Genet. Res., Camb. (2002), **80**, *pp.* 177–185. *With* 1 *figure.*
DOI: 10.1017/S0016672302005864 Printed in the United Kingdom

Characterization and isolation of novel microsatellites from the *Drosophila dunni* subgroup

J. A. WILDER[a]†, T. DIAZ[a], R. J. W. O'NEILL[a]‡, J. KENNEY[a]§ AND H. HOLLOCHER[b]*
[a] *Department of Ecology and Evolutionary Biology, Princeton University, Princeton NJ 08544, USA*
[b] *Department of Biological Sciences, University of Notre Dame, Notre Dame IN 46544, USA*

(*Received 21 February 2002 and in revised form 2 July 2002*)

Summary

We have isolated and characterized 77 novel microsatellites from two species, *Drosophila dunni* and *Drosophila nigrodunni*, which are closely related Caribbean-island endemics from the *Drosophila cardini* species group. These species are very distantly related to all other *Drosophila* from which microsatellites have previously been characterized. We find that the average length of microsatellites isolated in these species is quite small, with an overall mean length of 9·8 repeat units for dinucleotide microsatellites in the two study species. The nucleotide composition of dinucleotides differs between the two species: *D. nigrodunni* has a predominance of $(AC/GT)_n$ repeats, whereas *D. dunni* has equal numbers of $(AC/GT)_n$ and $(AG/CT)_n$ repeats. Tri- and tetranucleotide repeats are not abundant in either species. We assayed the variability of eight microsatellites in a closely related third species, *Drosophila arawakana*, using wild-caught individuals from the island of Guadeloupe. We found the microsatellites to be extremely variable in this population, with observed heterozygosities ranging from 0·541 to 0·889. DNA amplification trials suggest that these eight microsatellites are widely conserved across the *D. cardini* group, with five of the eight producing amplification products in every species tested. However, the loci are very poorly conserved over greater phylogenetic distances. DNA amplification of the microsatellite loci was unreliable in members of the closely related *Drosophila quinaria*, *Drosophila calloptera*, *Drosophila guarani* and *Drosophila tripunctata* species groups. Furthermore, these microsatellites could not be detected in the genome of *Drosophila melanogaster*, despite the conservation of microsatellite flanking regions at some loci. These data indicate that *Drosophila* microsatellite loci are quite short lived over evolutionary timescales relative to many other taxa.

1. Introduction

Microsatellites are hypervariable DNA sequences that are composed of tandem arrays of short nucleotide motifs. These repetitive sequences are highly polymorphic in their repeat number, making them one of the most widely used genetic markers for studies of population structure, gene mapping and parentage analysis (reviewed in Schlötterer & Pemberton, 1994; Jarne & Lagoda, 1996; Goldstein & Schlötterer, 1999).

* Corresponding author. Tel: +1 574 631 4765. e-mail: Hope.Hollocher.1@nd.edu
Current addresses:
† Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721, USA.
‡ Department of Molecular and Cell Biology, University of Connecticut, Storrs, CT 06269, USA.
§ Department of Biology, Washington University, St Louis, MO 63130, USA.

Microsatellites are ubiquitous among eukaryotes and have been described from a great variety of taxa. They are particularly well documented among *Drosophila* species, with studies primarily focusing on members of the *Drosophila melanogaster*, *Drosophila obscura* and *Drosophila virilis* groups. Of these, the microsatellites of *D. melanogaster* are the best characterized. In this species, both standard DNA library screens (e.g. England *et al.*, 1996; Schlötterer *et al.*, 1997; Harr *et al.*, 1998; Schug *et al.*, 1998*b*) and screens of publicly available sequence data (e.g. Goldstein & Clark, 1995; Michalakis & Veuille, 1996; Schug *et al.*, 1998*b*; Bachtrog *et al.*, 1999) have been used to document the characteristics of microsatellites in the *D. melanogaster* genome. From these studies, *D. melanogaster* has been described as having a high frequency of dinucleotide microsatellites relative to tri-, or tetranucleotide

repeats (Schug *et al.*, 1998 *b*). Of the dinucleotides, $(AC/GT)_n$ repeats are much more frequent than $(AG/CT)_n$, $(AT/TA)_n$ or $(CG/GC)_n$ repeats (England *et al.*, 1996; Schug *et al.*, 1998 *b*), a trait that is shared by many taxa, including some fish (Estoup *et al.*, 1993) and many mammals (e.g. Stallings *et al.*, 1991; Beckman & Weber, 1992; Ellegren, 1992). In addition, the microsatellites of *D. melanogaster* contain fewer repeat units than those from most other taxa (Kruglyak *et al.*, 1998; Schug *et al.*, 1998 *b*; Bachtrog *et al.*, 1999), which might significantly affect the level of variation because repeat length at a locus is positively correlated with the mutation rate (Jin *et al.*, 1996; Wierdl *et al.*, 1997; Schlötterer *et al.*, 1998; Schug *et al.*, 1998 *a*). Recent studies of other *Drosophila*, including *Drosophila simulans* (Hutter *et al.*, 1998), *Drosophila pseudoobscura* (Noor *et al.*, 2000), *Drosophila subobscura* (Pascual *et al.*, 2000) and *D. virilis* (Schlötterer & Harr, 2000), show that these species all have the same predominance of dinucleotide microsatellites over tri-, and tetranucleotides, and the same higher frequency of $(AC/GT)_n$ repeat motifs relative to $(AG/CT)_n$, $(AT/TA)_n$ or $(CG/GC)_n$ repeats. However, these species all have somewhat longer microsatellites than *D. melanogaster*, and those of *D. virilis* also appear to be more variable (Schlötterer & Harr, 2000).

