amplified the cloned DNA fragment from which they were characterized. Some samples were not tested for genetic variation in samples from natural populations.

These studies suggest that extensive physical and genetic maps, and studies of population structure, may provide insight into the degree to which natural selection affects DNA sequence variation across broad regions of the genome with different rates of recombination.

Microsatellites are the most commonly used class of DNA markers for genetic mapping because of their high frequency in the genome, high levels of variation relative to other genetic markers, and the technical ease of assaying genotypes. They have been used extensively for studies of population structure and demographic history in natural populations of many animals because of the ease of genotyping and high mutation rates (Jarne & Lagoda, 1996; Ellegren, 2000). Furthermore, we know that dinucleotide repeat microsatellites are likely to be abundant and highly variable in *D. ananassae* because virtually every *Drosophila* species genome studied to date shows these characteristics (Schug *et al.*, 1998*a*; Hutter *et al.*, 1988; Bachtrog *et al.*, 1999; Pascual *et al.*, 2000; Noor *et al.*, 2000; Schlotterer & Harr, 2000; Wilder *et al.*, 2002; Ross *et al.*, 2003).

We screened a DNA library to isolate and characterize a novel set of dinucleotide repeat microsatellite loci distributed at regular intervals across the three major chromosomes of *D. ananassae*. Here we report the characteristics of 85 *D. ananassae* dinucleotide repeat microsatellites with the purpose of comparing their general characteristics with those identified in other species of *Drosophila* and evaluating their potential use as genetic markers for mapping and population genetic analysis in natural populations.

2. Methods

A subgenomic DNA library was constructed using an enrichment protocol described by Hamilton *et al.* (1999) from *D. ananassae* DNA. Briefly, we enriched 300–600 bp fragments of *D. ananassae* DNA for (CA) and (TG) repeats in a hybridization reaction using biotinylated oligonucleotide probes $(CA)_{14}$, $(TG)_{14}$ bound to streptavidin-coated magnetic Dynabeads. Fragments were then cloned into pUC 18, transformed into *E. coli* competent cells and purified plasmids were sequenced using a LiCor automated DNA analyser. All the clones we sequenced contained microsatellites. Duplicate clones with identical or overlapping sequences were discarded. We designed primers for cloned DNA fragments that contained at least 24 bp of DNA sequence flanking the dinucleotide repeats. Loci were labeled as '*Dan*#' where # represents the number of the clone picked from the LB plate. Primers were designed to amplify the repeat regions using Primer v. 3.0 (Rozen & Skaletsky, 2000). We used a step-down amplification protocol for all PCR reactions. The protocol starts with a 60 °C

https://doi.org/10.1017/S0016672303006542 Published online by Cambridge University Press

annealing temperature and steps down 3 °C every 3 cycles until it reaches a 50 °C annealing temperature. At this stage 26 cycles of a 50 °C annealing temperature are performed. We have found the step-down protocol consistently produces high-quality genotypes with most of the primers. All PCR fragments were first analysed on an agarose gel. Genotyping was performed in 10 μl PCR reactions using a tailed-primer protocol to label with IRD700 or IRD800 and fragments were analysed on a LiCor automated DNA analyser.

Genetic variation was assayed for 63 loci in isofemale lines established from females sampled from a population in Java, Indonesia by Muneo Mutsuda. Because each isofemale line is highly inbred, we scored only one allele for each individual, choosing an allele randomly on the rare occasions when a heterozygous genotype was present. We calculated expected heterozygosity for all of the 63 loci. However, variance in repeat unit, number of alleles, and estimates of average and mean repeat unit length are highly sensitive to sample sizes (Pritchard & Feldman, 1996). Thus, we only calculated these measures of genetic variation for samples with 20 or more chromosomes. For loci with fewer samples, measures of heterozygosity are not meant to be used for population genetic analysis, but rather to provide an idea of the potential usefulness of the particular locus for genetic mapping and population genetic studies with larger sample sizes. All measures of genetic variation were calculated using MSA2 (Dieringer & Schlotterer, 2003). All PCR reactions were performed using the 50Enhance protocol as described above and scored on a LiCor automated DNA analyser using at least four size standards and GeneImage IR software (Scanalytics, Inc.). We re-ran gels and re-scored samples of loci to judge the accuracy of scoring among gels and scoring analyses and found very few ambiguities.

