Genet. Res., Camb. (2004), **84**, *pp.* 27–40. *With 3 figures.* © 2004 Cambridge University Press
DOI: 10.1017/S0016672304006986 Printed in the United Kingdom

Microsatellite variation among divergent populations of stalk-eyed flies, genus *Cyrtodiopsis*

TIMOTHY F. WRIGHT[1], PHILIP M. JOHNS, JAMES R. WALTERS[2], ADAM P. LERNER, JOHN G. SWALLOW[3] AND GERALD S. WILKINSON*
Department of Biology, University of Maryland, College Park, MD 20742, USA

(*Received 18 November 2003 and in revised form 27 April 2004*)

Summary

Microsatellite primers are often developed in one species and used to assess neutral variability in related species. Such analyses may be confounded by ascertainment bias (i.e. a decline in amplification success and allelic variability with increasing genetic distance from the source of the microsatellites). In addition, other factors, such as the size of the microsatellite, whether it consists of perfect or interrupted tandem repeats, and whether it is autosomal or X-linked, can affect variation. To test the relative importance of these factors on microsatellite variation, we examine patterns of amplification and allelic diversity in 52 microsatellite loci amplified from five individuals in each of six populations of *Cyrtodiopsis* stalk-eyed flies that range from 2·2% to 11·2% mitochondrial DNA sequence divergence from the population used for microsatellite development. We find that amplification success and most measures of allelic diversity declined with genetic distance from the source population, in some cases an order of magnitude faster than in birds or mammals. The median and range of the repeat array length did not decline with genetic distance. In addition, for loci on the X chromosome, we find evidence of lower observed heterozygosity compared with loci on autosomes. The differences in variability between X-linked and autosomal loci are not adequately explained by differences in effective population sizes of the chromosomes. We suggest, instead, that periodic selection events associated with X-chromosome meiotic drive, which is present in many of these populations, reduces X-linked variation.

1. Introduction

Microsatellites are short DNA sequences consisting of 1–5 base pair motifs repeated in tandem. Repeat number often exhibits high levels of polymorphism, with mutation rates estimated from 10^{-3} to 10^{-4} in vertebrates and 10^{-6} in *Drosophila* (Hancock, 1999). This variability makes microsatellite loci particularly useful as genetic markers for a range of applications (Goldstein & Schlötterer, 1999). However, mutations can alter allelic identity and priming sites, and thereby reduce microsatellite utility (Colson & Goldstein, 1999; Noor *et al.*, 2001) or bias the characteristics of these loci when amplified in other species, a phenomenon termed 'ascertainment bias' (Ellegren *et al.*, 1997*a*).

Ascertainment bias was first invoked to explain systematic differences in microsatellite allele lengths between species. These differences were initially interpreted as evidence of mutation rate differences between species, with shorter allele lengths indicating lower mutation rates (Rubinsztein *et al.*, 1995). Ellegren and colleagues (1997*b*) subsequently pointed out that shorter allele lengths in one of two species could be caused by the isolation of all loci in the other species. They hypothesized that, because microsatellite isolation protocols generally select loci with unusually long repeat lengths in the source species, any tendency for microsatellites to mutate towards shorter lengths

* Corresponding author. Tel: 001 301 405 6942. e-mail: wilkinso@umd.edu
[1] Current affiliation: Genetics Program, NMNH, Smithsonian National Zoo, 3001 Connecticut Ave., Washington DC 20008, USA
[2] Current affiliation: Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY 14853-2701, USA
[3] Current affiliation: Department of Biology, University of South Dakota, Vermillion, SD 57069, USA

would result in these loci exhibiting fewer repeats in related species. Likewise, because longer alleles tend to be more variable than shorter alleles (Weber, 1990), microsatellites in the source species should be more polymorphic and exhibit higher heterozygosity. This critique led to several studies that compared microsatellite allele lengths and variability reciprocally between two species using loci developed in both species (Ellegren *et al*., 1997*a*; van Treuren *et al*., 1997; Cooper *et al*., 1998; Crawford *et al*., 1998; Hutter *et al*., 1998). Such reciprocal studies, do not, however, allow the estimation of how ascertainment bias changes with increasing genetic distance from the source of the microsatellites. Such estimates can be obtained by comparing patterns of microsatellite amplification in populations or species at a range of genetic distances from a source population.

In addition to ascertainment bias, variation at microsatellite loci may also be influenced by their location in the genome. For example, selection on linked loci could affect microsatellite variability, with the effect depending on whether selection occurs against deleterious mutations, causing background selection, or for beneficial mutations, resulting in positive selection and hitchhiking. Background selection can eliminate linked neutral polymorphisms when deleterious recessive mutations reach appreciable frequency before removal. Deleterious recessive mutations should be removed more quickly from X chromosomes than from autosomes owing to hemizygous expression of X-linked mutations in males. Consequently, there should be a greater proportion of X chromosomes than autosomes that are free of deleterious recessive mutations, allowing higher neutral polymorphism on the X than on autosomes (Charlesworth *et al*., 1993; Aquadro *et al*., 1994; Betancourt *et al*., 2002). Conversely, under positive selection, beneficial recessive mutations should become fixed at a higher rate on the X chromosome than on any autosome, again because of hemizygous expression of X chromosomes in males. Fixation of beneficial mutations is expected to lead to a loss of neutral variability in linked loci, owing to genetic hitchhiking, thus higher rates of fixation on the X caused by positive selection would lead to lower neutral variability on the X than the autosomes (Maynard Smith & Haigh, 1974; Charlesworth *et al*., 1987). Thus, the two forms of selection make contrasting predictions for levels of neutral variability for X-linked versus autosomal loci.

In addition, meiotic drive systems (in which driving alleles of a given gene are preferentially transmitted over non-driving alleles) might mimic the hitchhiking effects of positive selection by increasing the rate of fixation in linked genes. Driving genes might be most common on the X chromosome, where they bias the sex ratio of offspring towards an excess of females by preventing the production of functional Y-bearing sperm in the father (Jaenike, 2001). Thus, species with X-linked male meiotic drive might also exhibit lower neutral polymorphism on the X chromosome than on autosomes (Derome *et al*., 2004).

Microsatellite variation should also differ between autosomal and X-linked loci owing to differences in the effective population sizes of the two types of chromosomes. With a 1 : 1 sex ratio, an X-linked gene has an effective population size (N_e) three-quarters that of an autosomal gene, which reduces the expected heterozygosity of X-linked loci relative to those of autosomes. Skew in the operational sex ratio can occur when males experience unequal mating success owing to sexual selection, or when X chromosome meiotic drive leads to female-biased population sex ratios. Female-biased sex ratios reduce N_e of males relative to females and thus reduce the apparent difference in N_e between X-linked and autosomal loci (Hedrick & Parker, 1997). Thus, the equilibrating effects of male-based meiotic drive on the N_e of X and autosomes would act counter to the decrease in neutral variability expected owing to hitchhiking with X-linked driving genes.

Stalk-eyed flies of the genus *Cyrtodiopsis* provide a useful system in which to examine how these factors influence microsatellite variation. Members of the genus occur throughout southeast Asia. Many species exhibit pronounced sexual dimorphism in the length of their protuberant eye stalks, and some of these species also possess a meiotic drive system that causes sex-ratio distortion (Presgraves *et al*., 1997; Wilkinson *et al*., 2003). The roles played by sexual selection and meiotic drive in the evolution of these eye stalks has been the focus of a range of behavioral, anatomical and genetic studies (Wilkinson, 2001). A recent study examined the degree of genetic isolation among populations of the species *Cyrtodiopsis dalmanni*, and its congeners *Cyrtodiopsis whitei* and *Cyrtodiopsis quinqueguttata* using 889 bp from the mitochondrial genes cytochrome oxidase II (COII) and the 16S ribosomal RNA (16S) (J. G. Swallow and G. S. Wilkinson, unpublished). This analysis found a well-supported tree in which all three species form monophyletic clades and individuals within sampled populations form monophyletic units (Fig. 1). When mated in the laboratory, all of these populations exhibit reproductive isolation, resulting from either the production of sterile hybrids or a failure to produce offspring (S. J. Christianson and G. S. Wilkinson, unpublished).