In this study, we report the results of the isolation and characterization of novel microsatellites from two additional species, *Drosophila dunni* and *Drosophila nigrodunni*, both members of the *Drosophila cardini* species group. These taxa are distantly related to all of the *Drosophila* species for which microsatellites have thus far been characterized. *D. virilis*, the only member of the *Drosophila* subgenus from which microsatellites have been described, is the most closely related to the *D. cardini* group, although the *virilis–repleta* radiation diverged from the remainder of the subgenus relatively early in the *Drosophila* radiation (Throckmorton, 1975), perhaps as long as 46 million years ago (Beverly & Wilson, 1984). The *D. cardini* group contains 19 species and subspecies restricted to tropical and subtropical climates in the new world (Heed & Russell, 1971). This group includes the *D. dunni* subgroup, which contains ten species and subspecies that are restricted to the islands of Eastern Caribbean (Heed & Krishnamurthy, 1959). The two species that we focus on, *D. dunni dunni* and *D. nigrodunni*, are members of this subgroup and are endemic to the islands of Puerto Rico and Barbados, respectively. It is our goal to develop microsatellite loci that can be used to describe the population structure of various members of the *D. dunni* subgroup. Our work with these species is part of our larger aim to elucidate the evolutionary history of these endemic island taxa and to understand the genetics of their remarkable interspecific cline in abdominal pigmentation (e.g. Hollocher *et al.*, 2000 *a*, 2000 *b*). Here, we have developed novel microsatellites from these species in order: (1) to compare the genomic composition of microsatellites from the *D. dunni* subgroup with that of other *Drosophila* taxa; (2) to assay microsatellite variation in a natural population of a third member of the *D. dunni* subgroup; and (3) to analyze the amplification properties of the novel microsatellites across the entire *D. cardini* group, as well as members of closely related species groups, in order to assess the breadth of utility of these genetic markers.

2. Methods

(i) Drosophila *strains*

The following representatives of the *D. cardini* species group were obtained from National *Drosophila* Species Resource Center in Bowling Green, Ohio, and used both in the creation of microsatellite libraries and to assess DNA amplification across the *D. cardini* species group: *Drosophila acutilabella* (15181-2171.2), *Drosophila antillea* (15182-2251.0), *Drosophila arawakana arawakana* (15182-2261.0), *Drosophila belladunni* (15182-2271.0), *D. cardini* (15181-2181.9), *Drosophila caribiana* (15182-2281.0), *D. dunni dunni* (15182-2291.0), *Drosophila neocardini* (15181-2201.0), *D. nigrodunni* (15182-2311.1) *Drosophila polymorpha* (15181-2231.2) and *Drosophila procardinoides* (15181-2241.0). Specimens of *Drosophila similis similis* were taken from the isofemale line Vermont 15(C) collected by H. Hollocher in July 1996 on the island of St Vincent.

The amplification of microsatellite loci from outside the *D. cardini* species group was performed using the following strains obtained from the National *Drosophila* Species Resource Center in Bowling Green, Ohio: *Drosophila ornatipennis* (*Drosophila calloptera* group, 15160-2121.0), *Drosophila guarani* (*D. guarani* group, 15172-2151.1) and *Drosophila crocina* (*Drosophila tripunctata* group, 15220-2341.0). Additionally, *Drosophila deflecta* (*Drosophila quinaria* group) was tested using a strain collected from Princeton, NJ, in August 1999 by J. Wilder and E. Dyreson.

The variability of the novel microsatellites was assessed in the species *D. arawakana arawakana*. We scored 25 individuals from each of two collecting sites (Matouba and L'Ermitage) on the island of Guadeloupe, French West Indies. Specimens were collected in the wild by J. Wilder during June and July of 1999. Flies were caught using a bait of rotting *Cucurbita moschata* (tropical pumpkin) and stored in 70% ethanol in the field.

(ii) *Microsatellite isolation from* D. nigrodunni *and* D. dunni dunni

We isolated microsatellites from two species, *D. nigrodunni* and *D. dunni dunni*, by separately screening

genomic libraries from each species using tandemly repetitive oligonucleotide probes, following a protocol similar to that initially described by Tautz & Renz (1984). Except where noted, the isolation procedure was identical for each of the species that we screened in this experiment. We extracted DNA from a pool of ~200 starved male and female flies. The DNA was digested with *Sau*3AI and *Rsa*I and then cloned into the *Bam*HI and *Hinc*II sites of the pBluescript II KS (+/−) cloning vector (Stratagene). Before cloning, a portion of the digested *D. nigrodunni* DNA was size selected for 200–600 bp fragments by electrophoresis and purification from a 1% agarose gel. Both this size-selected portion of the *D. nigrodunni* genomic library, and the non-size-selected *D. nigrodunni* and *D. dunni dunni* genomic libraries were transformed into competent DH5 *Escherichia coli* cells (Gibco).

Cells from each of the genomic libraries were fixed to membranes using two different techniques. The size-selected portion of the *D. nigrodunni* library was lifted directly onto Magna nitrocellulose membranes. For the remainder of the *D. nigrodunni* library, and the entire *D. dunni dunni* library, colonies were picked at random from the plates and individually placed into wells containing 100 μl TB on a 96-well microtiter plate. After overnight growth, cells were transferred from the microtiter plates to Hybond N+ membranes using a dot-blot apparatus. Cells were fixed to the membranes in 0·4 M NaOH.