Cross-species amplifications were performed using the same protocol as above except the primers were not labelled and the PCR products were separated by electrophoresis on a 2% agarose gel and stained with ethidium bromide adjacent to a 100 bp size marker. Negative controls with no DNA were included in all experiments.

3. Results and Discussion

We have isolated, sequenced and developed primers for 85 $(AC)_n$ dinucleotide repeat loci (Table 1). Of these, 69 (81%) are perfect repeats and 16 (19%) are compound repeats. We performed blast searches against the *D. melanogaster* genome sequence to determine the homologous physical location. Based on previous blast searches of single-copy nuclear gene coding sequences, homology between *D. ananassae* and *D. melanogaster* is approximately 80% (personal observation). Because most dinucleotide repeats are in non-coding sequence, we have included any sequences with 65% or higher homology in our table. We believe this is conservative because homologous intron sequences between *D. ananassae* and *D. melanogaster* are highly divergent, precluding potential alignment (A. Das & W. Stephan, personal communication). Precise physical locations will require *in situ* hybridizations to polytene chromosomes, a project currently under way in our laboratory. If our assumption is correct, the microsatellites we isolated are distributed evenly across the genome, though there are clearly clusters in some regions (Table 1). We are currently sequencing and characterizing more microsatellites with the goal of obtaining a genetic map based on one microsatellite approximately every 3 cM. PCR primers, chromosomal location and levels of genetic variation will be available at http://www.uncg.edu/~mdschug.

(i) *Repeat unit length*

The mean repeat unit length measured as the longest stretch of perfectly repeated units within a cloned DNA fragment is 8·75 repeat units – short relative to most other taxa (Schug *et al*., 1998*a*). Using the total repeat unit length including repeated units flanking an interrupted repeat increases the number only slightly to 9·15. The short length of microsatellites in most *Drosophila* species for which microsatellites have been characterized is well documented (Table 2). The average repeat unit length of dinucleotide repeats in this study is shorter than in many other *Drosophila* species (Table 2). Repeat unit length may be influenced by mutation rate, different constraints on maximum repeat length due to selection (reviewed in Ellegren, 2000) or a bias from our DNA library screen. We believe the shorter average repeat unit length relative to other *Drosophila* species may reflect a bias in the techniques we used for identifying the microsatellites rather than a shorter average repeat unit length of dinucleotide repeats in the genome of *D. ananassae*.

DNA library enrichment procedures are likely to bias selection of microsatellites towards identifying longer repeat units because of the nature of the hybridization reaction during the enrichment procedure. However, there are two reasons to believe that this procedure may bias estimates of microsatellite repeat unit length towards fewer repeats than a traditional subgenomic DNA library screen without enrichment. First, since most clones in the subgenomic DNA library contain microsatellites there is no need to screen colonies by hybridization to identify clones that contain repeat units as is typically performed in a standard DNA library. Screening colonies involves a hybridization reaction in which the DNA from colonies is transferred to a nylon membrane and

https://doi.org/10.1017/S0016672303006542 Published online by Cambridge University Press

Table 2. *Comparison among* Drosophila *species for repeat unit length and genetic variation at dinucleotide repeat microsatellites*

Species	No. dinucleotide repeat loci	Mean repeat length[a]	Mean heterozygosity	Reference
D. ananassae	85	8·09	0·61	Present study
D. melanogaster	41	10·30	0·57–0·61	Schug *et al.* (1998*b*)
D. melanogaster	10	12·2		Schlotterer & Harr (2000)
D. simulans	55	10·05	0·59–0·64	Hutter *et al.* (1998)
D. pseudoobscura	35	11·7	0·84	Noor *et al.* (2000)
D. subobscura	95	14·9	0·77	Pascual *et al.* (2000)
D. virilus	26	12·7	0·70	Schlotterer & Harr (2000)
D. nigrodunni	33	10·8		Wilder *et al.* (2002)
D. dunni dunni	25	8·6		Wilder *et al.* (2002)
D. arizonae	114	9·05		Ross *et al.* (2003)
D. mojavensis	258	8·86		Ross *et al.* (2003)
D. pachea	174	9·8		Ross *et al.* (2003)
D. neotestacea	114	8·13		Ross *et al.* (2003)
D. recens	51	7·55		Ross *et al.* (2003)