In this study we describe 52 microsatellite loci developed from a laboratory strain derived from populations of *C. dalmanni* that were collected near the village of Ulu Gombak, peninsular Malaysia (Wilkinson *et al*., 2003). We compare amplification of these loci in *C. dalmanni* sampled at the Gombak site, in three *C. dalmanni* populations sampled at other sites

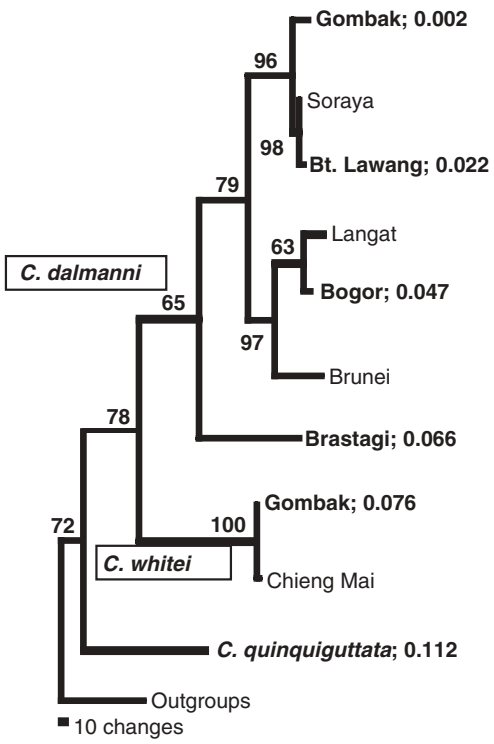


Fig. 1. A phylogeny of populations of stalk-eyed flies in the Genus *Cyrtodiopsis* constructed using maximum parsimony from mitochondrial genes COII and 16S (J. G. Swallow and G. S. Wilkinson, unpublished). Populations examined in the present study are in bold. Numbers to the right of population names are the genetic distances (as a proportion of nucleotide variation) from the Gombak population that was the source for the selected lines from which microsatellite loci were developed. The Gombak distance represents variation among the five individuals sampled at that site.

and in single populations of the congeners *C. whitei* and *C. quinqueguttata*. These five populations range from 2·2% to 11·2% mtDNA sequence divergence from the *C. dalmanni* Gombak population. We examine the effects of mtDNA sequence divergence, repeat array size and repeat array type on microsatellite amplification success and allelic diversity. In addition, we use observed heterozygosity as a means of comparing the level of selectively neutral variation on X-linked loci to that of autosomal loci.

2. Materials and methods

(i) *Microsatellite isolation and primer design*

As part of a project mapping quantitative trait loci for eyespan and sex ratio distortion, we designed microsatellite primers using DNA isolated from laboratory lines of *C. dalmanni* selected for altered eyespan-to-body-length ratio for over 50 generations (Wilkinson, 1993; Wolfenbarger & Wilkinson, 2001). We extracted DNA from one male and two females from a line selected for increased eyespan-to-body ratio (high lines), and from two males and one female from a line selected for decreased eyespan-to-body ratio (low lines). Flies were ground with a micropestle and DNA was extracted following the insect protocol in the DNeasy Tissue kit (Qiagen, Valencia, CA). DNA from these six flies was pooled and digested to completion with *Nhe*I, *Xmn*I, *Alu*I and *Bam*HI (New England Biolabs (NEB), Beverley, MA). The digest was enriched for GT repeats following standard protocols (Hamilton *et al.*, 1999). Briefly, the digest was ligated to SNX linkers and then amplified using the polymerase chain reaction (PCR) using complementary SNX primers. The amplified digest was then hybridized to a 3′ biotin-labeled 30 bp GT-repeat oligonucleotide and the hybridized fragments captured with streptavidin-coated magnetic beads. After another PCR amplification using SNX primers, GT-enriched products were ligated into *Xba*I-digested pBluescript SK+ plasmid vectors (Stratagene, La Jolla, CA) and transformed into *Escherichia coli* Supercompetent cells (Stratagene) for cloning.

We performed PCR in 30 μl volumes containing 50–100 ng DNA from each colony, 0.5 U Vent polymerase (NEB), 1 × Thermopol buffer (NEB), 0·2 mM of each dNTP, and 8 μM of the T3 and T7 primers. PCR reactions followed a profile of 96 °C for 5 minutes followed by 30 cycles of 96 °C for 45 s, 51 °C for 60 s and 72 °C for 120 s. PCR products from colonies containing genomic DNA inserts of 70–1000 bp were purified using QIAquick minicolumns (Qiagen) and sequenced in one direction using the ABI BigDye ready reaction kit (Applied Biosystems, Foster City, CA). Colonies containing microsatellite repeats were sequenced in the reverse direction using the same procedure and the resulting clone sequences compiled in Sequencher 4.1 (Gene Codes, Ann Arbor, MI).

We designed flanking primers for all clone sequences containing microsatellites with greater than three repeat units using the program Primer 3 (Rozen & Skaletsky, 2000). We tested primers for amplification with the genomic DNA used to create the enrichment using either a gradient or touchdown cycle on a PTC-200 Programable Thermal Cycler (MJ Research, Waltham, MA). The gradient cycle had duplicate reactions at annealing temperatures of 45 °C, 51 °C and 58 °C, whereas the touchdown cycle had an initial annealing temperature of 65 °C and decreased 0·7 °C per cycle to a final annealing temperature at 40·5 °C. Primer pairs that amplified products visible on an agarose gel were chosen for further study.

(ii) *Assessing microsatellite variability across populations*

To test amplification across populations of *C. dalmanni*, and its congeners *C. whitei* and *C. quinqueguttata*, we extracted DNA from field-collected

flies that were either preserved in 95% ethanol before extraction or used as fresh tissue. The four collection locations for *C. dalmanni* were: (1) near Ulu Gombak in peninsular Malyasia (3° 12′ N, 101° 42′ E); (2) near Bukit Lawang in northern Sumatra (3° 35′ N, 98° 6′ E); (3) at a forestry research station in Bogor, Java (6° 34′ S, 106° 50′ E); and (4) near Brastagi in northern Sumatra (3° 11′ N, 98° 28′ E). We captured *C. whitei* and *C. quinqueguttata* at site 1 near Ulu Gombak, Malaysia. We extracted DNA from five female flies from these six populations representing three species using either a DNeasy Tissue Kit (Qiagen) or phenol–chloroform (Baker *et al.*, 2001). PCR reactions were performed in 5·5 μl reactions containing 1 μl template DNA, 0.125 U *Taq* polymerase (Invitrogen, Carlsbad, CA), 1 × PCR buffer (Invitrogen), 0·2 mM of each dNTP, 2·5 mM $MgCl_2$ and 0·5 μM of each primer. We used either fluorescently labeled primers or 0·1 μM fluorescently labeled dCTP to label products. Labeled primers were used for loci that showed fixed fragment size differences between high and low selected lines in an initial screen of four individuals from each population; fluorescent dCTP were used for all other loci. We used a touchdown cycle of 17 cycles with initial annealing at 63 °C and decreases of 1 °C per cycle followed by 20 cycles with annealing at 47 °C. We removed excess fluorescent dinucleotides from PCR reactions by adding 0·25 μl shrimp alkaline phosphatase (1 units μl^{-1}; USB, Cleveland, OH) and 0·5 μl dilution buffer to each completed 5·5 μl PCR reaction followed by incubation at 37 °C for one hour and 65 °C for 15 minutes. Fluorescently labeled PCR products were separated on a 3100 DNA Analyzer (Applied Biosystems) using Pop-4 polymer and evaluated with Genescan 3.1.2 software (Applied Biosystems). Alleles were sized and scored using Genotyper 2.5 software (Applied Biosystems).

