

# Population structure and host-plant specialization in two *Scaptodrosophila* flower-breeding species

JSF Barker

School of Rural Science and Agriculture, University of New England, Armidale NSW 2351, Australia

In contrast to phytophagous insect species, little attention has been paid to the possibility of host races in the Drosophilidae, although flower-breeding species, where courtship and mating take place on the flowers, are likely candidates. Two species of *Scaptodrosophila*, *S. hibisci* and *S. acclinata*, are restricted to flowers of *Hibiscus* species (section *Furcaria*), and the *Furcaria* specialization likely predated the separation of *S. hibisci* and *S. acclinata*. In all, 20 microsatellite loci were analysed in nine populations of *S. hibisci* and five of *S. acclinata*. For two pairs of *S. hibisci* populations in close proximity, but breeding on different *Hibiscus* species, differentiation between the populations of each of these pairs was similar to that between the populations that were from the same *Hibiscus* species, but geographically distant, suggesting the early stages of host-

race formation. Genetic variability was significantly less in *S. acclinata* than in *S. hibisci*, suggesting greater drift effects in the former. However, of 253 alleles detected, 82 were present in both species, 160 in *S. hibisci* only and 11 in *S. acclinata* only, indicating that *S. acclinata* was derived from *S. hibisci*, following a strong bottleneck at the time of separation – possibly 40 000 years BP. Analyses and interpretation of Hardy–Weinberg equilibrium and F statistics needed to account for null alleles known to be present at eight loci in *S. hibisci*, and possibly present at other loci. The results emphasize the need for caution in studies where the presence of null alleles is inferred only from population data. *Heredity* (2005) 94, 129–138. doi:10.1038/sj.hdy.6800592  
Published online 8 September 2004

**Keywords:** *Scaptodrosophila*; flower breeders; host races; null alleles; population structure; microsatellites

## Introduction

Insect species with host races occurring on different plant species have provided model systems for study of the evolution of specialization and speciation. Some cases of host-race formation, where new hosts have been introduced into the range of the insect within historical times, are well documented (eg the tephritid fly *Rhagoletis pomonella* – Feder *et al*, 1998). Much of the interest in these systems has been in the context of sympatric speciation, where it is assumed that specialization on different, but sympatric host-plant species may result indirectly in significant reproductive isolation (Via, 2001; Berlocher and Feder, 2002). In many other cases, insect species have been found to comprise populations that are locally adapted (specialized) to different plant species (Funk *et al*, 2002; see reviews and references in Mopper and Strauss, 1998). Where there is this high level of host-plant specialization, these species allow detailed analyses of its genetic basis – genetic differentiation, host-plant preferences, mate choice and fitness on the different hosts (eg Janz *et al*, 2001; Thomas *et al*, 2003).

In contrast to the phytophagous insect species where host races commonly have been found, most species of the Drosophilidae are saprophagous, feeding on yeasts and bacteria and the decaying tissue of plants and fungi.

Among such species, most are generalists with unknown breeding sites, and only *D. mojavensis* is known to have evolved host races (Etges, 2002). On the other hand, species that breed in the flowers of more than one plant species may provide more opportunity for the evolution of host-plant specialization. Two closely related species of *Scaptodrosophila*, *S. hibisci* (Bock in Cook *et al*, 1977) and *S. acclinata* McEvey and Barker, 2001 have been recorded breeding in flowers of *Hibiscus* species in Australia, the former in five species in eastern Australia (Starmer *et al*, 1997; Wolf *et al*, 2000a) and the latter in 11 species in the Northern Territory (Wolf *et al*, 2000a; JSF Barker, unpublished). As both *Scaptodrosophila* species show an apparently strict host-plant association with endemic *Hibiscus* species of the section *Furcaria* (Barker, 2003), but each uses a range of different *Hibiscus* species (only *H. meraukensis* being common to both), it seems likely that *Furcaria* specialization predated the speciation of *S. hibisci* and *S. acclinata*. As courtship and mating take place on the flowers, adults could potentially respond to different physical or chemical cues from the flowers of different *Hibiscus* species, while larval growth and survival may be subject to differential selection by host-plant chemistry. Thus, further specialization may have evolved within each species, leading to genetic differentiation and host-race formation on different *Hibiscus* species.

As compared with the genus *Drosophila*, very little is known of the biology and genetics of *Scaptodrosophila* species, but our already completed studies of *S. hibisci* include ecological aspects, quantitative genetic analyses and reproductive biology (Starmer *et al*, 1997, 1998, 2000; Polak *et al*, 1998, 2001; Wolf *et al*, 2000a, b). Both *S. hibisci*

Correspondence: Dr JSF Barker, School of Rural Science and Agriculture, University of New England, Armidale NSW 2351, Australia.