Once affixed to membranes, the libraries were probed for di-, tri- and tetranucleotide repeats. Membranes were allowed to hybridize with 20 ng of each of the following probes, which were end-labeled with [^{33}P]-dATP: $(AG)_{15}$, $(AC)_{15}$, $(CAG)_{10}$, $(CGG)_{10}$, $(AAAC)_7$, and $(AAAT)_7$. Overnight hybridizations of di-, tri-, and tetranucleotides were each performed separately in 30 ml of 55 °C Church's Buffer. After hybridization, membranes were washed for 30 min in 55 °C 2 × SSC/0·1% SDS solution and then exposed for 48 h on BioMax MR autoradiography film (Kodak). Oligonucleotides were stripped from each membrane between hybridizations by a 30 min wash in boiling 0·5% SDS. Inserts from positive colonies were sequenced with pBluescript T3 and T7 primers using either an AmpliCycle manual sequencing kit (Perkin Elmer) or the ABIPRISM 377 automated DNA sequencer (Perkin Elmer) maintained by the Princeton University synthesizing /sequencing facility.

(iii) *Microsatellite primer design and DNA amplification conditions*

We designed primers in microsatellite flanking regions using the program Primer 3.0 (http://www-genome.wi.mit.edu/genome–software/other/primer3.html). Primers were designed to produce DNA amplification products 100–300 bp long. Primer conditions were optimized in the species of origin for each microsatellite. Conditions were optimized by varying amplification conditions until only a single clear band was visible when the products were electrophoresed on a 2% agarose gel. Template DNA for each amplification was isolated from males using single-fly squish preps (Gloor *et al*., 1993). Each 10 μl amplification reaction contained 200 μM dNTPS, 1 l 10 × buffer, 1 U T_{AQ} polymerase, 1–1·5 μl from a 50 μl squish DNA preparation and 1·0 μM of each primer. DNA amplification reactions consisted of an initial incubation at 94 °C for 3 min, followed by 30–35 cycles of 45 sec at 94 °C, 1 min at 52–55 °C (Table 1) and 30 sec at 72 °C.

Heterozygosity at each locus was assessed by scoring the amplification products from wild-caught *D. arawakana arawakana* specimens. Primer concentrations were altered to contain 0·5 μM upper primer labeled with [^{33}P]-dATP, 0·5 μM unlabelled upper primer, and 1·0 μM unlabelled lower primer. Amplification products were separated by electrophoresis on a 6% polyacrylimide gel. Gels were visualized by exposure to BioMax MR film (Kodak) for 1–4 days.

(iv) *Cross-species microsatellite utility*

Each primer pair was tested in each of the 12 species from the *D. cardini* group listed above. Although this set of species does not constitute the entire *D. cardini* group, representatives from each major clade were used in this analysis (Heed & Russell, 1971; Hollocher *et al*., unpublished data). DNA amplifications for each species were performed at the temperatures listed in Table 1. The resulting products were sized on a 2% agarose gel using a 100 bp ladder (Promega). Positive results were considered to be any DNA amplification that produced a product within 50 bp of the fragment size from the focal species. Primer pairs were also tested using the same protocol for members of the *D. calloptera*, *D. guarani*, *D. tripunctata* and *D. quinaria* species groups as listed above.

The complete sequences of the clones containing the eight microsatellites for which we designed primers were also screened against the *D. melanogaster* genome using a BLAST search in order to check for conservation of microsatellites or microsatellite flanking regions in this species.

3. Results

(i) *Microsatellite isolation and characterization from the* D. dunni *species group*

Of the 1056 *D. nigrodunni* clones and 768 *D. dunni dunni* clones that we screened in our genomic libraries, 134 (12·7%) and 95 (12·4%), respectively, showed homology to one or more oligonucleotide probes. We

Table 1. *Characteristics of microsatellites isolated and successfully amplified in this study. Repeat motif is based on the sequence from the species of origin, as is the PCR fragment size. Number of alleles and expected* (H_e) *and observed* (H_o) *heterozygosity were estimated from 50 wild-caught* D. arawakana *individuals. Loci isolated from* D. dunni dunni *start with the prefix 'D', whereas loci isolated from* D. nigrodunni *start with the prefix 'ND'*

Locus	GenBank Accession	Repeat motif	Size (bp)	Melting point (°C)	No. of alleles	H_o	H_e	Primer sequences (5′→3′)
ND10	AF453651	$(AC)_{13}$	129	55·0	7	0·818	0·806	GTCAAAAGCGTTTAGTGTGTGG
								GCTTTGTGTCAGTCAGTTTTGT
ND43	AF317292	$(AC)_9$	211	54·5	13	0·541	0·589	CCGAAAGAAGATAAGAAAGGAAA
								GAGAGTAATGGAGAGCAGCAAAC
ND46	AF453653	$(AC)_{10}$	122	55·0	5	0·654	0·637	CACAACGTGGCGTATGAATAAT
								CGACACAGCGTTTTACATGAAT
ND21	AF453654	$(AC)_{12}$	189	53·5	17	0·769	0·853	CTGTTGTTATTAGCCTCTGAGC
								GATGTTGTGTTAAAAAGGTTTCTG
D7B2	AF453656	$(AG)_8TG(AG)_5$	282	54·0	12	0·659	0·821	GGCATTTCATTTGTTGCCTT
								AAATTTCCACCACCTTTCCC
D4G6	AF453652	$(AG)_9$	181	52·5	13	0·857	0·862	GCAGAAACCAAATAGATACAGGG
								GTTGTCCTCCCTTGGTCACTTTG
D3F11	AF453657	$(GA)_8AA(GA)_{20}$	232	52·1	11	0·889	0·941	CGCTGACCAAATCCAAGTGC
								CAAGAAACCAAAATAGGCATTC
ND9F7	AF453655	$(AC)_3AA(AC)_8$	256	54·8	20	0·800	0·915	TAGGCAGGTAAACAGACAGG
								GTTTTCATCGCCAGGACTC

sequenced 40 positive clones from *D. nigrodunni* and 42 positive clones from the *D. dunni dunni* library. Sequence data were obtained first from clones that hybridized strongly with di-, tri- and tetranucleotide repeat types, followed by those that hybridized with at least two of the three types, and finally with those that hybridized strongly with dinucleotide repeats. The sequence data showed that 25 positive clones from each species contained at least one microsatellite with a repeat length of four or more, making the overall efficiency of the screens ~60% in each species. The efficiency of the subset of *D. nigrodunni* clones that we size-selected to contain only genomic inserts from 200–600 bp showed a similar efficiency of 15 out of 24 clones (62·5%). However, microsatellites isolated from these size-selected clones tended to be near the ends of the inserts, limiting our ability to design usable primers in many cases.