For each locus, we calculated the proportion of individuals with amplified products, the number of alleles amplified, the expected heterozygosity (H_E) and the observed heterozygosity (H_O) in each population. We calculated the proportion of loci that were polymorphic in each population, excluding those loci that did not amplify in the population. For each locus, we calculated the median and range of repeat array sizes within each population by taking the PCR product size of each allele and subtracting the size of the region flanking the repeat array as determined from the sequence of the isolated clone. For genetic distances between populations, we used the average proportion of nucleotide sequence difference in 889 bp of mtDNA COII and 16S genes for five individuals from each population (J. G. Swallow and G. S. Wilkinson, unpublished). The five non-source populations ranged from 2·2% to 11·2% sequence divergence from the *C. dalmanni* Gombak population (Fig. 1).

(iii) *Testing for effects of repeat array size and repeat array type*

We classified loci according to the size and nature of the repeat array. We divided loci into two size classes based on the median allele size when amplified in the Gombak source population: large loci had median repeat arrays of 30 bp or greater and small loci had median repeat arrays shorter than 30 bp. We used a cut off of 30 bp for two reasons: (a) repeats of that length or longer are expected to be preferentially isolated in our enrichment protocol; and (b) arrays of 15 repeat units or larger (e.g. $\geqslant$30 bp of a $[GT]_{15}$ repeat) have been found to harbor significantly more variability than shorter arrays in humans (Brinkmann *et al.*, 1998) and *Drosophila melanogaster* (Harr & Schlötterer, 2000). We classified repeat arrays as perfect if they consisted of uninterrupted tandem repeats in the isolated clone sequence and imperfect if the tandem repeats were interrupted by other motifs.

(iv) *Testing for the effect of chromosomal location on observed heterozygosity*

To determine whether loci were located on the X chromosome or autosomes, we extracted DNA and amplified loci from six male and five female hybrid *C. dalmanni* produced by crossing flies from the Gombak population with flies captured near the Soraya field station in northern Sumatra (2° 52′ N, 97° 54′ E). Loci that exhibited fixed allelic differences between the populations, heterozygous female F1 progeny, and homozygous F1 male progeny were scored as X-linked. Examination of haplotype data from an F2 intercross of two selected lines of *C. dalmanni* indicates that recombination is absent from males of this species (P. M. Johns and G. S. Wilkinson, unpublished). We applied two different adjustments to correct observed heterozygosity measures for the effective population sizes of X-linked and autosomally linked markers. The first was a uniform correction applied to all species in which we multiplied observed heterozygosity for X-linked markers by 1·33, based on an assumption of a 1 : 1 effective sex ratio (Begun & Whitley, 2000). The second was a species-specific correction based on brood sex ratios produced in captive populations (Wilkinson *et al.*, 2003), which correspond closely to sex ratios observed at mating aggregations in the field (Wilkinson & Reillo, 1994). These data indicate that the frequency and intensity of X chromosome meiotic drive do not differ between populations of the same species, but do differ between species (Wilkinson *et al.*, 2003), and results in average brood sex ratios of 0·373 for *C. whitei*, 0·453 for *C. dalmanni* and 0·5 for *C. quinqueguttata*. Using the formula in Kauer *et al.* (2002), these sex ratios produce correction factors (k) of 1·22 for *C. whitei*, 1·29 for *C. dalmanni* and 1·33 for *C. quinqueguttata*.

(v) *Statistical analysis*

Our experimental design, in which each locus was amplified in every population, creates a situation in which measures of allelic variation at each locus are not statistically independent. To avoid artificially inflating our degrees of freedom, we used the mean value for each measure of variation within each of the six populations for statistical tests. We used linear regression to examine the effect of genetic distance on measures of amplification success and allelic variability. Logarithmic regressions of these measures on genetic distance gave similar significance values and fit to the data, and are not reported here. We used analysis of covariance (ANCOVA) to examine the effect of array size and array type on our allelic diversity measures while controlling for the effect of genetic distance, and to test the effect of chromosomal location on observed heterozygosity while controlling for genetic difference. ANCOVA's with non-significant interaction terms were rerun with the interaction term removed. When testing for significance of a particular effect across several measures, we applied the sequential Bonferroni corrections to P values within each set of tests to correct for the possibility of type-I errors (Rice, 1989). All statistical tests were conducted in Statview 5.1 (SAS Institute, Cary, NC).

3. Results

(i) *Microsatellite isolation*

We successfully sequenced 433 transformed colonies, of which 200 contained repeat arrays with three or more repeat units. Of these, 114 contained sufficient flanking regions to design primers. We designed primers for 69 loci and obtained consistent amplification with genomic DNA from the high and low selected lines for 52 primer pairs (Table 1, clone sequences available as GenBank accession numbers AY542904–AY542955). These 52 microsatellite loci were used in all subsequent population tests.

(ii) *Effects of genetic distance from the source population*

Amplification success and most measures of genetic diversity declined with increasing mtDNA sequence divergence from the source population (Fig. 2). Linear regressions on the mean values for the six populations showed that the percent of successful amplification declined significantly with genetic distance ($r^2=0{\cdot}97$, $P=0{\cdot}0004$, Bonferroni-corrected $P<0{\cdot}05$), as did the percent of polymorphic loci ($r^2=0{\cdot}90$, $P=0{\cdot}004$, Bonferroni-corrected $P<0{\cdot}05$) and expected heterozygosity within populations ($r^2=0{\cdot}83$, $P=0{\cdot}01$, Bonferroni-corrected $P<0{\cdot}05$). Both the number of alleles amplified per population ($r^2=0{\cdot}80$, $P=0{\cdot}02$, Bonferroni-corrected P not significant) and observed heterozygosity ($r^2=0{\cdot}75$, $P=0{\cdot}02$, Bonferroni-corrected P not significant) showed no relationship with genetic distance after sequential Bonferroni correction for multiple tests. An ANCOVA with genetic distance as a covariate found no difference between observed and expected heterozygosity after removal of the non-significant interaction term, but did find a strong effect of genetic distance (heterozygosity type, $F=4{\cdot}3$, $P=0{\cdot}07$; genetic distance, $F=34{\cdot}0$, $P=0{\cdot}0002$). The size of the repeat array did not change with genetic distance from the source population ($r^2=0{\cdot}25$, $P=0{\cdot}3$, Bonferroni-corrected P not significant), nor did the size range of the repeat array ($r^2=0{\cdot}42$, $P=0{\cdot}2$, Bonferroni-corrected P not significant).

(iii) *Effects of repeat array size and type*

The median array size for our 52 loci in the source population averaged 25·5 bp. We classified 17 loci in the large size class ($\geqslant 30$ bp median array size in Gombak) and 35 loci in the small size class (<30 bp). ANCOVA on our allelic variation measures with size class and genetic distance revealed that loci in the larger size class had significantly more allelic variation by most measures (Table 2). Polymorphism, the number of alleles amplified and expected heterozygosity were all higher for loci with larger arrays; observed heterozygosity did not differ between the two size classes.

We classified 13 loci as perfect repeats and 39 as imperfect based on the array structure in the initial sequence isolated from the cloned enrichment fragments. Perfect arrays were more likely to fall into the smaller (<30 bp) size class (Fisher's exact test, $P=0{\cdot}04$). As found for smaller arrays, arrays containing perfect repeats were less variable than those with imperfect repeats for some measures of allelic variation when compared by ANCOVA (Table 3). These measures included amplification success and number of alleles amplified. Observed heterozygosity did not differ between the two classes of repeat arrays. Both the median repeat array size and the range in repeat array sizes differed strongly between the two classes, with imperfect arrays having larger values for both measures (Table 3).