E-mail: sbarker@pobox.une.edu.au

Received 28 March 2003; accepted 21 July 2004; published online 8 September 2004

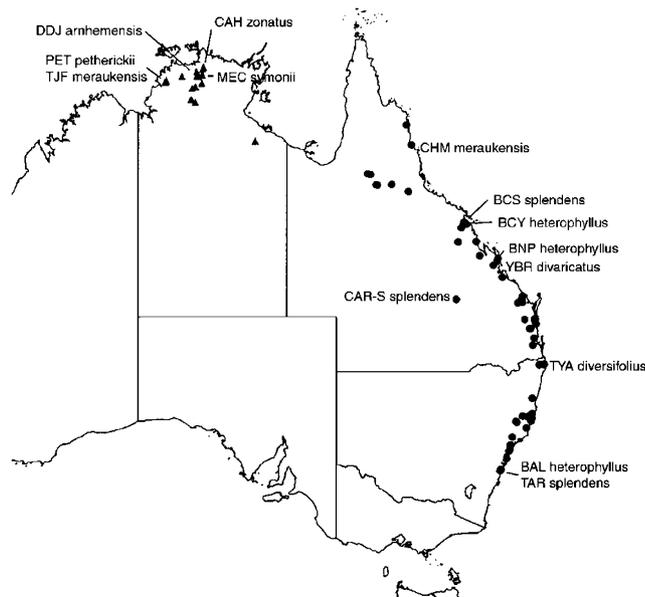
and *S. acclinata* have fragmented, patchy distributions, with distances between neighbouring populations ranging from 1 to 2 km to probably 100 km or more. The distribution of *S. hibisci* is broader than that of *S. acclinata* (McEvey and Barker, 2001), but for both species, local population sizes may vary by orders of magnitude, as the breeding habitat at a locality varies from just one or two to hundreds of plants.

I present here a comparative analysis of population structure in the two species and tests for host-plant specialization in *S. hibisci*, based on microsatellite assay of nine populations of *S. hibisci* and five of *S. acclinata*. Prior to these population assays, segregation studies had shown null alleles to be present at a number of the microsatellite loci (Wilson *et al.*, unpublished), and their treatment in the analyses has been given special consideration.

## Materials and methods

Flies were collected directly from *Hibiscus* flowers in the field, and either put immediately into absolute ethanol, or kept alive in vials with sugar-agar medium until return to the laboratory, where they were frozen at  $-80^{\circ}\text{C}$ . The collection localities and *Hibiscus* species at each are shown in Figure 1.

DNA extraction and microsatellite typing for 21 males and 21 females from each population were performed using the protocols of Wilson *et al.* (2002) that were developed for *S. hibisci*. Of the 20 loci assayed, five (Sh9, Sh72, Sh74, Sh78 and Sh89) are X linked in *S. hibisci* (Wilson *et al.*, 2002), and 16 (Sh8i, Sh8ii, Sh9, Sh29c, Sh34d, Sh36c, Sh38d, Sh49, Sh74, Sh81, Sh88, Sh88<sup>+</sup>, Sh89, Sh90, Sh90<sup>+</sup> and Sh94) have null alleles segregating in one or more populations of *S. hibisci* (14 identified in Wilson *et al.*, unpublished) and two more in this paper). Only Sh9, Sh36c and Sh90 had null alleles in *S. acclinata* (reported here).



**Figure 1** Localities where *S. hibisci* (●) and *S. acclinata* (▲) have been collected. Populations assayed here indicated by code name and *Hibiscus* species at that locality.

## Allele frequency and heterozygosity

For those autosomal loci where null homozygotes were segregating, genotype and allele frequencies were estimated using the computer program GENEPOP Version 3.3, option 8,1 (Raymond and Rousset, 1995), separately for each population where null homozygotes were detected. For the X-linked loci, where there were null individuals in males only in any population (Sh9 in three populations of *S. hibisci*) or none segregating, allele frequencies were computed directly. Where null females were present (Sh74 and Sh89 in *S. hibisci*), GENEPOP was used to estimate genotype and allele frequencies for females, and overall (sex combined) allele frequencies were then computed. Thus, for seven loci in *S. hibisci* (Sh8ii, Sh34d, Sh36c, Sh74, Sh88<sup>+</sup>, Sh89 and Sh90) and two in *S. acclinata* (Sh36c and Sh90), Hardy–Weinberg equilibrium is assumed for the estimation of null heterozygote frequencies.

Tests for deviations from Hardy–Weinberg equilibrium were carried out using the exact tests of GENEPOP, with significance levels for each test determined by applying the sequential Bonferroni procedure (Hochberg, 1988; Lessios, 1992) over all loci within each population to the probability estimates calculated by GENEPOP. As the raw genotype data were used for these tests, estimated deviations from Hardy–Weinberg equilibrium will be inflated for those loci where null homozygotes were segregating. The number of alleles per locus (all loci) and observed heterozygosity (for the 12 loci where no null homozygotes were detected in any population) were compared among populations of each species with Kruskal–Wallis analysis of variance (Sokal and Rohlf, 1981).

## Statistical analyses

Tests for pairwise linkage (genotypic) disequilibria, separately for each population and overall, were carried out using GENEPOP, and then applying the sequential Bonferroni procedure over the pairwise tests to determine significance levels. Genotypic differentiation among populations was tested using GENEPOP, both for overall and pairwise differentiation, with the sequential Bonferroni procedure applied over population pairs for the latter in determining significance levels.

Two of the microsatellites (Sh8i and Sh8ii) are juxtaposed, being separated by 150 base pairs (Wilson *et al.*, 2002). The allele size range for microsatellites is generally presumed to be constrained (Nauta and Weissing, 1996), and juxtaposed microsatellites might be expected to provide insights into the mechanisms of any such constraints. To determine any association of allele sizes at these two loci, haplotypes were estimated using the method of Stephens *et al.* (2001), as implemented in the program PHASE. The correlation between allele sizes of the two loci in the haplotypes was computed, and a permutation test was carried out to compare the observed covariance of the repeat numbers with that expected by chance (Pritchard and Feldman, 1996).