[a] Repeat unit length is tabulated as the longest stretch of perfect repeats in a fragment.

hybridized with a labelled probe containing the microsatellite repeat of interest. Colonies for which hybridization occurs are detected by exposure to X-ray film, or in a colorometric reaction. Plasmids within the colonies that contain longer inserts usually produce a more intense signal appearing as a darker or more intensely coloured spot, on the membrane. Because of the likelihood of a percentage of the colonies being false positives, the technician usually chooses the darkest or most intensely coloured colonies first, which contain fragments with the longest microsatellites in the pool of potentially positive clones. In contrast, the enriched library clones are typically chosen at random because without a hybridization screening using labelled microsatellite probes, it is not possible to give priority to any of the colonies containing plasmids with microsatellites. By the nature of these different procedures, enriched subgenomic DNA library screens may produce shorter microsatellites on average than standard subgenomic DNA library screens. The similarly short average repeat unit lengths between our enriched DNA library screen and that of Ross *et al.* (2003) in five *Drosophila* subgenus species is consistent with this hypothesis.

Second, sample size of microsatellites isolated from DNA library screens may influence estimates of average repeat unit length. The distribution of microsatellite repeat unit lengths in the genomes of many organisms follows a Poisson distribution with a long tail representing longer repeat unit lengths (Kruglyak *et al.*, 1998; Bachtrog *et al.*, 1999). As more samples of microsatellites are identified, the likelihood of sampling those with shorter repeat units increases. Thus, longer average microsatellite repeat unit lengths may be estimated from smaller samples than from larger samples. For example, Schlotterer & Harr (2000) estimated the average repeat unit length of dinucleotide repeats in *D. melanogaster* from 10 clones identified in a traditional subgenomic DNA library screen as 12·1 repeat units, whereas Schug *et al.* (1998*b*) estimated the average repeat unit length using similar techniques from 67 clones as 10·3 repeat units. Together with the screening procedure above, which tends to sample from the upper tail of the distribution of repeat unit lengths in standard subgenomic DNA library screens, and the larger number of random samples of microsatellites identified during the enriched subgenomic DNA libraries screen reported here and by Ross *et al.* (2003), samples from the latter may be biased towards shorter repeat units on average than the former.

We can only directly compare our results with those of Ross *et al.* (2003) who used the same enriched DNA library screening protocol because of the biases inherent in the sampling procedure and DNA library screening methods. In their study, the average repeat unit length for the five species examined is similar to that of *D. ananassae*. Because we identified very few microsatellites longer than 12 repeat units, we can be fairly certain that average repeat unit length of *D. ananassae* is not significantly longer than in other *Drosophila* species. However, we cannot be sure whether the shorter repeat unit length we estimate here is comparable to estimates from studies of other *Drosophila* species using different DNA library screening techniques. More accurate estimates of average repeat unit and frequency of microsatellites across the genome will make use of long stretches of DNA sequence, which are not yet available for *D. ananassae*.

https://doi.org/10.1017/S0016672303006542 Published online by Cambridge University Press

(ii) *Genetic variation*

We assayed genetic variation in a single population of *D. ananassae* from Java, Indonesia for 63 of the loci (Table 1). Of these, 62 (95%) were polymorphic and three were monomorphic (5%). Furthermore, in a separate study of additional populations, we assayed one of the monomorphic loci (*Dan11*) in a different population sample from India. Heterozygosity is high, suggesting that even the few monomorphic loci may be highly variable in populations of *D. ananassae* outside of Indonesia. Such loci that are monomorphic in one population but polymorphic in other populations may be located near or within a gene that was a target of selection. We can thus use such markers as potential indicators of chromosomal regions that may have been influenced by natural selection (Schlotterer & Wiehe, 1999). The high percentage of polymorphic microsatellites we have identified indicates that the enriched subgenomic DNA library screen is thus an excellent technique for identifying highly variable microsatellites.