(iv) *Effects of chromosomal location on observed heterozygosity*

We determined chromosome location for 45 of the 52 microsatellite loci by comparing male and female hybrid genotypes obtained by crossing individuals from two populations. Of these 45 loci, 37 were autosomal and eight were X-linked (Table 1). ANCOVA with chromosomal type and genetic distance (Table 4)

Table 1. *Characteristics of 52 microsatellites developed from* C. dalmanni

Locus	Repeat sequence*	Product length* (bp)	Chromosome type	Median repeat array size in Gombak population†	Observed heterozygosity in Gombak population§	Cross-amplification in non-source populations (%)¶	Primer sequence 5′–3′**
ms-012	1. $[GT]_6$,AT $[GT]_1$ 2. $[GT]_7CT[GT]_1T[GT]_2$	255	Autosome	33	1·00	64	F: TGTTATTTTCATCTCGATCTGCAT R: CGGACAAACAGACAAGCGTA
ms-014	$[AC]_{18}CC[AC]_2[CATA]_3$	208	Autosome	28	1·00	56	F: FAM-CTTCGGCGTTAGGTAGTA R: ACGGGTAGTGTCTGTATGTCT
ms-015	$[AT]_4$	202	Undetermined	8	0·00	88	F: TCCTCCTGCATTTGCTATGTT R: TTCGAGCGAACAGTGGTATG
ms-021	$[GT]_1AG[GT]_9AT[GT]_2CAG[GT]_2$	386	Autosome	16	1·00	60	F: ACCATTTAGTTAGTGAGTTGACA R: TCCTTTGATTTTGCCATTT
ms-039	$[AC]_2AA[AC]_4GCW$ $[CA]_3A[AC]_7AT[AC]_1TC[AC]_2$	147	Autosome	34	0·80	64	F: NED-AATCACAACGCTAACGAGTCA R: ATGCTTCAACGCTTACCTACC
ms-054	$[AC]_2ATTAT[AC]_{10}AT[AC]_1$	165	X	40	0·53	84	F: FAM-ACGGAAGTAACACAAAAAGATAAA R: TCAGCGCTACTCACAGAACTAACT
ms-059	$[GT]_3AT[GT]_1CTA[GT]_{11}$	129	Autosome	36	1·00	88	F: TGGAGCATATCATTACGTACAAAA R: GTGCGCTTAATAGTTGGCAACAC
ms-070	$[TG]_4CT[TG]_2TT[TG]_2TT[TG]_2TA$ $[TG]_2CCTC[TG]_{11}TC[TG]_1$	168	Autosome	62	0·80	76	F: CTCGCTTGCTTTATGAACGA R: TGCCATTTGAACGTAATTGAC
ms-071	$[AC]_{13}$	179	X	24	0·00	20	F: AGTAACTCTTCCGTCACCTT R: AACCGATTTCTTATACACATTT
ms-090	$[GT]_{11}GA[GT]_3GG[GT]_4AT[GT]_1$	197	Autosome	24	1·00	28	F: NED-TCTTGCCTTTGCCACACTAA R: TGGGAAATGTGAGTTTACTTAAACAGT
ms-095B	$[TG]_{11}TA[TG]_1TA[TG]_2A[TG]_2$	175	Autosome	34	0·80	76	F: HEX-TTGGTCTGCTTGCAGATTGT R: CCATTTTGCTTAGAGCAGGAC
ms-106	$[TG]_3CG[TG]_4CG[TG]_3TTA[TG]_3$	235	X	33	0·89	64	F: GCAACTGACCAACCTCACTTC R: GGCAGCAGTGCATAAAAATG
ms-116	$[GT]_2ATA[TG]_8AG[TG]_4$	197	Autosome	28	0·80	68	F: FAM-TGAGTTTATATGCATGCTAATGGTT R: TCTGCATTTCGTTCCACAAG
ms-122	$[GT]_2ATACATATCA[GT]_8AT$ $[GT]_5$	213	Autosome	28	0·80	76	F: TGTGCGTCGTATGCTGTTG R: TTTAGCAAAACTTCCGATTCAA
ms-125	$[GT]_{14}$	153	X	32	0·44	60	F: FAM-TGGTGTTAATGAACGAGTGACTTC R: TGCCATTCATGCAAGTCTTC
ms-157	$[AC]_8$	248	Autosome	15	0·50	68	F: NED-CTGGCTCAGTATTGCTGCAT R: TCGTCGTTGTTGTTTGCATT
ms-167	$[AC]_1TC[AC]_9$	222	X	30	0·53	28	F: GCTGCGAGCTGTAAAACAGA R: GGCAGTGACAATGGCAGTAA
ms-174	$[GT]_1AT[GT]_{13}$	211	Autosome	21	1·00	100	F: FAM-TGCATAGTGGAATTAGGTCACG R: TGTCCTTCGGCGTTAGGTAG
ms-177	$[GT]_4$	170	Autosome	5	0·80	84	F: AAAATAATATGCGGTCGGTCA R: AATTGAAGCCACATGCTCCT
ms-217A	$[GT]_3AT[GT]_9$	169	Autosome	30	0·33	80	F: FAM-TGTTACTTCAGCGAAAAGAGCA

ms-221	$[GT]_{15}$	159	Autosome	No amplification	No amplification	36	R: AGAAGCCTGCCAACTTCACT F: TCATTTGCATGAGTTTTGCAG R: TCTGTGTTGTGTTCACTGTTGG
ms-223	$[CA]_2[AC]_1AAG[AC]_{10}$	188	Autosome	26	0·75	88	F: TTGCAACATGGGATACATTTTT R: CGTCTCCGGTATGGCTTAAC
ms-238	$[AT]_7$	157	X	4	0·67	32	F: CGTTCACAAAGGCAAACAGA R: GACGTTTGGCGTTTGGTATT
ms-244	$[GTT]_4[GT]_1C[GT]_{10}$	223	Autosome	30	0·60	56	F: NED-GAGCATGAGAACGTGGATGA R: AAGGGACAAAGGCTATGCAA
ms-249	$[AT]_3[GT]_{13}C[GT]_1C[GT]_2$	214	Autosome	36	0·80	100	F: HEX-TGATGCTGCAGTGGTAAAGG R: GAGCGAGCACGACAATTAAA
ms-262Y	$[GT]_{13}$	259	Autosome	20	0·80	72	F: CGAGTTGGACACTTCGCTCT R: AAGCGCACAAAATGGCTACT
ms-262Z	$[CT]_3[GTT]_5GAT[GTT]_1$	158	Autosome	26	1·00	80	F: FAM-GTAGCCATTTTGTGCGCTTT R: CAACAGCGACAGCAGTAGGA
ms-277	$[AC]_{10}GA[AC]_2$	194	Autosome	12	1·00	56	F: HEX-AGTTCATCGTTGCTCCGTTC R: TGCAAGTCAGAGGATGATCG
ms-282	$[AT]_4$	196	Autosome	8	0·00	96	F: ACGATGAGCAACGAAGGTTT R: CGTCGCTCCACTAAAAACG
ms-291	$[GT]_7AT[GT]_3GA[GT]_2$	198	Autosome	16	0·40	76	F: GATTGGGGTAGCACGTCTGT R: CTCCATAAACATGAACATACCAAA
ms -295	$[AC]_1AT[AC]_{11}$	167	Autosome	18	0·40	32	F: GAAGCCTTTAAGCGATAATGC R: TTTTGATGTCGCTTTGTTATTATTG
ms-301A	$[AC]_8AT[AC]_3TC[AC]_2$	138	Autosome	15	0·75	96	F: FAM-TTCAGCACTAAATGCAGCAGA R: GCACTTAACATGCGATGAGG
ms-325	1. $[AT]_4$ 2. $[AT]_3$	260	Undetermined	−39	0·00	20	F: CCTCGTCCATTGCATTTTTAC R: CGTGTCTGTTCGAAAATGGA
ms-336	$[AC]_1C[AC]_4AT[AC]_2AA[AC]_5$ GC$[AC]_1$	211	Autosome	37	0·40	72	F: GCACAAACTCGTCCCAAACT R: TGCCACTTCGTCATAAGCAA
ms-357	$[AT]_7AC[AT]_4CA[AT]_1$	155	Autosome	29	0·00	20	F: AATCACTACCGACGGCATTT R: CCCAATGTTCCGATTGTTGT
ms-361A	$[TG]_2T[TG]_4CG[TG]_4TA[TG]_2$	149	Autosome	28	0·00	80	F: TGGCTGGCTGTCTTTACACTC R: CTCATTTGTTATCGGGCGTA
ms-387	$[GT]_3AT[GT]_1ATGC[GT]_8$	106	Undetermined	22	0·60	72	F: TTTTCACATCTGCAAAAATCTGTT R: TGCAGTATGCACAATTCACA
ms-389	$[AT]_1C[AT]_2[GAT]_2GT[GAT]_1$	189	Undetermined	8	0·40	64	F: CGAGAAGGATGATAAAGAATTGC R: AGTTTGTGTTGGCAATGCAG
ms-392	$[GT]_1AT[GT]_2AT[GT]_2AT[GT]_4$	227	Autosome	25	0·75	68	F: HEX-TTGGCATGGTGTGATTCCTA R: GACACTGAACGTACAGAGCAAC
ms-393	$[CA]_{13}$	186	Autosome	28	0·40	48	F: CAAACTTGCAGAAGGCATGA R: GCCGGCTTTAAAGTCTGTCT
ms-395	$[GT]_{10}AT[GT]_1$	200	X	23	0·53	68	F: HEX-CGAGTAGAGCACTTTGAAGATACA R: TTGCGGTTGTAGAAGTTTGC
ms-397	$[TG]_{13}TRTAG[TG]_1C[TG]_2$ TTC$[TG]_1T[TG]_2$	198	Autosome	42	1·00	24	F: GACGAACGTGTCGATTTGTG R: AGAAACACCGACAACAGCAA
ms-398	$[CA]_{11}TA[CA]_2[TA]_3[CA]_2$	146	Autosome	40	0·60	68	F: NED-GCGAAATACAAACAAGCGAGA R: CGAATGCGAATACGATTGTG