Several clones contained more than one microsatellite, giving us a total of 77 microsatellites (45 from *D. nigrodunni* and 32 from *D. dunni dunni*). Of these, 57 (74·0%) were dinucleotide repeats, 12 (15·6%) were trinucleotides and eight (10·4%) were tetranucleotides (Table 2). Of the eight tetranucleotides isolated, two were found within *mini-me* retroposons (Wilder & Hollocher, 2001), limiting their use as unique genomic markers. The lengths of microsatellites that we isolated among *D. dunni dunni* and *D. nigrodunni* were relatively short, with mean lengths of 9·9, 4·5 and 8·4 repeat units for di-, tri- and tetranucleotides, respectively. In *D. nigrodunni*, 26 of 33 (78·8%) dinucleotide microsatellites were composed of the motif $(AC/TG)_n$, whereas six (18·2%) were $(AG/TC)_n$, and two (6·1%) were $(AT/TA)_n$. *D. dunni dunni* dinucleotides were distributed more evenly among repeat motifs, with nine of 25 (36·0%) $(AC/TG)_n$, nine (36·0%) $(AG/TC)_n$, six (24·0%) $(AT/TA)_n$ and one (4·0%) $(CG/GC)_n$.

(ii) *Microsatellite variability*

We developed DNA amplification primers for eight microsatellites isolated from separate clones in the genomic libraries of *D. nigrodunni* and *D. dunni dunni* (Table 1). Primers were developed for three more loci, but these were rejected because they did not produce amplification products that could be unambiguously scored in *D. arawakana arawakana*. Each of the loci that we scored was found to be highly polymorphic, with a range of five to 20 alleles (mean of 12·3 alleles) in the 50 individuals sampled from Guadeloupe. The observed heterozygosities (H_o) of each locus were high, ranging from 0·59 to 0·89, with a mean of 0·76. Variability at each of these loci showed few obvious deviations from a basic stepwise mutation pattern. Two loci, ND9F7 and ND21, each had a single allele of

Table 2. *The number and mean repeat lengths of microsatellites isolated in this study. Average dinucleotide lengths are also provided for other* Drosophila *species for which data are available*

Species	No. double repeats	Mean double-repeat length	No. triple repeats	Mean triple-repeat length	No. quadruple repeats	Mean quadruple-repeat length
Drosophila nigrodunni	33	10·8	8	5·0	4	8·5
Drosophila dunni dunni	25	8·6	4	3·6	4	8·3
*Drosophila melanogaster**	41	10·1	–	–	–	–
Drosophila pseudoobscura†	35	11·7	–	–	–	–
Drosophila subobscura‡	96	14·9	–	–	–	–
Drosophila virilis§	26	12·7	–	–	–	–

* From Schug *et al.* (1998 *b*).
† From Noor *et al.* (2000).
‡ From Pascual *et al.* (2000).
§ From Schlötterer & Harr (2000).

a size class inconsistent with the addition or subtraction of dinucleotide repeats. These odd-sized alleles each appeared at a low frequency (only once out of the 50 individuals sampled). The odd-sized allele at locus ND9F7 is unusually large, indicating that it might be the product of an insertion mutation in the microsatellite flanking region. Locus D3F11 showed several deviations from a stepwise mutation pattern, with many odd-sized alleles and several unusually large alleles, indicating either non-stepwise mutations in the repeat tract or size polymorphisms in the microsatellite flanking regions.

We use the mean estimated value of H_o (H) across loci to predict the effective population size (N_e) of *D. arawakana arawakana* from the island of Guadeloupe. Under the stepwise model of microsatellite mutation (SMM; Ohta & Kimura, 1973), the relationship between H and N_e is: $1-H=1/\mu(1+8N_e\mu)$. The per generation mutation rate (μ) of microsatellites has been empirically estimated in *D. melanogaster* at $9{\cdot}3\times10^{-6}$ for dinucleotide repeats (Schug *et al.*, 1998 *b*). Assuming that this is the mutation rate in *D. arawakana* as well, we estimate N_e to be $\sim2{\cdot}3\times10^5$. Under an infinite-alleles model (IAM) of microsatellite mutation (Kimura & Crow, 1964), the relationship of N_e to H is $N_e=H/[4(1-H)]$. Under this model, our estimate of N_e is $\sim8{\cdot}5\times10^4$.