Accurate estimates of genetic variation such as heterozygosity, variance in repeat unit and number of alleles commonly quantified for microsatellites in natural populations are sensitive to sample size (Pritchard & Feldman, 1996; Zhivotovsky *et al.*, 2001). The sequence of our isolation and screening procedure was used first to identify potential microsatellites by amplification of the cloned DNA fragment, and then to test for amplification in a small sample of individuals from Java, Indonesia. A subset of these (31 loci to date) were then assayed in a larger sample from the same population for more accurate measures of genetic variation. Our initial population screen is useful for identification of potentially useful DNA markers, but less useful for population genetic analysis. We thus restricted our analysis and discussion of genetic variation to the 31 loci for which we assayed at least 20 individuals (Table 1).

Heterozygosity ranges from 0 to 1 (mean $H=0{\cdot}51$), similar to estimates for *D. melanogaster* ($H=0{\cdot}61$; Schug *et al.*, 1998*b*) and *D. simulans* ($H=0{\cdot}64$; Hutter *et al.*, 1998), and substantially lower than for *D. subobscura* ($H=0{\cdot}77$; Pascual *et al.*, 2000), *D. pseudoobscura* ($H=0{\cdot}84$; Noor *et al.*, 2000) and *D. virilus* ($H=0{\cdot}70$; Schlotterer & Harr, 2000). Studies of mutation rate at microsatellite loci in yeast, *Drosophila*, mice and humans have demonstrated that repeat unit length is associated with mutation rate such that longer repeat unit lengths have higher mutation rates (e.g. Wierdl *et al.*, 1997; Kruglyak *et al.*, 1998, 2000). These results predict that longer microsatellites with higher mutation rates have higher levels of variation in natural populations than microsatellites with shorter repeat unit lengths, all other factors being equal. In *Drosophila*, such a relationship has been demonstrated in *D. melanogaster* (Schug *et al.*, 1998*b*) and *D. subobscura* (Pascual *et al.*, 2000), but did not exist for *D. pseudoobscura* (Noor *et al.*, 2000). For the Indonesian population of *D. ananassae* in this study, the correlation between all measures of genetic variation and both mean and maximum repeat unit length are positive and significant (Table 3). Maximum repeat unit length shows the strongest correlation with variance in repeat unit, as was also the case for *D. melanogaster* (Schug *et al.*, 1998*b*) and *D. subobscura* (Pascual *et al.*, 2000).

Table 3. *Spearman Rank Order correlations between measures of genetic variation and repeat unit length in population samples of* D. ananassae *from Jakarta, Indonesia*

	Mean	Max.
Variance in repeat number	0·36*	0·78**
Heterozygosity	0·48**	0·66**
No. of alleles	0·49**	0·71**

*$P<0{\cdot}05$, **$P<0{\cdot}01$.

It is widely believed that longer microsatellites have higher mutation rates than shorter microsatellites because the likelihood of DNA slippage during replication increases as the repeat unit length of a microsatellite increases (Ellegren, 2000). Our assays of microsatellite variation for loci with average repeat unit length of 8·75 are similar to levels of microsatellite variation with average repeat unit lengths of 10·3 and 10·1 for *D. melanogaster* and *D. simulans*, respectively. Kimura & Crow (1964) demonstrated theoretically the following relationship among heterozygosity (H), mutation rate (μ) and effective population size (N_e) assuming an infinite allele model, where each new mutation is a new length allele: $H=4N_e\mu/(1+4N_e\mu)$, where μ is the mutation rate of a new length allele per generation. For a stepwise mutation model, which incorporates mutations to previous mutant allele sizes, $H=1-[1/\sqrt{(1+8N_e\mu)}]$ (Ohta & Kimura, 1973). Though we do not have direct estimates of mutation rate for dinucleotide repeats in *D. ananassae*, if we assume they are similar to the mutation rate of *D. melanogaster*, we can estimate N_e by substituting our empirical estimates of heterozygosity and mutation rates inferred from *D. melanogaster* and solving for N_e. Using these estimates, the effective population size of *D. ananassae* is 142 648 (range = 0–2 661 290) for the infinite alleles model and 4 545 602 (range = 0–134 395 161) for the stepwise mutation model. Similar calculations can be used to estimate N_e based on variance in repeat unit length (V) using a stepwise mutation model. Slatkin (1995) demonstrated theoretically that $V=4N_e\mu$. Substituting the values of V from the Indonesian population and assuming a mutation rate of $9{\cdot}3\times10^{-6}$ leads to an estimate of

https://doi.org/10.1017/S0016672303006542 Published online by Cambridge University Press