Table 1. (*Cont.*)

Locus	Repeat sequence*	Product length* (bp)	Chromosome type	Median repeat array size in Gombak population†	Observed heterozygosity in Gombak population§	Cross-amplification in non-source populations (%)¶	Primer sequence 5′–3′**
ms-402A	1. $[CAA]_2ATA[CAA]_8CAG[CAA]_2$ 2. $[AC]_8$	205	Autosome	56	0·60	84	F: NED-CCAAATGGGCCACATTATTC R: AGGAAAGTGGATGCATTCGT
ms-402B	$[GT]_4AA[GT]_{10}G[GT]_2$	153	Autosome	38	0·60	88	F: HEX-GGCCGAATTCAATTTAAGCA R: CCACAAAATATTTGCCAACAA
ms-415	$[AT]_1A[AT]_4$	186	Autosome	1	0·25	76	F: HEX-ATGCAGATCTGGGGTCAAAT R: ATGCGAAGGGTACAAGGAAA
ms-421	$[AC]_{13}$	120	Undetermined	−12	1·00	76	F: CGGAGGTACCACAAAAAGAAAAA R: CGAGATGGACACTTCGCTCT
ms-422	$[CA]_7CG[CA]_3$	182	Autosome	22	0·00	84	F: TTGGCGAATTTCTTGGCTAC R: TGGCTTACAGTTGTGGCTGA
ms-442	$[CCA]_3CCG[CCA]_4$	207	Undetermined	20	0·25	28	F: CAGTACCTGTTGAAGCCGAAG R: GTGGTCCACCTTCTGCTGAT
ms-470	$[AT]_3GCA[TC]_3TG[TC]_2$	202	Undetermined	15	0·75	64	F: TTTGCCAAGTCATCACCAAC R: GCTGACCACGCTAATTACTGC
ms-478	1. $[CT]_3$ 2. $[GT]_3$	208	X	11	0·00	88	F: CTCGATTGTCACGCCAGTAG R: GCAAACAAAAACAAAGTGACG
ms-480	$[GT]_{10}[AT]_2[GT]_1[AT]_1[GT]_1$	114	Autosome	26	0·60	60	F: NED-CAGCCATGAAATTCAGTACCAA R: TGTTGGCTGTACGTACTTCGTC

* Repeat array and product size determined from colony boil sequences.
† Negative values imply deletions in the flanking regions.
§ Observed heterozygosity values for X-linked loci are multiplied by 4/3 to correct for difference in expected population size relative to autosomal loci.
¶ Amplification success averaged across the five non-source populations.
** F, forward primer; R, reverse primer.

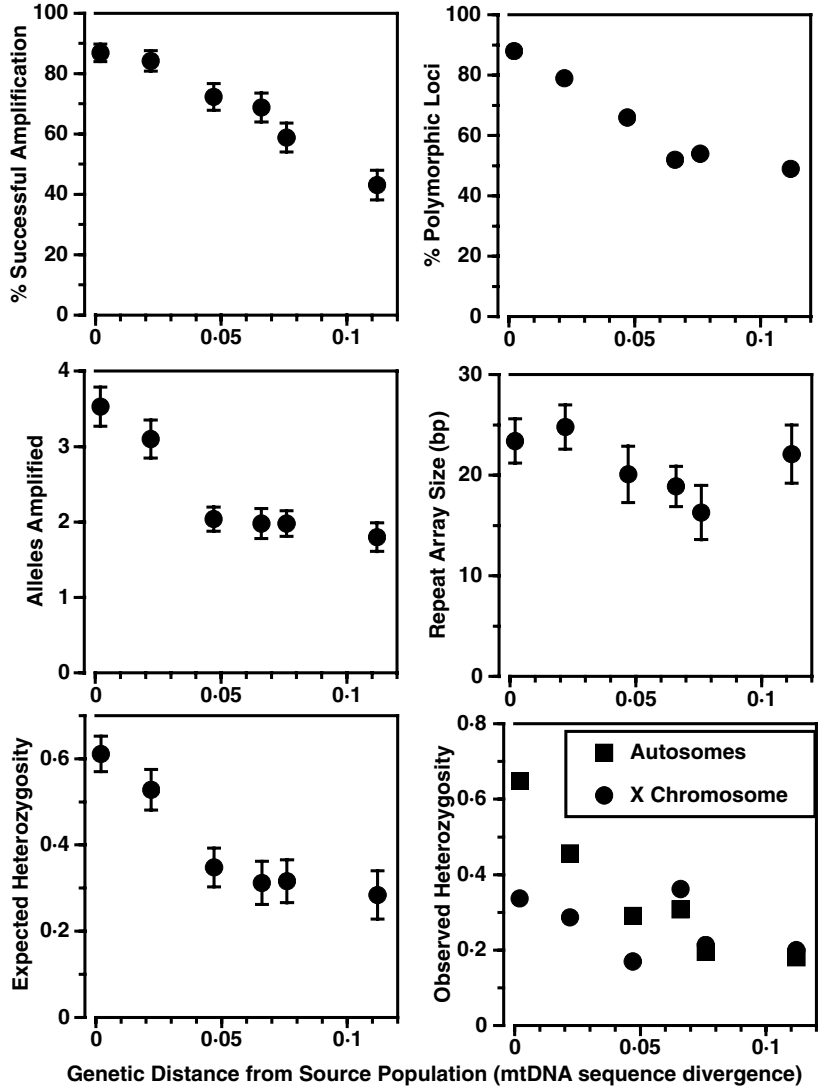


Fig. 2. Plots of mean ± SE for each population for measures of microsatellite amplification success and allelic variation. Populations are arrayed by genetic distance from the source population as measured by mitochondrial DNA sequence divergence. Measures are the proportion of successful amplification, the porportion of polymorphic loci, the number of alleles amplified, the median size of the microsatellite repeat array, expected heterozygosity and the observed heterozygosity. The means for observed heterozygosity are shown for autosomal and X-linked loci separately to illustrate the significant interaction between genetic distance and chromosomal location for this measure.

revealed that autosomally linked loci had higher observed heterozygosity than did X-linked loci. Observed heterozygosity also showed a significant interaction between genetic distance and chromosome type, with autosomal loci showing a decline with genetic distance and X-linked loci exhibiting low levels of observed heterozygosity in all populations (Fig. 2). The difference in observed heterozygosity between autosomal and X-linked loci, and the interaction of chromosome type with genetic distance were no longer significant after application of either the uniform correction for effective population size or the species-specific correction (Table 4). The effect of these corrections, however, was not uniform across the populations. Observed heterozygosity values for X-linked loci remained lower than autosomal values after these corrections in the source and closely related populations, whereas X-linked values for populations genetically distant from the source population were higher than autosomal values after correction (Fig. 3).