For the loci where no null homozygotes were detected in any population, F-statistics and their significance were determined using FSTAT Version 2.9.3 (Goudet, 2001), again with the sequential Bonferroni procedure applied over loci in deriving significance levels. F-statistics for

loci with null alleles segregating were computed using the methods of Weir and Cockerham (1984), as implemented in the computer program DIPLOID (Weir, 1990), but with this program adapted (B Tier, personal communication) for input of the numbers of each genotype at each locus (program *diplinta.f* available at <http://ansc.une.edu.au/ansc/genetics/> – link through Barker JSF). The genotype numbers used were those estimated as above, accounting for null homozygotes.

Mean gene diversity (ie expected heterozygosity – computed following correction for null heterozygote frequencies) for each locus and for each population and the  $D_A$  distances among populations (Nei *et al*, 1983) were obtained using the DISPAN program (T Ota, personal communication). As Takezaki and Nei (1996) showed that this distance is generally the best for inferring the correct topology, it was used to construct a neighbour-joining tree including both species (Saitou and Nei, 1987), using the DISPAN program.

Isolation by distance was tested using data on the 12 loci in *S. hibisci* and 10 in *S. acclinata* where no null homozygotes were detected. Estimates of  $\theta$  (Weir and Cockerham, 1984) for each pair of populations were computed using the FSTAT program, and pairwise  $\theta/(1-\theta)$  values were then regressed on log (geographic distance) between each pair of populations (Rousset, 1997), and the significance of the association determined using Mantel's (1967) permutation test.

Given the patchy distribution of the host plants and the variation in the number of plants at each locality, the BOTTLENECK program (Cornuet and Luikart, 1996) was applied to determine if there had been past bottlenecks in population size at any locality. In this analysis, only those loci that did not deviate from Hardy–Weinberg equilibrium in each population were used, and for the TPM (two-phase model) the options recommended by Piry *et al* (1999) were applied.

### Divergence time of the two species

Genetic distances between the two species were estimated for Nei's standard distance ( $D_S$  – Nei, 1972, 1978) and the Tomiuk and Loeschke distance ( $D_{TL}$  – Tomiuk and Loeschke, 1991, 1995), using the POPDIST program available at <http://genetics.sh.dk/bg/popgen>. Takezaki and Nei (1996) showed that  $D_S$  was more appropriate than other measures for estimating evolutionary time, while Tomiuk *et al* (1998) found  $D_{TL}$  to be more stable and more closely correlated linearly with evolutionary time than  $D_S$ . Two approaches were used with the 17 loci assayed in both species: (i) including all 14 populations, and computing the average of the 45 interspecific pairwise distances, and (ii) pooling data from all populations of each species, and computing a single distance.

## Results

Allele frequencies for each population and the primary data files (genotypes of all individuals) are available at <http://ansc.une.edu.au/ansc/genetics/> link through Barker JSF. One locus (Sh90<sup>+</sup>) did not amplify satisfactorily in two *S. hibisci* populations, and this and two other loci (Sh29c and Sh88<sup>+</sup>) did not amplify in *S. acclinata*. For the remaining 17 loci in the two species, the mean numbers of alleles/locus/population were, for

*S. hibisci* and *S. acclinata* respectively, 5.38 and 1.24 for the X-linked loci, and 9.16 and 5.18 for the autosomal, both significantly less for *S. acclinata* ( $P < 0.01$ , Kruskal–Wallis ANOVA). For seven loci (Sh8, Sh13, Sh34d, Sh48, Sh78, Sh81 and Sh94), the same allele was at the highest frequency in all or most populations of both species, but for other loci allele frequency distributions were often markedly different. There were no significant differences among populations of either species for either number of alleles per locus or observed heterozygosity.

Over the 17 loci in both species, a total of 253 alleles were detected (Table 1), with 82 in both species and 160 only in *S. hibisci*. However, the alleles found in *S. acclinata* were not just a subset of those in *S. hibisci*, as 11 alleles (in five loci) were found only in *S. acclinata*. One allele (at locus Sh8ii) was fixed or near to fixation in all *S. acclinata* populations, while four were present at low frequency in all five *S. acclinata* populations.

Average gene diversity (expected heterozygosity) within populations also was significantly less for *S. acclinata* ( $P < 0.01$ , Kruskal–Wallis ANOVA), although five loci (Sh8, Sh13, Sh49, Sh88 and Sh94) had similar estimates in both species (Table 1). The reduction in average number of alleles and in heterozygosity for *S. acclinata* was much greater for the X-linked loci than for the autosomal. Using Mann–Whitney tests, the differences between X-linked and autosomal loci for number of alleles and expected heterozygosity were significant (both  $P < 0.05$ ), in contrast to no significant differences for *S. hibisci*. Average expected heterozygosities for all populations within each species were very similar – *S. hibisci* range: 0.625–0.699, and *S. acclinata* range: 0.323–0.363.

Significance of deviations from Hardy–Weinberg expectations was tested using the raw genotype frequencies (ie not corrected for unrecognized null heterozygotes), so results for those locus/population combinations with null homozygotes present are biased toward heterozygote deficiency (Tables 2 and 3). While the proportion of significant tests for these loci in *S. hibisci* ( $35/70 = 0.5$ ) is greater than for the other loci ( $28/108 = 0.259$ ), a number of locus/population combinations without nulls are significant, while others with nulls are not significant (Table 2). Across all loci, all significant tests show an observed heterozygote deficiency, while global tests over the loci where no null homozygotes were detected were highly significant ( $P < 0.001$ ) for all populations except TYA ( $P < 0.05$ ). In all populations, all or most loci showed observed heterozygote deficiency. For the X-linked loci, genotype frequencies in females were generally in Hardy–Weinberg equilibrium (26/27 for loci without nulls, 12/18 for loci with nulls), but allele frequencies differed significantly between the sexes in 18.5% (5/27) of cases for loci without nulls, and 11.1% (2/18) for loci with nulls.