$N_e = 304\,362$ (range = 0–2 917 473) for the Indonesian population. The estimates based on heterozygosity may be skewed towards a high value because of a single locus, *Dan9*, with a heterozygosity of 0·99. Because the relationship between heterozygosity and effective population size does not increase linearly, such high values of heterozygosity may inflate the average. Omitting this locus leads to average N_e estimates of 58 692 and 217 284 for the infinite alleles model and stepwise mutation model, respectively.

Estimates of N_e have been calculated using a similar method for a variety of *Drosophila* species (Schug *et al*., 1998*a*, *b*; Noor *et al*., 2000; Pascual *et al*., 2000; Wilder *et al*., 2002). Our estimates are close to those of the African population of *D. melanogaster*: 80 823 and 328 278 for the IAM and SMM, respectively (Schug *et al*., 1998*b*). Because the repeat unit length of the loci in our study of *D. ananassae* is shorter than the loci in the *D. melanogaster* study (mean RU = 10·3) the mutation rate of the *D. ananassae* microsatellites we assayed may be slightly lower than those assayed in the *D. melanogaster* study. The true N_e of the Indonesian *D. ananassae* population may thus be slightly higher, but is unlikely to be lower than the estimate from this study. The similar estimates of N_e for *D. ananassae* and *D. melanogaster* may reflect a similar history and cosmopolitan distribution.

(iii) *Tests for historic population fluctuations*

Ancestral populations of *D. ananassae* are believed to be from Southeast Asia or the South Pacific islands. Because *D. ananassae* is a human commensal, it may have colonized islands in Southeast Asia during the recent past accompanying humans. We thus examined the data from the Java, Indonesia population for evidence of a population bottleneck that may represent a colonization event on the island. We used the software Bottleneck (Cornuet & Luikart, 1996) to test for evidence of excess or deficiency of heterozygosity

Table 4. *Results of tests for heterozygosity*(*H*) *excess or deficiency for 30 dinucleotide repeat loci in a population sample of* Drosophila ananassae *from Jakarta, Indonesia using Bottleneck* (*Cornuet & Luikart, 1996*)

Statistical test	IAM	TPM	SMM
Sign test			
Expected no. of loci with H excess	16·9	17·04	17·25
Observed no. of loci with H excess	7	7	4
Observed no. of loci with H deficiency	22	22	25
P	0·0019	0·0002	0
Standardized differences test			
T2	−7·323	−10·485	−17·818
P	0	0	0
Wilcoxon test			
P (one-tailed for H deficiency)	0·0012	0·0001	0

One locus, *Dan11*, was omitted because it is monomorphic. Statistical tests were performed to test for deviations from the expected heterozygosity at mutation–drift equilibrium using an infinite alleles model (IAM), two-phase mutation model with 30% stepwise mutations and 70% infinite allele mutations (TPM), and a stepwise mutation model (SMM).

Table 5. *Cross-species PCR amplification of dinucleotide repeat microsatellites isolated and characterized from a* Drosophila ananassae *DNA library*

Locus *DAN*	11	12	20	26	27	31	32	33	40	45	59	69	70	71	73	76	78	79	83	120
Ananassae subgroup																				
ananassae complex																				
D. ananassae	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
D. pallidosa	+	+	+	+	+	+		+	+	+	+	+	+	+	+	+	+	+	+	+
D. varians															+					
bipectinata complex																				
D. bipectinata				+	+					+					+	+		+		
D. parabipectinata															+					
D. malerkotliana										+			+		+	+		+		
D. pseudoananassae					+					+					+	+		+		
D. ercepeae															+					
Elegans subgroup																				
D. elegans															+					
Eugracilis subgroup																				
D. eugracilu	+										+									
Ficusphila subgroup																				
D. ficusphila															+					