(iii) *Cross-species microsatellite utility*

The eight microsatellites that we developed from *D. nigrodunni* and *D. dunni dunni* performed extremely well in tests of cross-species utility within the *D. cardini* species group. Although we could not measure population-level variability in other species, we did check for the presence of DNA amplification products of the expected length in each member of the *D. dunni* subgroup and in representatives of each of the more distantly related clades of the *D. cardini* group (Fig. 1). Each locus amplified successfully in all members of the *D. dunni* subgroup, with the exception of microsatellite ND21, which produced no DNA amplification product in *D. dunni dunni*. When the primer pairs were extended for use among more distantly related species within the group, DNA amplification products were reliably produced in every species for five of the eight microsatellites tested. One microsatellite locus, D7B2, failed to amplify in one species, *D. polymorpha*. D9F7 produced an amplification product in all species but *D. belladunni*, *D. acutilabella*, and *D. polymorpha*. Locus ND21 did not perform well outside the *D. dunni* subgroup (or in *D. dunni dunni*, as discussed above), failing to amplify in *D. dunni dunni*, *D. belladunni*, *D. acutilabella*, *D. procardinoides*, *D. neocardini*, *D. polymorpha* and *D. cardini*.

Outside the *D. cardini* species group, the apparent utility of the microsatellites became much less. In general, microsatellites did not amplify consistently in any of the test species. Only two loci amplified successfully in the *D. quinaria* and *D. calloptera* representatives, three loci amplified in *D. guarani* and five loci amplified in the *D. tripunctata* representative. The microsatellite loci varied considerably in their cross-group utility. One locus, ND46 worked broadly in every species surveyed. Two loci, ND10 and ND43a failed to amplify in any non-*cardini*-group species. The remaining five loci amplified successfully in either one or two of the test species, as detailed in Fig. 1.

To check for the presence of the microsatellites over even greater phylogenetic distances, we performed BLAST searches of each of the eight microsatellite-containing sequences against the entire *D. melanogaster* genome. Only two cases, loci D4G6 and ND10, identified similar sequence regions in *D. melanogaster* (both had E values $<10^{-9}$). Locus D4G6 corresponds to an intron region of the *nemo* gene on chromosome 3L (map region 66A22-66B5) and locus ND10 corresponds to an unannotated region near the predicted gene *CG14247* on chromosome 3R (map region

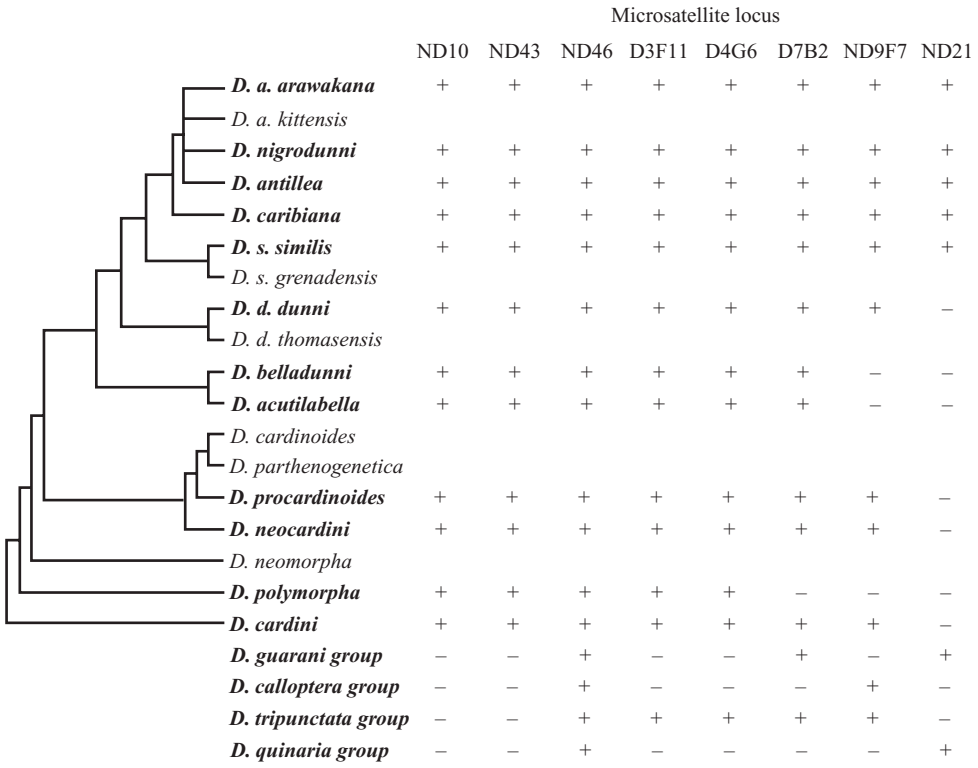

Fig. 1. The results of amplification trials for 12 species from the *Drosophila cardini* species group and four species from the *Drosophila calloptera* (represented by *Drosophila ornatipennis*), *Drosophila guarani* (represented by *Drosophila guarani*), *Drosophila tripunctata* (represented by *Drosophila crocina*) and *Drosophila quinaria* (represented by *Drosophila deflecta*) species groups. The phylogeny of the *cardini* species group is shown on the left (Hollocher *et al*., unpublished data), with species used in this analysis shown in bold type. Although all taxa are members of the same *Drosophila* lineage (Throckmorton, 1975), the relationships between groups are uncertain.

97C3-97C4). At both loci, it was possible to align portions of the microsatellite flanking regions on either side of the repeat tract. These conserved regions retained approximately the same spacing as the microsatellite flanking regions in the *D. dunni* subgroup but the microsatellite itself was completely absent in *D. melanogaster*. The portions that appeared conserved between *D. melanogaster* and the test species were generally non-repetitive in nature, indicating that regions with lower overall levels of DNA slippage might be more evolutionarily stable and thus more likely to be conserved among taxa.