4. Discussion

(i) *Ascertainment bias and genetic distance from the source population*

We found evidence of ascertainment bias in several measures that directly affect the utility of microsatellite loci when amplified from divergent populations and species. In general, the severity of the bias

Table 2. *Results from analyses of covariance of amplification and allelic diversity measures with genetic distance and size class of the repeat array*

Measure	Means ± SE		*F* values*	
	<30 bp	⩾30 bp	Size class	Genetic distance
Amplification (%)	67·3 ± 7·3	72·5 ± 5·6	3·5	100·8†
Polymorphism (%)	57·7 ± 6·6	77·8 ± 7·4	25·2†	52·5†
Alleles amplified (no.)	2·1 ± 0·28	2·9 ± 0·37	10·9†	30·4†
Expected heterozygosity	0·34 ± 0·05	0·51 ± 0·06	21·0†	40·2†
Observed heterozygosity	0·30 ± 0·06	0·36 ± 0·06	1·3	18·8†
Array size range	7·2 ± 0·9	10·9 ± 1·2	10·3†	7·5†

* All interaction terms were non-significant and were removed from the models. Residual d.f. = 9 for all models.
† $P < 0{\cdot}05$ after sequential Bonferroni correction.

Table 3. *Results from analyses of covariance of amplification and allelic diversity measures with genetic distance and type of repeat array*

Measure	Means ± SE		*F* values*	
	Imperfect	Perfect	Repeat type	Genetic distance
Amplification (%)	70·6 ± 6·9	64·4 ± 6·2	10·6†	225·3†
Polymorphism (%)	69·3 ± 7·5	50·6 ± 7·3	6·3	10·6†
Alleles amplified (no.)	2·6 ± 0·35	1·8 ± 0·18	11·0†	18·6†
Expected heterozygosity	0·43 ± 0·06	0·31 ± 0·04	7·9	23·0†
Observed heterozygosity	0·34 ± 0·07	0·27 ± 0·04	2·6	18·6†
Repeat array size	23·7 ± 1·8	12·1 ± 1·2	24·6†	0·09
Array size range	9·8 ± 1·2	4·3 ± 0·5	24·8†	4·3

* All interaction terms were non-significant and were removed from the models. Residual d.f. = 9 for all models.
† $P < 0{\cdot}05$ after sequential Bonferroni correction.

was proportional to the genetic distance from the population in which the loci were originally isolated. Amplification success, proportion of polymorphic loci and expected heterozygosity all declined with genetic distance from the source population. In many cases, these measures declined substantially with relatively small increases in genetic distance. For example, the proportion of successful amplifications declined by half with 11% mtDNA sequence divergence, dropping from 87% amplification in the source Gombak population of *C. dalmanni* to 43% amplification in *C. quinqueguttata*. Expected heterozygosity, meanwhile, dropped from 0·61 in Gombak to 0·35 in Bogor, two populations separated by 4·5% mtDNA sequence divergence. Overall, both within *C. dalmanni* and across related species, measures of allelic variability declined more rapidly than amplification success.

Qualitatively similar patterns have been found for microsatellite cross-amplification in vertebrates, although the decline in both amplification success and allelic diversity with genetic distance is more gradual than in stalk-eyed flies. Two separate studies of cross-amplification of primers in birds found that both the proportion of microsatellite loci that amplified and the proportion of polymorphic loci declined logarithmically with increasing genetic distance from the source species, with polymorphisms declining faster than amplification success (Primmer *et al.*, 1996; Galbusera *et al.*, 2000). Primmer *et al.* (1996) estimated that amplification success and polymorphism would decline to 50% at distances of 7 $\Delta T_M H$ and 5 $\Delta T_M H$ (as measured by DNA–DNA hybridization) respectively. Using calibrations of 1·0 ΔT_M equals 2·3 million years (MY) for passerines and 4·6 MY for non-passerines (Sibley & Ahlquist, 1990), these distances correspond to divergence times between 16 MY and 32 MY for 50% amplification success and between 11 MY and 23 MY for a twofold reduction in polymorphism. By contrast, in the stalk-eyed flies, we found a twofold reduction in both amplification success and polymorphism between our source population and its congener *C. quinqueguttata*, which exhibit 11% sequence divergence in COII. If

Table 4. *Results from analyses of covariance of the effect of chromosomal location and genetic distance on observed heterozygosity for the microsatellite loci*

Measure	Means ± SE		*F* values*		
	Autosome	X	Chromosome type	Genetic distance	Interaction
Observed heterozygosity	0·35 ± 0·07	0·26 ± 0·03	10·1†	16·9†	6·5†
Observed heterozygosity, uniform correction	0·35 ± 0·07	0·35 ± 0·04	0·002	10·5†	ns
Observed heterozygosity, species-specific correction	0·35 ± 0·07	0·34 ± 0·04	0·02	10·5†	ns

* Residual d.f. = 8 for models with a significant interaction term; residual d.f. = 9 for models with non-significant interaction terms removed.
† $P < 0·05$.

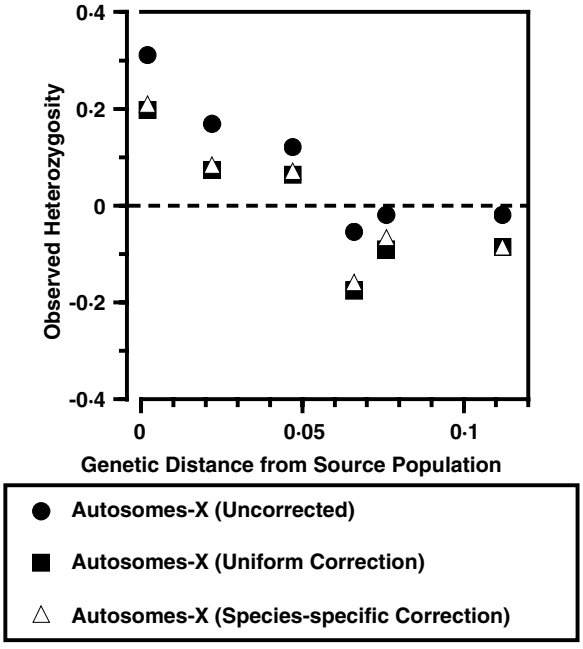


Fig. 3. Plots of the difference in mean values for observed heterozygosity between autosomal loci and X-linked loci by genetic distance from the population source when observed values of heterozygosity for X-linked loci are uncorrected (circles), corrected for effective population size with a uniform 4/3 correction for X-linked loci that assumes a balanced sex ratio (squares) and with a species-specific correction for observed effective population sizes of male and female stalk-eyed flies (triangles). Both types of corrections reduce the difference in observed heterozygosity between autosomal and X-linked loci in the source population and closely related populations but increase this difference in populations more distantly related to the source population.

we assume that stalk-eyed fly substitution rates equal 2·8% COII sequence divergence per million years as in *Drosophila* (Spicer & Pitnick, 1996), the divergence time between *C. dalmanni* and *C. quinqueguttata* would be only 2·3 MY. Although these are only rough estimates, they do suggest that both amplification success and polymorphism decline an order of magnitude more quickly in stalk-eyed flies than in birds.

(ii) *Causes of ascertainment bias*

Ascertainment bias in microsatellite amplification is generally ascribed to high mutation rates in flanking regions leading to mismatch of primers to the template during PCR. We found evidence for large-scale deletions in flanking regions at some loci resulting in the calculation of negative repeat array sizes for those alleles (Table 1). Other studies have observed similar mutations in microsatellite flanking regions (Ortí *et al*., 1997; Hutter *et al*., 1998; Colson & Goldstein, 1999; Noor *et al*., 2001).