In strong contrast to the results for *S. hibisci*, only 8/55 (14.5%) of tests for deviation from Hardy–Weinberg equilibrium were significant for *S. acclinata* (Table 3 – three in TJF and three for Sh90, where null homozygotes were segregating), with all significant tests again showing an observed heterozygote deficiency. Global tests were not significant for MEC, but were significant for the other populations (PET –  $P < 0.05$ , CAH –  $P < 0.01$ , DDJ and TJF –  $P < 0.001$ ). None of the significant tests

**Table 1** Number of alleles and average expected heterozygosity per locus for *S. hibisci* and *S. acclinata*

Locus	Number of alleles			Total	Expected heterozygosity <sup>a</sup>	
	<i>S. hibisci</i>	<i>S. acclinata</i>	<i>S. acclinata only</i>		<i>S. hibisci</i>	<i>S. acclinata</i>
<i>X-linked loci</i>						
Sh9	18	3		18	0.696	0.142
Sh72	9	1		9	0.661	0.000
Sh74	5	2		5	0.423	0.013
Sh78	7	2		7	0.446	0.007
Sh89	18	1		18	0.732	0.000
<i>Autosomal loci</i>						
Sh8i	9	8	3	12	0.745	0.744
Sh8ii	20	4	4	24	0.840	0.019
Sh13	12	10		12	0.818	0.822
Sh29c	12	—			0.673	—
Sh34d	7	2		7	0.546	0.344
Sh36c	23	6		23	0.859	0.538
Sh38d	17	5		17	0.781	0.606
Sh48	4	2		4	0.018	0.009
Sh49	35	14		35	0.883	0.821
Sh81	13	1		13	0.594	0.000
Sh88	26	18	1	27	0.883	0.815
Sh88*	13	—			0.606	—
Sh90	8	5	1	9	0.707	0.241
Sh94	11	9	2	13	0.603	0.660
Mean ( $\pm$ SD)	14.2 (8.4) <sup>b</sup>	5.5 (4.9)		14.9 (8.5)	0.659 (0.206)	0.340 (0.344)
X-linked loci – mean ( $\pm$ SD)	11.4 (6.2)	1.8 (0.8)		11.4 (6.2)	0.592 (0.146)	0.032 (0.062)
Autosomal loci – mean ( $\pm$ SD)	15.4 (9.1) <sup>b</sup>	7.0 (5.1)		16.3 (9.1)	0.683 (0.223)	0.468 (0.331)

<sup>a</sup>Expected heterozygosity for X-linked loci estimated from female data only.

<sup>b</sup>Excluding Sh29c, Sh88\*.

were for X-linked loci, but allele frequencies differed significantly between the sexes in 12.5% (1/8) of cases.

For both species, no genotypic linkage disequilibrium tests within populations were significant after Bonferroni correction.

Over all nine *S. hibisci* populations (378 individuals), 79 different haplotypes for Sh8i and Sh8ii were estimated. The correlation coefficient between allele sizes of the two loci was  $-0.020$  ( $P=0.59$ ), and for the permutation test of the covariance, the probability of getting the observed value by chance was 0.305.

As with the Hardy–Weinberg tests within populations, the overall F-statistics show generally positive  $f$  ( $F_{IS}$ ) values (Tables 4, 5). Mean  $f$  values are not significantly different between those loci where no null homozygotes were detected and those where null heterozygote frequencies were estimated assuming Hardy–Weinberg equilibrium (*S. hibisci*: 0.170 and 0.152, Table 4; *S. acclinata*: 0.107 and 0.045, Table 5). It would be expected that the latter loci should have  $f$  values not significantly different from zero, but there could still be deficiencies of heterozygotes for the non-null alleles. There were more loci with significant  $f$  estimates, and higher mean values for *S. hibisci* than for *S. acclinata* (Tables 4 and 5). For both species,  $\theta$  ( $F_{ST}$ ) estimates show significant differentiation among populations, but the magnitude of the differentiation is 2–3 times greater for *S. hibisci*. Neither species showed a significant pattern of isolation by distance ( $P=0.332$  and 0.070 for *S. hibisci* and *S. acclinata*, respectively). For both species, genotypic differentiation of populations was highly significant ( $P<0.001$ ). Pairwise  $\theta$  estimates (based on the loci in part 1 of Tables 4 and 5) show for *S. hibisci* (Table 6) that two populations (CHM

and CAR-S) are each significantly differentiated from each other and from all other populations. In contrast, three of the central Queensland populations (BCY, BNP, YBR) are not significantly different from each other, nor from the two southernmost populations (BAL, TAR). Similarly for *S. acclinata*, where geographic distances among the populations are much less, some that are geographically close (MEC, DDJ) are significantly different, while more distant populations are not (Table 6). Pairwise estimates of genotypic differentiation (Table 6) showed more differentiation among populations, but were generally in accord with the estimates of  $\theta$ .