+represents amplification successful.

https://doi.org/10.1017/S0016672303006542 Published online by Cambridge University Press

from that expected if the population was in mutation–drift equilibrium. Heterozygosity excess is typical following a population bottleneck and heterozygosity deficiency is typical of a recent population expansion. The test is an extension of the Ewens–Waterson test originally developed for allozymes using coalescent simulations based on an infinite alleles model (Maruyama & Fuerst, 1985) and also incorporating a stepwise and two-phase mutation model.

For the 31 loci we assayed, the number of loci showing heterozygosity deficiency deviates significantly from expectations if the population is at mutation–drift equilibrium for all mutation models (Table 4). These data indicate that the Java, Indonesia population is in a stage of an expansion. Further studies of microsatellites and DNA sequences of single-copy nuclear genes and mitochondrial DNA of additional populations will be necessary to determine the extent to which the population expansion is a characteristic of the Indonesian population, or if it extends further through the geographic range.

(iv) *Cross-species amplification*

It is common to use primers developed in one species to amplify the homologous loci in other, closely related species for use in population genetic and phylogenetic analyses, and studies of genome evolution. We used the same genotyping PCR protocol to test for amplification of the homologous DNA fragments for 20 of the microsatellite loci in all species within the Ananassae, Elegans, Eugracilis and Ficusphila subgroups available in the Tucson fly stock centre. PCR fragments were separated on a 2% agarose gel and were considered positive if they amplified a single fragment identifiable as a sharp band. The phylogenetic relationship of these subgroups is not well established and the species status is based largely on morphological characteristics (e.g. Throckmorton, 1962). Most of the species exist in much smaller geographic regions in Asia than *D. ananassae*. *D. pallidosa* is a sister species identified in a single population from Samoa. There is chromosomal evidence that it may have hybridized with *D. ananassae* in nature (Futch, 1966).

Results of successful amplification across species is shown in Table 5. With the exception of only one locus (*Dan32*), the primers all amplify a DNA fragment in *D. pallidosa* similar in length to the DNA fragment in *D. ananassae*. Only one of the loci (*Dan73*) amplified in all species. Amplification was successful in the Bipectinata subgroup more often than in *D. varians*, a member of the Ananassae subgroup, suggesting that this species may be misclassified. Though it is difficult to estimate the exact size of the PCR products on an agarose gel, most of the PCR fragments amplified in species more distantly related than *D. pallidosa* were more than 40 bases different in length from *D. ananassae*. Such large differences in DNA fragment length most probably represent additional insertions or deletions in the flanking regions of the repeated unit. Thus, although amplification of homologous loci is possible in many cases, it is unclear whether the homologous region in other species contains microsatellites. Assays of potential genetic variation in these species will require sampling natural populations, as multiple individuals are not available from the *Drosophila* stock centre. Nevertheless, our data suggest that the majority of dinucleotide repeat microsatellite loci we report in this study will not be useful for species more distantly related than *D. ananassae* and *D. pallidosa*.

Our data demonstrate that highly variable dinucleotide repeat microsatellites are abundant in the genome of *D. ananassae*. Despite the short repeat unit length of the loci we isolated, more than 90% are polymorphic in a population from Java, Indonesia. The high level of variation in natural populations makes these ideal DNA markers for genetic mapping experiments. This large collection of DNA markers thus promises to be a valuable tool for genetic mapping studies, construction of physical maps using *in situ* hybridization to polytene chromosomes, and for studies of historic demography and population structure in natural populations.

We thank G. Acosta, K. Pate and J. Stafford for their help in the laboratory. E. Regulski was supported by an Undergraduate Research Award from the Office of the Provost at UNCG. This project was funded by a grant from the National Science Foundation.

References

Bachtrog, D., Weiss, S., Zangerl, B., Brem, G. & Schlotterer, C. (1999). Distribution of dinucleotide microsatellites in the *Drosophila melanogaster* genome. *Molecular Biology and Evolution* **16**, 602–610.