4. Discussion

We have developed a set of highly polymorphic microsatellite markers from the *D. cardini* species group. The results of our microsatellite screens of *D. dunni dunni* and *D. nigrodunni* show that the overall composition of microsatellites in these species is similar to that described for *D. melanogaster*. Microsatellites in these species tend to be quite short, with the average length of dinucleotide microsatellites for the two species isolated in our study (9·8 repeat units) being only slightly less than those isolated using similar methods from *D. melanogaster*. As shown in Table 2, other *Drosophila* species appear to have longer microsatellites, with dinucleotide repeats isolated from *D. subobscura* (Pascual *et al*., 2000) averaging 50% larger than those isolated in this study. Despite this variability, all *Drosophila* examined so far and, indeed, most insects in general, have relatively short microsatellites relative to other taxa (reviewed in Schug *et al*., 1998 *b*). For *D. melanogaster*, the difference has been attributed to a low mutation rate, estimated at $9{\cdot}3 \times 10^{-6}$ (Schug *et al*., 1998 *a*), which is up to three orders of magnitude lower than estimates for other organisms (e.g. Dallas, 1992; Weber & Wong, 1993; Ellegren, 1995). Because the microsatellites that we have isolated from the *D. cardini* species group share similar size characteristics to those from *D. melanogaster*, they probably also share this low mutation rate.

The nucleotide composition of microsatellites in *D. nigrodunni* and *D. dunni dunni* also appear to be generally similar to that described in other *Drosophila*. In all other *Drosophila* species from which microsatellites have been isolated, dinucleotides are far more common than either tri- or tetranucleotides. Among dinucleotides, $(AC/GT)_n$ and $(AG/CT)_n$ account for

the great majority of microsatellites identified through *Drosophila* DNA library screens, with $(AC/GT)_n$ dinucleotides being many times more common that $(AG/CT)_n$. Screens of GenBank data have also shown that $(AT)_n$ repeats are common in *D. melanogaster*, but it is thought that the self-complementary nature of $(AT)_n$ oligonucleotide probes inhibits their isolation from plasmid libraries (Schug *et al*., 1998*b*). Our screens of *D. dunni dunni* and *D. nigrodunni* DNA libraries showed the same predominance of dinucleotide repeats, accounting for 75% of the microsatellites identified. In the case of *D. nigrodunni*, we also observed the same bias seen in other *Drosophila* species in favor of $(AC/GT)_n$ microsatellites. *D. dunni dunni*, however, showed a departure from this pattern. In this species, we isolated equal numbers of $(AC/GT)_n$ and $(AG/CT)_n$ microsatellites (nine of each repeat type). It is unlikely that the observed difference in the frequency of these two repeat types between the two species in our study is due to chance (*G* test, $P=0{\cdot}022$), nor were there any procedural differences that might have affected our library screens. Therefore, it appears that $(AG/CT)_n$ microsatellites are relatively more abundant in *D. dunni dunni* than in other *Drosophila* species. Bachtrog *et al*. (2000) have shown that $(AG/CT)_n$ microsatellites have a mutation rate of 0·71 relative to $(AC/GT)_n$ microsatellites in *D. melanogaster*. It is interesting to consider that the relatively smaller size of dinucleotide microsatellites isolated from *D. dunni dunni* (Table 2) might be influenced by the higher proportion of $(AG/CT)_n$ microsatellites.

It was our goal to develop microsatellite loci that would be useful across the entire *D. dunni* species radiation and also broadly useful in the *D. cardini* species group. The microsatellites that we present here generally fit the criteria of working universally across the *D. dunni* subgroup. In only one case did a primer pair fail to work in a species from the subgroup (ND21 in *D. dunni dunni*). In addition, we wanted to ensure that the microsatellites that we identified were polymorphic in as many species as possible. To this end, we assayed the variability of our loci in *D. arawakana*, rather than in one of the species of origin of the microsatellites. Many studies of microsatellite variability across taxa are subject to what is known as an 'ascertainment bias', where microsatellites are more variable in the species from which they are derived than in other species (Ellegren *et al*., 1995; Forbes *et al*., 1995; Rubinsztein *et al*., 1995; Goldstein & Pollock, 1997; Hutter *et al*., 1998). Although we could not mitigate this effect entirely (because part of our criterion in choosing microsatellites was that they have relatively long repeat tracts), we assayed polymorphism in a third species in order to minimize the probability of isolating loci that were variable in only the species of origin. Our study found that all eight microsatellites repeats for which we designed primers were highly polymorphic in *D. arawakana arawakana*, indicating that they probably have broad utility across the subgroup.

Based on the observed levels of heterozygosity at the eight loci analyzed in *D. arawakana arawakana*, we estimate the effective population size of this species on the island of Guadeloupe to be between $8{\cdot}5\times10^4$ and $2{\cdot}3\times10^5$. Because the microsatellites isolated in this study do not behave exactly as predicted by the IAM or SMM, as evidenced by the many odd-sized alleles at locus D3F11, a true estimate of N_e might not be exactly described by either value. At present, there are no other genetic data to corroborate this estimate for *D. arawakana arawakana*, but this method of estimating N_e using microsatellite data has been explored for several *Drosophila* species, including *D. pseudoobscura* (Noor *et al*., 2000), *D. subobscura* (Pascual *et al*., 2000) and *D. melanogaster* (Schug *et al*., 1998*a*). Estimates from these species have each produced results that are similar to independent estimates derived from single-copy nuclear genes, indicating that microsatellites produce reliable estimates of effective population size. Our values of N_e are much smaller than those from any of the other *Drosophila* species, which might be explained by the widespread, weakly structured distribution of these other species compared with the restricted range and spatially structured distribution of *D. arawakana arawakana* on the islands of Guadeloupe and Montserrat. The lower values of N_e might also reflect population bottlenecks associated with island colonization events, recent volcanic activity or hurricane-related disturbances. Further work using these microsatellites will help to characterize the population-genetic parameters of *D. arawakana arawakana* and other island endemics from the *D. cardini* species group.