The BOTTLENECK analysis (results not shown), using the TPM model that is most appropriate for microsatellites, found that, rather than heterozygosity excess (ie expected heterozygosity greater than that expected from the observed number of alleles – Cornuet and Luikart, 1996) as expected with past bottlenecks, all populations of *S. hibisci* (except YBR) have a significant heterozygosity deficiency, while no populations of *S. acclinata* are significant (except TJJ for the Wilcoxin test).

The neighbour-joining tree including both species (Figure 2) shows a clear separation of the species, and for *S. hibisci* no clustering of populations from the same *Hibiscus* species. CAR-S and CHM, the populations most genetically distinct from the others, also are the most geographically separated from the other populations.

For the two pairs of *S. hibisci* populations from different *Hibiscus* species but in close proximity, both pairwise genotypic differentiation and  $D_A$  genetic distances show differentiation between the populations of each pair similar to that between these populations that were from the same *Hibiscus* species, but far distant

**Table 2** Significant tests of deviation from Hardy–Weinberg equilibrium for *S. hibisci* (females only for X-linked loci): (1) 12 loci where no null homozygotes were detected (Sh9 null individuals in males only), and (2) eight loci where null homozygotes were segregating (including Sh90<sup>+</sup>, which was not assayed in two populations)

Locus	CHM	BCS	BCY	BNP	YBR	CAR-S	TYA	BAL	TAR
(1)									
Sh8i	0.682***	0.602***		0.172*	0.467***	0.369**		0.438***	
Sh9 <sup>x</sup>			N <sup>a</sup>			N, S <sup>b</sup>			N
Sh13									
Sh29c	0.441**	0.324**	0.335***	0.274*		0.231*		0.116***	0.463**
Sh38d						0.282**			0.316*
Sh48	H <sup>c</sup>		H	H	M <sup>d</sup>	M	H	M	H
Sh49	0.585***		0.179*			0.324***			
Sh72 <sup>x</sup>	S*		S*			S*			
Sh78 <sup>x</sup>						0.709***			
Sh81						S**		0.231*	
Sh88		0.360***		0.242*	0.275*			0.396***	0.309***
Sh94	0.427***		0.403**	0.365**					
Global (over loci) probability	***	***	***	***	***	***	*	***	***
(2)									
Sh8ii		0.582***		0.260**		0.249***			
		N		N	N		N		
Sh34d	0.408**	0.524**	0.715***	0.420**	0.469**			0.651***	0.542***
	N				N			N	N
Sh36c	0.547***	0.415***	0.352***	0.474***	0.484***	0.520***	0.450***	0.512***	0.551***
	N						N		N
Sh74 <sup>x</sup>	0.750***				1.000***		0.489*	0.748*	0.720**
	N	N		N, S**	N	N, S**		N	N
Sh88 <sup>+</sup>	0.358***							0.609***	
	N	N			N			N	
Sh89 <sup>x</sup>								0.489***	
			N	N				N	
Sh90	0.680***	0.412***	0.404***	0.259**	0.419***	0.203**		0.668***	
	N	N	N	N	N	N	N	N	N
Sh90 <sup>+</sup>	0.610***								
	N		N		N	na	na		

Figures given for significant tests are *f* values (Weir and Cockerham, 1984).

na – not assayed. \**P* < 0.05, \*\**P* < 0.01, \*\*\**P* < 0.001.

The superscript letter 'x' denotes X-linked loci.

<sup>a</sup>N – null allele segregating.

<sup>b</sup>S – for X-linked loci, significant genic differentiation between the sexes.

<sup>c</sup>H – only one heterozygote in sample.

<sup>d</sup>M – monomorphic.

( $\cong 1500$  km) from each other (Table 7). Four loci showed significant pairwise genotypic differentiation for one of the nearby but different *Hibiscus* species comparisons, and also significant differentiation for one or other of the distant but same *Hibiscus* species pairs. For each of two loci (Sh38d and Sh89) the nearby BCY and BCS populations showed significant differentiation (both *P* < 0.01), but differentiation was not significant for either of the distant but same *Hibiscus* species comparisons. Pairwise  $\theta$  estimates show strong differentiation for the nearby BCY and BCS populations, but not for BAL and TAR.

#### Divergence time of the two species

The  $D_S$  distance estimates for the average over 45 pairwise distances and for the pooled populations of each species were 0.620 and 0.603, respectively. For  $D_{TL}$ , the estimates were 0.501 and 0.365. These estimates were made using the raw genotype data, but an estimate of  $D_S$  also was obtained using the allele frequencies after correction for the presence of null alleles, and the average over the 45 pairwise distances was 0.606.

Of the 17 loci, four are pure dinucleotide repeats and the other 13 are interrupted pure, compound (one interrupted) or complex (one interrupted) (nomenclature according to Chambers and MacAvoy, 2000). Based on the estimate of Schug *et al* (1998) of a dinucleotide mutation rate of  $9.3 \times 10^{-6}$  in *D. melanogaster*, with tri- and tetranucleotide rates 6.4–8.4 times slower, and the Zhu *et al* (2000) finding of a lower rate for imperfect repeats, a mutation rate of  $1 \times 10^{-6}$  is assumed. Assuming 10 generations per year, divergence time estimates were: for the average of 45 pairwise distances and for pooled populations –  $D_S$  31 000 and 30 150 years, respectively, and for  $D_{TL}$  25 050 and 18 250 years. For five generations per year, these estimates would be doubled.