Chen, Y., Marsh, B. J. & Stephan, W. (2000). Joint effects of natural selection and recombination on gene flow between *Drosophila ananassae* populations. *Genetics* **155**, 1185–1194.

Cornuet, J. M. & Luikart, G. (1996). Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics* **144**, 2001–2014.

Dieringer, D. & Schlotterer, C. (2003). Microsatellite Analyzer (MSA): a platform independent analysis tool for large microsatellite data sets. *Molecular Ecology Notes* **3**, 167–169.

Dobzhansky, T. & Dreyfus, A. (1943). Chromosomal aberrations in Brazilian *Drosophila ananassae*. *Proceedings of the National Academy of Sciences of the USA* **29**, 301–305.

Ellegren, H. (2000). Microsatellite mutations in the germline: implications for evolutionary inference. *Trends in Genetics* **16**, 551–558.

Futch, D. G. (1966). A study of speciation in South Pacific populations of *Drosophila ananassae*. *University of Texas Publications* **6615**, 79–120.

https://doi.org/10.1017/S0016672303006542 Published online by Cambridge University Press

Hamilton, M. B., Pincus, E. L., DiFiore, A. & Fleischer, R. C. (1999). Universal linker and ligation procedures for construction of genomic DNA libraries enriched for microsatellites. *BioTechniques* **27**, 500–507. Protocol found at: http://bioserver.georgetown.edu/faculty/hamilton/

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.

Jarne, P. & Lagoda, J. L. (1996). Microsatellites, form molecules to populations and back. *Trends in Ecology and Evolution* **11**, 424–429.

Johnson, F. M. (1971). Isozyme polymorphisms in *Drosophila ananassae*: genetic diversity among island populations in the South Pacific. *Genetics* **68**, 77–95.

Kimura, M. & Crow, J. F. (1964). The number of alleles that can be maintained in a finite population. *Genetics* **49**, 725–738.

Kruglyak, S., Durrett, R. T., Schug, M. D. & Aquadro, C. F. (1998). Equilibrium distributions of microsatellite repeat length resulting from a balance between slippage events and point mutations. *Proceedings of the National Academy of Sciences of the USA* **95**, 10774–10778.

Kruglyak, S., Durrett, R. T., Schug, M. D. & Aquadro, C. F. (2000). Distribution and abundance of microsatellites in the yeast genome can be explained by a balance between slippage events and point mutations. *Molecular Biology and Evolution* **17**, 1210–1219.

Lynch, M. & Crease, T. J. (1990). The analysis of population survey data on DNA sequence variation. *Molecular Biology and Evolution* **7**, 377–394.

Maruyama, T. & Fuerst, P. A. (1985). Population bottlenecks and nonequilibrium models in population genetics. II. Number of alleles in a small population that was formed by means of a recent bottleneck. *Genetics* **111**, 675–689.

Noor, M. A. F., Schug, M. D. & Aquadro, C. F. (2000). Differentiation among *Drosophila pseudoobscura* populations at microsatellite loci. *Genetical Research* **75**, 25–35.

Ohta, T. & Kimura, M. (1973). A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. *Genetical Research* **22**, 201–204.

Prakash, R., Shamina, J. P. & Yadav, J. P. (1994). Evidence for alcohol dehydrogenase cline in natural populations of *Drosophila ananassae* from India. *Cell and Chromosome Research* **17**, 12–19.

Pascual, M., Schug, M. D. & Aquadro, C. F. (2000). High density of long microsatellites in *Drosophila subobscura*. *Molecular Biology and Evolution* **17**, 1259–1267.

Pritchard, J. K. & Feldman, M. W. (1996). Statistics for microsatellite variation based on coalescence. *Theoretical Population Biology* **50**, 325–344.

Ross, C. L., Dyer, K. A., Erez, T., Miller, S. J., Jaenike, J. & Markow, T. A. (2003). Rapid divergence of microsatellite abundance among species of *Drosophila*. *Molecular Biology and Evolution* **20**, 1143–1157.