The causes of ascertainment bias in measures of allelic diversity are more complex. One hypothesis is that many protocols favor the isolation of microsatellites with long repeat arrays, which tend to be more variable than shorter arrays (Ellegren *et al*., 1997*b*). Longer repeat arrays also tend to be rarer than short arrays, perhaps because of biases in mutation mechanisms that favor smaller repeat array sizes (Kruglyak *et al*., 1998; Eisen, 1999; Estoup & Cornuet, 1999; Falush & Iwasa, 1999; Harr & Schlötterer, 2000). Thus, there should be a tendency for long microsatellites isolated in the source population to decline in both length and variability in more distant populations and species (Harr & Schlötterer, 2000; Pascual *et al*., 2000). Our enrichment protocol included a hybridization step that preferentially selects for GT arrays of 15 units or more. Despite this step, we isolated many loci with fewer than 15 repeat units in the source population. Furthermore, although our larger loci ($\geqslant$30 bp) were more variable in several measures of allelic diversity (Table 4), we did not see a significant decline in median allele length with increasing genetic distance from the source population (Fig. 1). Although some previous studies have

detected ascertainment bias in repeat length (Primmer *et al*., 1996; Ellegren *et al*., 1997*a*, *b*), others have not (Hutter *et al*., 1998; Zhu *et al*., 2000), and many of the studies that did detect some ascertainment bias found differences in allelic diversity and allele lengths between species regardless of the source of the microsatellites (Cooper *et al*., 1998; Crawford *et al*., 1998; Amos *et al*., 2003). A reciprocal contrast between loci developed in the *C. dalmanni* Gombak population and loci developed in a different *Cyrtodiopsis* population or species would provide a stronger test for ascertainment bias in repeat length.

A second hypothesis to explain ascertainment bias in allelic diversity is that the microsatellite isolation process emphasizes the detection of perfect (i.e. uninterrupted) repeat arrays in the source population. In a reciprocal comparison of *D. melanogaster* and *Drosophila simulans* (Hutter *et al*., 1998), perfect repeats were both positively correlated with heterozygosity, and more common in the source population. In our set of loci, perfect repeats were relatively uncommon, comprising only 25% of all loci, and all but one of these was shorter than 30 bp. This bias was probably responsible for our finding that loci with imperfect arrays tended to be more, rather than less, variable than loci with perfect repeat arrays (Table 3). Furthermore, it has been found that interruptions in microsatellite repeat arrays can be removed during replication, resulting in the regeneration of a perfect array from an imperfect one (Harr *et al*., 2000). Because we did not sequence all microsatellite alleles, we cannot be certain that repeat arrays identified as imperfect in the original isolates were imperfect in all populations.

A third hypothesis for the reduction in allelic variability with genetic distance from the source population is that null alleles (i.e. alleles that do not amplify owing to primer-template mismatch) are more common in non-source populations. Rates of both substitutions and insertions/deletions appear to be relatively high in microsatellite flanking regions and could increase the incidence of null alleles (Glenn *et al*., 1996; Ortí *et al*., 1997; Hutter *et al*., 1998; Colson & Goldstein, 1999; Noor *et al*., 2001). An increase in null alleles with increasing genetic distance could account for some of our observed declines in number of alleles amplified, polymorphism and heterozygosity, particularly if mutation rates in the tandem repeat region were correlated with mutation rates in the flanking regions as proposed for crocodilians (Glenn *et al*., 1996). If null alleles were more common in populations distant from the source, we would expect to see departure from Hardy–Weinberg equilibrium owing to an excess of homozygotes in these populations. We did not, however, detect any difference between observed and expected heterozygosity, and the interaction with genetic distance was also not significant, suggesting that null alleles are not more common in genetically distant populations.

(iii) *Effects of chromosomal location on observed heterozygosity*

We found that observed heterozygosity was generally low for X-linked loci in all populations. By contrast, observed heterozygosity for autosomal loci declined with increasing genetic distance in a manner similar to other measures of allelic diversity (Fig. 2). Corrections for the effective population sizes of the two types of chromosomes reduced this difference somewhat, but values for X-linked loci remained lower than autosomal loci in the Gombak source population and closely related populations of *C. dalmanni*. Although these results come from a relatively small sample of X-linked loci, they do suggest that reduced heterozygosity at X-linked loci may be obscured by ascertainment bias in more distantly related populations and is the result of factors other than reduced effective population size.

Lower variability in X-linked loci might result from genetic hitchhiking with genes under selection. There are two scenarios under which hitchhiking might have this effect. The first is that selection for beneficial mutations will be more efficient on the X chromosome than on autosomes and will lead to lower variability of linked neutral variants on the X (Maynard Smith & Haigh, 1974; Charlesworth *et al*., 1987). This hypothesis was proposed to explain the finding that recently derived non-African populations of *D. melanogaster* had lower observed heterozygosity for X-linked microsatellite loci than for autosomal loci (Kauer *et al*., 2002). Studies of recently derived populations of *D. melanogaster* and *D. simulans* using other genetic markers have found similar patterns (Begun & Whitley, 2000; Andolfatto, 2001; but see Betancourt *et al*., 2002), suggesting that positive selection and genetic hitchhiking might play an important role when populations are expanding into novel environments.

An alternative hitchhiking scenario is that selfish elements on an X chromosome will reduce variability at linked neutral loci owing to rapid evolution and fixation of meiotic drive chromosomes, as has recently been reported for *D. simulans* (Derome *et al*., 2004). Several lines of evidence suggest this scenario might be particularly likely in stalk-eyed flies. First, all populations of *C. dalmanni* and *C. whitei* sampled to date possess X-linked factors that cause males carrying them to produce female-biased progeny sex ratios (Wilkinson *et al*., 2003). Second, a recent genetic mapping study based on an F2 intercross in which the sex ratio factor segregates indicates that X-linked microsatellite loci segregate with the drive factor, suggesting that one or more chromosomal inversions

prevent recombination between the driving X and the wild-type X in heterozygotes (P. M. Johns and G. S. Wilkinson, unpublished). Third, among Gombak *C. dalmanni*, extreme female-biased sex ratios are always produced by males carrying either of two X haplotypes, whereas several other X haplotypes are either associated with normal sex ratios or partially female-biased sex ratios, as expected if these are older drive haplotypes that are now susceptible to segregating suppressor loci (G. S. Wilkinson and P. M. Johns, unpublished). A further test of this hypothesis would be to examine X-linked versus autosomal variability in microsatellite loci developed in a stalk-eyed fly species without apparent sex ratio distortion, such as *C. quinqueguttata*.

(iv) *Implications of ascertainment bias*

It has been widely recognized that direct comparisons of allelic diversity and heterozygosity among different species using microsatellites developed in one of these species are subject to ascertainment bias (Ellegren *et al*., 1997a; Cooper *et al*., 1998; Hutter *et al*., 1998). The data presented here, which come from six reproductively isolated populations of stalk-eyed flies drawn from three closely-related species, suggest that this concern should, in some cases, extend to comparisons across populations within a species. The magnitude of this bias is likely to be greater for divergent populations that exhibit reproductive isolation, such as these stalk-eyed fly populations, than for populations with ongoing gene flow. Further studies are required to determine whether the high degree of ascertainment bias observed in *Cyrtodiopsis* stalk-eyed flies are restricted to this group or are representative of a wider range of taxa.

We thank S. Lance, E. Askt and R. Fleischer for generous advice with microsatellite development, K. Madden and S. Prager for fly maintenance, and M. Dambach and S. Christianson for assistance in lab. Research was supported by NSF grants DEB-9807937 and DEB-0077878 (G.S.W.), NIH grant NRSA F32-GM20377-0 (J.G.S.), and an American Philosophical Society grant (J.G.S.). P.M.J. received fellowship support from the Neuroethology: Neurobiology, Evolution and Behavior training grant at the University of Maryland (NIH T32-MH020048). T.F.W. received support during the preparation of this manuscript from the Smithsonian Office of Fellowships and Grants, and the Friends of the National Zoo as a postdoctoral fellow in the Genetics Program.