## Discussion

### Genetic variability analyses

Variability is much higher for *S. hibisci*, both for number of alleles/locus and expected heterozygosity (Table 1). Although the expected heterozygosity ( $H_E$ ) depends on

**Table 3** Significant tests of deviation from Hardy–Weinberg equilibrium for *S. acclinata* (females only for X-linked loci)

Locus	MEC	CAH	PET	DDJ	TJF
Sh8i		0.324*			
Sh8ii	M <sup>a</sup>	H <sup>b</sup>	H	M	
Sh38d					0.302*
Sh48	M		M	M	M
Sh49		0.258*			0.176*
Sh90	0.353*	0.702***	H		0.483*
	N <sup>c</sup>	N			N
Sh94				0.414***	

Figures given for significant tests are *f* values (Weir and Cockerham, 1984).

Sh13, Sh34d, Sh36c and Sh88 – no significant deviation in any population.

Sh9 – null allele segregating in PET and DDJ, significant sex difference in allele frequencies in DDJ.

Sh72, Sh81, Sh89 – monomorphic in all populations.

Sh74 – females monomorphic in all populations, one null male in each of CAH and PET.

Sh78 – one heterozygote in TJF, other populations monomorphic.

<sup>a</sup>M – monomorphic.

<sup>b</sup>H – only one heterozygote in sample.

<sup>c</sup>N – null allele segregating.

\**P* < 0.05, \*\*\**P* < 0.001.

**Table 4** F-statistics analyses for *S. hibisci*: (1) 12 loci where no null homozygotes were detected in any population (significance from permutation tests in the FSTAT program) and (2) seven loci where null homozygotes were segregating (significance from *t*-test using estimated standard deviation)

Locus <sup>a</sup>	<i>f</i> ( <i>F<sub>IS</sub></i> )	$\theta$ ( <i>F<sub>ST</sub></i> )	<i>F</i> ( <i>F<sub>IT</sub></i> )
(1)			
Sh8i	0.354 (0.065)***	0.000 (0.004)	0.354 (0.065)***
Sh9 <sup>x</sup>	0.064 (0.028)	0.011 (0.010)**	0.074 (0.026)
Sh13	0.026 (0.027)	0.004 (0.005)	0.029 (0.028)
Sh29c	0.288 (0.038)***	0.051 (0.022)***	0.324 (0.043)***
Sh38d	0.159 (0.033)***	0.002 (0.005)	0.161 (0.033)***
Sh48	−0.004 (0.004)	−0.002 (0.003)	−0.006 (0.002)
Sh49	0.189 (0.061)***	0.029 (0.017)***	0.214 (0.072)***
Sh72 <sup>x</sup>	0.040 (0.047)	0.012 (0.008)	0.051 (0.047)
Sh78 <sup>x</sup>	0.139 (0.120)	0.039 (0.025)***	0.174 (0.131)*
Sh81	0.160 (0.032)***	0.008 (0.007)	0.167 (0.034)***
Sh88	0.241 (0.036)***	0.005 (0.004)	0.245 (0.037)***
Sh94	0.185 (0.068)***	0.001 (0.004)	0.186 (0.070)***
Mean <sup>b</sup>	0.170 (0.032)***	0.014 (0.005)***	0.181 (0.033)***
(2)			
Sh8ii	0.100 (0.040)*	0.027 (0.015)	0.125 (0.044)**
Sh34d	0.188 (0.082)*	0.107 (0.056)	0.274 (0.079)***
Sh36c	0.310 (0.077)***	0.027 (0.008)**	0.328 (0.072)***
Sh74 <sup>x</sup>	0.146 (0.091)	0.049 (0.017)**	0.188 (0.094)*
Sh88 <sup>+</sup>	0.114 (0.037)**	0.124 (0.043)**	0.225 (0.056)***
Sh89 <sup>x</sup>	0.181 (0.052)***	0.007 (0.014)	0.187 (0.051)***
Sh90	−0.013 (0.010)	0.031 (0.017)	0.018 (0.023)
Mean <sup>b</sup>	0.152 (0.044)***	0.049 (0.016)**	0.193 (0.043)***

The superscript letter 'x' denotes X-linked loci.

<sup>a</sup>Standard deviation in parentheses – estimate from jackknife over populations.

<sup>b</sup>Standard deviation in parentheses – estimate from jackknife over loci.

\**P* < 0.05, \*\**P* < 0.01, \*\*\**P* < 0.001.

allele frequencies, increasing number of alleles generally leads to increased *H<sub>E</sub>*. Thus, differences in *H<sub>E</sub>* between the two species for any given locus are largely a function of differences in the number of alleles.

**Table 5** F-statistics analyses for *S. acclinata*: (1) 10 loci where no null homozygotes were detected in any population (significance from permutation tests in the FSTAT program) and (2) two loci where null homozygotes were segregating (significance from *t*-test using estimated standard deviation)