Although the microsatellites that we have developed will clearly be useful for many applications within the *D. cardini* species group, our study indicates that these microsatellites are poorly conserved in species outside the group. The results of our DNA-amplification test show that the loci that we have developed are only sporadically amplified in species from closely related groups. The *D. cardini* group and three of the test groups (*D. guarani*, *D. calloptera* and *D. tripunctata*) are closely related members of the *tripunctata* radiation, each with largely neotropical distributions (Throckmorton, 1975). The fourth group (*D. quinaria*) is not a member of the *tripunctata* radiation but is still in the same major evolutionary lineage as the other groups tested (Throckmorton, 1975). Despite the very close relationships between these taxa, it does not appear that microsatellites developed here will be widely useful at these deeper phylogenetic levels. Based on our results, it also appears that microsatellite utility cannot necessarily be predicted on the basis of phylogenetic distance alone. As shown in Fig. 1, some

microsatellites appear to be lost in lineages that are closely related to the *D. dunni* subgroup, such as locus ND9F7 in *D. belladunni* and *D. acutilabella*, but retained in more distantly related species. Thus, microsatellite conservation (or at least the conservation of microsatellite priming sites) might decline in a relatively stochastic manner, such that distantly related species generally share fewer useful microsatellite loci, but the decrease in shared sites is not always proportional to phylogenetic distance.

In addition to a lack of amplification success in closely related species groups, screening the microsatellites we describe here against the genome of *D. melanogaster* reveals that the microsatellite repeat tracts can be entirely absent in more distantly related *Drosophila* species. Even in cases where homologous flanking regions could be identified between *D. melanogaster* and the *D. cardini* group, the microsatellite tracts themselves were not shared between species. The lack of microsatellite conservation between members of the *D. cardini* group and *D. melanogaster* is not entirely expected based on previous studies of microsatellite conservation across taxa. The *Drosophila* and *Sophophora* subgenera (which, respectively, contain the *D. cardini* group and *D. melanogaster*) diverged approximately 60 million to 65 million years ago (Berverly & Wilson, 1984; Spicer, 1988), a timescale that does not preclude the existence of conserved polymorphic microsatellites. Variable microsatellite loci have been found to be conserved in a variety of taxa, such as polistine wasps, turtles and some fish, for as long as 144–400 million years (Fitzsimmons *et al*., 1995; Rico *et al*., 1996; Ezenwa *et al*., 1998). Compared to these cases, it appears that *Drosophila* microsatellites are much more short-lived, even when the short generation time of *Drosophila* is accounted for.

The short lifespan of microsatellites in the *Drosophila* genome might be due to overall selection on genome size and content. For instance, it has been shown that the *Drosophila* genome has a high rate of DNA loss relative to other organisms owing to selection for relatively small genome size (Petrov & Hartl, 1997; Petrov *et al*., 2000). Although no studies have shown microsatellites to be specific targets for removal as 'junk DNA', their proclivity for expansion might make their excision beneficial when selection favors a small genome. Thus, their frequent removal from the genome might cause the lack of conservation among taxa that we observe. In addition to their short persistence time, *Drosophila* microsatellites appear to be unstable with regard to repeat-motif content. Studies from the *D. obscura* species group have shown that the composition of microsatellite repeat tracts can change quite rapidly among very closely related *Drosophila* species (Noor *et al*., 2001). Together, the observations of a short lifetime in the genome and instability with respect to repeat-tract composition indicate that the mutational processes affecting the lifespan of *Drosophila* microsatellites might be different in many respects from that of other organisms.

Studies of the genus *Drosophila* have provided great insights into the variation in microsatellite content that exists between closely related species. Differences between *Drosophila* species appear to be relatively modest. Although variation exists, microsatellites tend to be quite short and evolutionarily unstable. By contrast, many other taxa have much longer microsatellites that can persist for many millions of years. These differences indicate that the processes governing microsatellite evolution vary considerably over evolutionary timescales and can greatly affect the genomic content of repetitive DNA.

We thank two anonymous reviewers for comments that greatly improved the manuscript. Additionally, we thank A. Sainz, B. Heller, E. Dyreson and E. Schielke for assistance collecting samples, and A. Sainz for helpful comments on the manuscript. This project was funded by awards from Princeton University to T.D., J.K. and H.H., and awards from the Alfred P. Sloan Foundation, the Henry Luce Foundation, the University of Notre Dame, and the National Science Foundation to H.H. In addition, the Program in Latin American Studies at Princeton University supplied funds for fieldwork to J.A.W.

References

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

Beckman, J. & Weber, J. (1992). Survey of human and rat microsatellites. *Genomics* **12**, 627–631.

Beverly, S. & Wilson, A. (1984). Molecular evolution in *Drosophila* and the higher Diptera. II. A time scale for fly evolution. *Journal of Molecular Evolution* **21**, 1–13.

Dallas, J. (1992). Estimation of microsatellite mutation rates in recombinant inbred strains of mouse. *Mammalian Genome* **3**, 452–456.

Ellegren, H. (1992). Cloning of highly polymorphic microsatellites in the horse. *Animal Genetics* **23**, 133–142.

Ellegren, H. (1995). Mutation rates at porcine microsatellite loci. *Mammalian Genome* **6**, 376–377.

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

England, P., Briscoe, D. & Frankham, R. (1996). Microsatellite polymorphisms in a wild population of *Drosophila melanogaster*. *Genetical Research* **67**, 285–290.

Estoup, A., Presa, P., Kreig, F., Vaiman, D. & Guymard, R. (1993). $(CT)_n$ and $(GT)_n$ microsatellites: a new class of genetic markers for *Salmo trutta* L. (brown trout). *Heredity* **71**, 488–496.