References

Amos, W., Hutter, C. M., Schug, M. D. & Aquadro, C. F. (2003). Directional evolution of size coupled with ascertainment bias for variation in *Drosophila* microsatellites. *Molecular Biology and Evolution* **20**, 660–662.

Andolfatto, P. (2001). Contrasting patterns of X-linked and autosomal nucleotide variation in *Drosophila melanogaster* and *Drosophila simulans*. *Molecular Biology and Evolution* **18**, 279–290.

Aquadro, C. F., Begun, D. J. & Kindahl, E. C. (1994). Selection, recombination, and DNA polymorphism in *Drosophila*. In *Non-neutral Evolution: Theories and Molecular Data* (ed. B. Golding), pp. 46–56. Chapman and Hall.

Baker, R. H., Wilkinson, G. S. & DeSalle, R. (2001). The phylogenetic utility of different types of molecular data used to infer evolutionary relationships among stalk-eyed flies (Diopsidae). *Systematic Biology* **50**, 87–105.

Begun, D. J. & Whitley, P. (2000). Reduced X-linked nucleotide polymorphism in *Drosophila simulans*. *Proceedings of the National Academy of Sciences of the USA* **97**, 5960–5965.

Betancourt, A. J., Presgraves, D. C. & Swanson, W. J. (2002). A test for faster X evolution in *Drosophila*. *Molecular Biology and Evolution* **19**, 1816–1819.

Brinkmann, B., Klintschar, M., Neuhube, R. F., Huhne, J. & Rolf, B. (1998). Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. *American Journal of Human Genetics* **62**, 1408–1415.

Charlesworth, B., Coyne, J. A. & Barton, N. H. (1987). The relative rates of evolution of sex chromosomes and autosomes. *American Naturalist* **130**, 113–146.

Charlesworth, B., Morgan, M. T. & Charlesworth, D. (1993). The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**, 1289–1303.

Colson, I. & Goldstein, D. B. (1999). Evidence for complex mutations at microsatellite loci in *Drosophila*. *Genetics* **152**, 617–627.

Cooper, G., Rubinsztein, D. C. & Amos, W. (1998). Ascertainment bias cannot entirely account for human microstallites being longer than their chimpanzee homologues. *Human Molecular Genetics* **7**, 1425–1429.

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

Derome, N., Métayer, K., Montchamp-Moreau, C. & Veuille, M. (2004). Signature of selective sweep associated with the evolution of sex-ratio drive in *Drosophila simulans*. *Genetics* **166**, 1357–1366.

Eisen, J. A. (1999). Mechanistic basis for microsatellite instability. *Microsatellites: Evolution and Applications*, pp. 34–48. Oxford University Press.

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

Ellegren, H., Primmer, C. R. & Sheldon, B. C. (1997*b*). Microsatellite evolution: directionality or bias in locus selection. *Nature Genetics* **11**, 360–362.

Estoup, A. & Cornuet, J.-M. (1999). Microsatellite evolution: inferences from population data. *Microsatellites: Evolution and Applications*, pp. 49–65. Oxford University Press.

Falush, D. & Iwasa, Y. (1999). Size-dependent mutability and microsatellite constraints. *Molecular Biology and Evolution* **16**, 960–966.

Galbusera, P., Van Dongen, S. & Matthysen, E. (2000). Cross-species amplification of microsatellite primers in passerine birds. *Conservation Genetics* **1**, 163–168.

Glenn, T. C., Stephan, W., Dessauer, H. C. & Braun, M. J. (1996). Allelic diversity in alligator microsatellite loci is

negatively correlated with GC content of flanking sequences and evolutionary conservation of PCR amplifiability. *Molecular Biology and Evolution* **13**, 1151–1154.

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

Hamilton, M., Pincus, E., Di Fiore, A. & Fleischer, R. C. (1999). Universal linker and ligation procedures for construction of genomic DNA libraries enriched for microsatellites. *Biotechniques* **27**, 500–507.

Hancock, J. M. (1999). Microsatellites and other simple sequences: genomic context and mutation mechanisms. In *Microsatellites: Evolution and Applications* (ed. D. B. Goldstein & C. Schlötterer), pp. 1–9. Oxford University Press.

Harr, B. & Schlötterer, C. (2000). Long microsatellite alleles in *Drosophila melanogaster* have a downward mutation bias and short persistence times, which cause their genome-wide underrepresentation. *Genetics* **155**, 1213–1220.

Harr, B., Zangerl, B. & Schlötterer, C. (2000). Removal of microsatellite interruptions by DNA replication slippage: phylogenetic evidence from *Drosophila*. *Molecular Biology and Evolution* **17**, 1001–1009.

Hedrick, P. W. & Parker, J. D. (1997). Evolutionary genetics and genetic variation of haplodiploids and X-linked genes. *Annual Review of Ecology and Systematics* **28**, 55–83.

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.

Jaenike, J. (2001). Sex chromosome meiotic drive. *Annual Review of Ecology and Systematics* **32**, 25–49.

Kauer, M., Zangerl, B., Dieringer, D. & Schlotterer, C. (2002). Chromosomal patterns of microsatellite variability contrast sharply in African and non-African populations of *Drosophila melanogaster*. *Genetics* **160**, 247–256.

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.

Maynard Smith, J. & Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genetical Research* **23**, 23–35.

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

Ortí, G., Pearse, D. E. & Avise, J. C. (1997). Phylogenetic assessment of length variation at a microsatellite locus. *Proceedings of the National Academy of Sciences of the USA* **94**, 10745–10749.

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

Presgraves, D. C., Severance, E. & Wilkinson, G. S. (1997). Sex chromosome meitotic drive in stalk-eyed flies. *Genetics* **147**, 1169–1180.

Primmer, C. R., Møller, A. P. & Ellegren, H. (1996). A wide-range survey of cross species microsatellite amplification in birds. *Molecular Ecology* **5**, 365–378.

Rice, W. R. (1989). Analyzing tables of statistical tests. *Evolution* **43**, 223–225.

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. Humana Press.

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

Sibley, C. G. & Ahlquist, J. E. (1990). *Phylogeny and Classification of Birds: A Study in Molecular Evolution*. Yale University Press.

Spicer, G. S. & Pitnick, S. (1996). Molecular systematics of the *Drosphila hydei* subgroup as inferred from mitochondrial DNA sequences. *Journal of Molecular Evolution* **43**, 281–286.

van Treuren, R., Kuittinen, H., Kärkkäinen, K., Baena-Gonzalez, E. & Savolainen, O. (1997). Evolution of microsatellites in *Arabis petra* and *Arabis lyrata*, outcrossing relatives of *Arabadopsis thaliana*. *Molecular Biology and Evolution* **14**, 220–229.

Weber, J. L. (1990). Informativeness of human $(dC–dA)_n$. $(dG–dT)_n$ polymorphisms. *Genomics* **7**, 524–530.

Wilkinson, G. S. (1993). Artificial sexual selection alters allometry in the stalk-eyed fly *Cyrtodiopsis dalmanni* (Diptera: Diopsidae). *Genetical Research* **62**, 213–222.

Wilkinson, G. S. (2001). Genetic consequences of sexual selection in stalk-eyed flies. *Model Systems in Behavioral Ecology*, pp. 72–91. Princeton University Press.

Wilkinson, G. S. & Reillo, P. R. (1994). Female choice response to artificial selection on an exaggerated male trait in a stalk-eyed fly. *Genetical Research* **255**, 1–6.

Wilkinson, G. S., Swallow, J. G., Christiansen, S. J. & Madden, K. (2003). Phylogeography of *sex ratio* and multiple mating in stalk-eyed flies from southeast Asia. *Genetica* **117**, 37–46.

Wolfenbarger, L. L. & Wilkinson, G. S. (2001). Sex-linked expression of a sexually-selected trait in the stalk-eyed fly, *Cyrtodiopsis dalmanni*. *Evolution* **55**, 103–110.

Zhu, Y., Queller, D. C. & Strassman, J. E. (2000). A phylogenetic perspective on sequence evolution in microsatellite loci. *Journal of Molecular Evolution* **50**, 324–338.