Rozen, S. & Skaletsky, H. J. (2000). Primer3 on the WWW for general users and for biologist programmers. In *Bioinformatics Methods and Protocols: Methods in Molecular Biology* (ed. S. Krawetz & S. Misener), pp. 365–386. Totowa, NJ: Humana Press. (Code available at http://www-genome.wi.mit.edu/genome_software/other/primer3.html.)

Schlotterer, C. & Harr, B. (2000). *Drosophila virilus* has long and highly polymorphic microsatellites. *Molecular Biology and Evolution* **17**, 1641–1646.

Schlotterer, C. & Wiehe, T. (1999). Microsatellites, a neutral marker to infer selective sweeps. In *Microsatellites: Evolution and Applications* (ed. D. B. Goldstein & C. Schlotterer), pp. 238–248. New York: Oxford University Press.

Schug, M. D., Wetterstrand, K., Gaudette, M. S., Lim, R. H., Hutter, C. H. & Aquadro, C. F. (1998*a*). The distribution and frequency of microsatellites in *Drosophila melanogaster*. *Molecular Ecology* **7**, 57–69.

Schug, M. D., Hutter, C. M., Wetterstrand, K. A., Gaudette, M. S., Mackay, T. F. C. & Aquadro, C. F. (1998*b*). The mutation rate of dinucleotide repeat microsatellites in *Drosophila melanogaster*. *Molecular Biology and Evolution* **15**, 1751–1760.

Singh, B. N. (1998). Population genetics of inversion polymorphism in *Drosophila ananassae*. *Indian Journal of Experimental Biology* **36**, 739–748.

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

Stephan, W. (1989). Molecular genetic variation in the centromeric region of the X chromosome in three *Drosophila ananassae* populations. II. The *Om (1D)* Locus. *Molecular Biology and Evolution* **6**, 624–635.

Stephan, W. & Langley, C. H. (1989). Molecular genetic variation in the centromeric region of the X chromosome in three *Drosophila ananassae* populations. I. Contrasts between the *vermilion* and *forked* loci. *Genetics* **121**, 89–99.

Stephan, W. & Mitchell, S. J. (1992). Reduced levels of DNA polymorphism and fixed between-population differences in the centromeric region of *Drosophila ananassae*. *Genetics* **132**, 1039–1045.

Stephan, W., Xing, L., Kirby, D. A. & Braverman, J. M. (1998). A test of the background selection hypothesis based on nucleotide data from *Drosophila ananassae*. *Proceedings of the National Academy of Sciences of the USA* **95**, 5649–5654.

Throckmorton, L. (1962). The problem of phylogeny in the genus Drosophila. *University of Texas Publications* **6205**, 207–344.

Tobari, Y. N. (1993). *Drosophila ananassae – Genetical and Biological Aspects*. Tokyo/Basel: Japan Scientific Societies Press/Karger.

Tomimura, Y., Matsuda, M. & Tobari, Y. N. (1993). Polytene chromosome variations of *Drosophila ananassae* and its relatives. In *Drosophila ananassae – Genetical and Biological Aspects* (ed. Y. N. Tobari), pp. 139–151. Tokyo/Basel: Japan Scientific Societies Press/Karger.

Wiehe, T. H. E. & Stephan, W. (1993). Analysis of a genetic hitchhiking model, and its application to DNA polymorphism data from *Drosophila melanogaster*. *Molecular Biology and Evolution* **10**, 842–854.

Wierdl, M., Dominska, M. & Petes, T. D. (1997). Microsatellite instability in yeast: dependence on the length of the microsatellite. *Genetics* **146**, 769–779.

Wilder, J. A., Diaz, T., O'Neill, R. J. W., Kennedy, J. & Hollocher, H. (2002). Characterization and isolation of novel microsatellites from the *Drosophila dunni* subgroup. *Genetical Research* **80**, 177–185.

Zhivotovsky, L., Goldstein, D. B. & Feldman, M. W. (2001). Genetic sampling error of distance and variation in mutation rate among microsatellite loci. *Molecular Biology and Evolution* **18**, 2141–2145.

https://doi.org/10.1017/S0016672303006542 Published online by Cambridge University Press