Ezenwa, V., Peters, J., Zhu, Y., Arevalo, E., Hastings, M., Seppa, P., Pederson, J., Zacchi, F., Queller, D. & Strassman, J. (1998). Ancient conservation of trinucleotide microsatellite loci in polistine wasps. *Molecular Phylogenetics and Evolution* **10**, 168–177.

Fitzsimmons, N., Moritz, C. & Moore, S. (1995). Conservation and dynamics of microsatellite loci over 300 million years of marine turtle evolution. *Molecular Biology and Evolution* **12**, 432–440.

Forbes, S., Hogg, J., Buchanan, F., Crawford, A. & Allendorf, F. (1995). Microsatellite evolution in congeneric mammals. *Molecular Biology and Evolution* **12**, 1106–1113.

Gloor, G., Preston, C., Johnson-Schlitz, D., Nassif, N., Phillis, R., Benz, W., Robertson, H. & Engels, W. (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. & Pollock, D. D. (1997). Launching microsatellites: a review of mutation processes and methods of phylogenetic inference. *Journal of Heredity* **88**, 335–342.

Goldstein, D. & Schlötterer, C. (1999). *Microsatellites: Evolution and Applications*. Oxford University Press.

Harr, B., Zangerl, B., Brem, G. & Schlötterer, C. (1998). Conservation of locus-specific microsatellite variability across species: a comparison of two sibling species, *D. melanogaster* and *D. simulans*. *Molecular Biology and Evolution* **15**, 176–184.

Heed, W. B. & Krishnamurthy, N. B. (1959). Genetic studies on the *cardini* group of *Drosophila* in the West Indies. *University of Texas Publications* **5914**, 155–179.

Heed, W. & Russell, J. (1971). Phylogeny and population structure in island and continental species of the *cardini* group of *Drosophila* studied by inversion analysis. *University of Texas Publications* **7103**, 91–130.

Hollocher, H., Hatcher, J. & Dyreson, E. (2000 *a*). Evolution of abdominal pigmentation differences across species in the *Drosophila dunni* subgroup. *Evolution* **54**, 2046–2056.

Hollocher, H., Hatcher, J. & Dyreson, E. (2000 *b*). Genetic and developmental analysis of abdominal pigmentation differences across species in the *Drosophila dunni* subgroup. *Evolution* **54**, 2057–2071.

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

Jarne, P. & Lagoda, P. (1996). Microsatellites, from molecules to populations and back. *Trends in Ecology and Evolution* **11**, 424–129.

Jin, L., Macubas, C., Hallmayer, J., Kimura, A. & Mignot, E. (1996). Mutation rate varies among alleles at a microsatellite locus: phylogenetic evidence. *Proceedings of the National Academy of Sciences of the USA* **93**, 15285–15288.

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

Kruglyak, S., Durrett, R., Schug, M. & Aquadro, C. (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.

Michalakis, Y. & Veuille, M. (1996). Length variation of CAG/CAA trinucleotide repeats in natural populations of *Drosophila melanogaster* and its relation to the recombination rate. *Genetics* **143**, 1713–1725.

Noor, M., Schug, M. & Aquadro, C. (2000). Microsatellite variation in populations of *Drosophila pseudoobscura* and *Drosophila persimilis*. *Genetical Research* **75**, 25–35.

Noor, M., Kliman, R. & Machado, C. (2001). Evolutionary history of microsatellites in the Obscura Group of *Drosophila*. *Molecular Biology and Evolution* **18**, 551–556.

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.

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

Petrov, D. A. & Hartl, D. L. (1997). Trash DNA is what gets thrown away: high rate of DNA loss in *Drosophila*. *Gene* **205**, 279–289.

Petrov, D., Sangster, T., Spencer Johnston, J., Hartl, D. & Shaw, K. (2000). Evidence for DNA loss as a determinant of genome size. *Science* **287**, 1060–1062.

Rico, C., Rico, I. & Hewitt, G. (1996). 470 million years of conservation of microsatellite loci among fish species. *Proceedings of the Royal Society of London Series B* **263**, 540–557.

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

Schlötterer, C. & Pemberton, J. (1994). The use of microsatellites for genetic analysis of natural populations. *Molecular Ecology and Evolution: Approaches and Applications*, pp. 203–214. Birkhauser Verlag.

Schlötterer, 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.

Schlötterer, C., Ritter, R., Harr, B. & Brem, G. (1998). High mutation rate of a long microsatellite allele in *Drosophila melanogaster* provides evidence for allele-specific mutation rates. *Molecular Biology and Evolution* **15**, 1269–1274.

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

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

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

Spicer, G. (1988). Molecular evolution among some *Drosophila* species groups as indicated by two-dimensional electrophoresis. *Journal of Molecular Evolution* **27**, 250–260.

Stallings, R., Ford, A., Nelson, D., Torney, D., Hildebrand, C. & Moyzis, R. (1991). Evolution and distribution of $(GT)_n$ repetitive sequences in mammalian genomes. *Genomics* **10**, 807–815.

Tautz, D. & Renz, M. (1984). Simple sequences are ubiquitous repetitive components of eukaryotic genomes. *Nucleic Acids Research* **12**, 4127–4138.

Throckmorton, L. (1975). The phylogeny, ecology and geography of *Drosophila*. *Handbook of Genetics: Invertebrates of Genetic Interest*, Vol. 3, pp. 421–469. Plenum Press.

Weber, J. & Wong, C. (1993). Mutation of human short tandem repeats. *Human Molecular Genetics* **2**, 1123–1128.

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

Wilder, J. & Hollocher, H. (2001). Mobile elements and the genesis of microsatellites in Dipterans. *Molecular Biology and Evolution* **18**, 384–392.