Locus <sup>a</sup>	<i>f</i> ( <i>F<sub>IS</sub></i> )	$\theta$ ( <i>F<sub>ST</sub></i> )	<i>F</i> ( <i>F<sub>IT</sub></i> )
(1)			
Sh8i	0.228 (0.030)**	0.001 (0.006)	0.229 (0.033)**
Sh8ii	0.633 (0.490)	−0.005 (0.003)	0.631 (0.491)*
Sh9 <sup>x</sup>	0.147 (0.172)	−0.025 (0.004)	0.125 (0.172)
Sh13	−0.047 (0.029)	0.012 (0.011)	−0.034 (0.032)
Sh34d	0.198 (0.073)	−0.006 (0.005)	0.192 (0.070)*
Sh38d	0.116 (0.059)	−0.008 (0.002)	0.108 (0.059)
Sh48	−0.022 (0.010)	0.022 (0.010)	0
Sh49	0.114 (0.054)**	0.012 (0.010)**	0.125 (0.045)**
Sh88	0.027 (0.030)	0.010 (0.006)	0.037 (0.031)
Sh94	0.194 (0.073)**	0.013 (0.008)*	0.205 (0.071)**
Mean <sup>b</sup>	0.107 (0.041)***	0.006 (0.003)***	0.112 (0.040)***
(2)			
Sh36c	0.045 (0.042)	0.002 (0.011)	0.047 (0.034)
Sh90	0.044 (0.015)**	0.069 (0.048)	0.109 (0.044)*
Mean <sup>b</sup>	0.045 (0.001)***	0.015 (0.040)	0.059 (0.037)

The superscript letter 'x' denotes X-linked loci.

<sup>a</sup>Standard deviation in parentheses – estimate from jackknife over populations.

<sup>b</sup>Standard deviation in parentheses – estimate from jackknife over loci.

\**P* < 0.05, \*\**P* < 0.01, \*\*\**P* < 0.001.

#### Hardy–Weinberg equilibrium and null alleles

A striking feature of the results is the large number of significant deviations from Hardy–Weinberg equilibrium in the *S. hibisci* populations (Table 2). However, these tests used the observed numbers of each genotype, so that for those loci where null homozygotes were found the number of heterozygotes is underestimated (ie heterozygotes for the null allele are not recorded). Thus, the test results were given separately for loci without and with null homozygotes present (Tables 2, 3). However, an analysis to check Mendelian segregation of the microsatellite loci in *S. hibisci* (Wilson *et al*, unpublished) found null alleles at loci additional to those seen in these populations, namely at Sh8i, Sh29c, Sh38d, Sh49, Sh81, Sh88 and Sh94. Thus, there could be undetected null alleles at these loci, particularly as all the seven show significant deviations from Hardy–Weinberg equilibrium in one or more populations.

If null alleles are suspected, but no null homozygotes are detected in the sample assayed, the putative null allele frequency can be estimated from the heterozygote deficiency (Brookfield, 1996) as:

$$(H_E - H_O)/(1 + H_E)$$

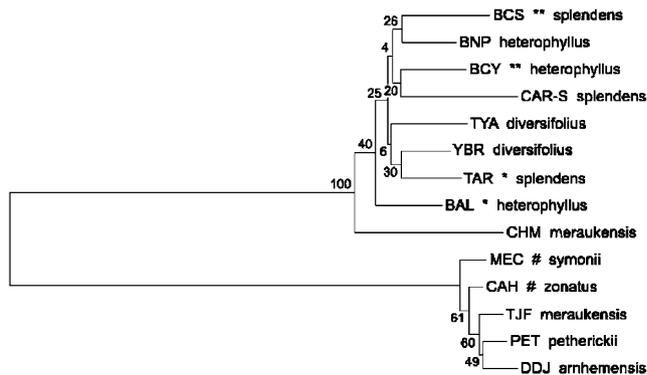
where *H<sub>O</sub>* and *H<sub>E</sub>* are the observed and expected heterozygosities, and panmixia is assumed. Using this formula, the estimated null allele frequencies for each of the 28 cases of significant deviation from Hardy–Weinberg equilibrium in Table 2, part 1 range from 0.046 to 0.288, which are generally within the range of estimated null allele frequencies at the loci where nulls were detected. At a null allele frequency of 0.154, one null homozygote would be expected in a sample of 42 individuals. Out of the 28 cases, 15 had an estimated null allele frequency less than this, so that these deviations could be due to undetected null alleles. At the extreme

**Table 6** Pairwise population differentiation ( $\theta$ ) estimates above the diagonal, and pairwise geographic distances (km) below

S. hibiscis									
	CHM	BCS	BCY	BNP	YBR	CAR-S	TYA	BAL	TAR
CHM		<b>0.037*</b>	<b>0.019*</b>	<b>0.034*</b>	<b>0.032*</b>	<b>0.032*</b>	<b>0.025*</b>	<b>0.019*</b>	<b>0.027*</b>
BCS	574		<b>0.014*</b>	0.008	<u>0.011*</u>	<b>0.011*</b>	<b>0.014*</b>	<b>0.011*</b>	<u>0.003</u>
BCY	574	1.2		0.013	<u>0.006</u>	<b>0.019*</b>	<b>0.015*</b>	0.008	<u>0.003</u>
BNP	876	304	304		0.005	<b>0.018*</b>	<b>0.011</b>	0.009	0.004
YBR	888	314	314	19		<b>0.018*</b>	<b>0.012</b>	0.006	-0.003
CAR-S	978	471	470	358	342		<b>0.019*</b>	<b>0.017*</b>	<b>0.012*</b>
TYA	1562	991	990	702	686	658		<b>0.004</b>	<b>0.006</b>
BAL	2046	1521	1521	1289	1270	1069	693		-0.001
TAR	2046	1521	1521	1288	1270	1069	692	2.1	
S. acclinata									
	MEC	CAH	PET	DDJ	TJF				
MEC		0.003	<b>0.011*</b>	<b>0.007*</b>	0.003				
CAH	6		0.003	<u>0.010*</u>	0.001				
PET	259	260		<u>0.007</u>	<b>0.007</b>				
DDJ	12	16	251		0.004*				
TJF	257	258	17	248					

Significant pairwise genotypic differentiation is also shown – indicated by  $\theta$  values in bold ( $P < 0.001$ ), italic ( $P < 0.01$ ) and underlined ( $P < 0.05$ ).

\* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$  – significance of  $\theta$  tested by permutation of genotypes among populations (10 000 times).



**Figure 2** Neighbour-joining tree for *S. hibiscis* and *S. acclinata* populations, and the *Hibiscus* species at each of these localities. Populations of *S. hibiscis* (*S. acclinata*) with the same superscript symbol are located within 2 km (# – 6 km) of each other.

estimated null allele frequency of 0.288, 3.5 null homozygotes would be expected in a sample of 42 individuals, but the probability that none would be seen is not particularly low.

This probability of getting no null homozygotes in a sample of 42 individuals from an infinite population with true null homozygote frequency of  $x$  is  $(1-x)^{42}$ . For true null homozygote frequencies of 0.0238, 0.0476 and 0.0714, corresponding to one, two or three such individuals expected in a sample of 42, the probabilities are 0.364, 0.129 and 0.045, and these probabilities are not much reduced by sampling from a large finite population, for example, sampling from 1000 individuals, the probability of not seeing one null homozygote, when the population frequency is  $3/42$ , is 0.042.

Similarly,  $f$  ( $F_{IS}$ ) values for those loci where no null homozygotes were detected (Tables 4 and 5 – part 1 in each) will be inflated if null alleles are present in the sample, but not detected. The significant  $f$  ( $F_{IS}$ ) values were only for those seven loci where null alleles were

found in the segregation analyses (Wilson *et al*, unpublished). Thus, it is likely that much of the significant deviation from Hardy–Weinberg equilibrium and significant positive  $f$  values at these loci can be attributed to undetected null alleles. For *S. acclinata*, all cases of significant deviations from Hardy–Weinberg equilibrium and significant  $f$  (except Sh90) could be attributed to undetected null alleles, assuming null alleles occur at the same loci in *S. acclinata* as in *S. hibiscis*.

Significant  $f$  ( $F_{IS}$ ) values for the loci where null homozygotes were detected (Tables 2 and 3 – part 2) would not be expected, as these estimates were obtained from genotype frequency data after estimation of null heterozygote frequencies. As this procedure assumes panmixia, the significant positive  $f$  values for these loci are likely due to deficiencies of heterozygotes for the non-null alleles. However,  $f$  values are global (over populations) estimates, and the effect of estimating null heterozygote frequencies will vary among loci, depending on the number of alleles, and on the number of populations where null homozygotes were found. Overall, it would seem most parsimonious to conclude that these populations of both species are panmictic. However, the results should raise a strong note of caution in population studies where significant deviations from Hardy–Weinberg equilibrium may be due to undetected null alleles, or where the presence of null alleles is inferred only from the population data.

#### Ascertainment bias

A number of studies have shown that microsatellite alleles are longer and more variable in the species from which they were cloned, and Ellegren *et al* (1995) and Forbes *et al* (1995) proposed that the identification of microsatellite loci by DNA library screens selects for longer loci (ascertainment bias). Here, however, the mean allele size is greater in *S. acclinata* for eight loci, and for *S. hibiscis* in nine, and the differences are not significant

**Table 7** Genetic differentiation between geographically close populations of *S. hibisci* breeding on different *Hibiscus* species, as compared with differentiation of these populations breeding on the same *Hibiscus* species, but geographically distant

	BAL/TAR het/spl <sup>a</sup>	BCY/BCS het/spl	BAL/BCY het/het	TAR/BCS spl/spl
Distance (km)	2.1	1.2	1521	1521
Pairwise differentiation <sup>b</sup> ( $\chi^2$ )				
	30.2	67.9***	36.3	47.3*
$D_A$	0.081	0.087	0.081	0.089
Pairwise $\theta$	-0.001	0.014*	0.008	0.003

<sup>a</sup>het = *H. heterophyllus*, spl = *H. splendens*.

<sup>b</sup>Genotypic differentiation – estimated using GENEPOP.

\* $P < 0.05$ , \*\*\* $P < 0.001$ .

for any locus. Further, the *S. acclinata*-specific alleles were equally at or near the upper size range and the lower size range. Thus, we have no evidence for ascertainment bias.

In a reciprocal test of ascertainment bias in *D. melanogaster* and *D. simulans*, Hutter *et al* (1998) found both number of alleles and expected heterozygosity to be significantly higher in the species from which the microsatellite loci were derived, although differences between these species were much less than for *S. hibisci* and *S. acclinata*. Thus, differences in population history for these latter species more likely contributed to the marked differences in levels of genetic variability.

#### Juxtaposed microsatellites

Previous studies (Pena *et al*, 1994; Dermitzakis *et al*, 1998) have found strong linkage disequilibrium between juxtaposed loci, but, in contrast to these studies, *S. hibisci* shows no linkage disequilibrium and a weak negative, nonsignificant, allele size association. Both previous studies suggest that juxtaposed microsatellites may not evolve independently, but the result for *S. hibisci* could indicate allele size homoplasy coupled with high rates of microsatellite evolution or long-lived alleles and rare recombination. The general evolutionary features of juxtaposed microsatellites remain to be elucidated.

#### Population structure

The BOTTLENECK tests showed that no populations of *S. acclinata* significantly deviated from expectation, but all populations of *S. hibisci* (except YBR) had a significant heterozygosity deficiency, suggesting that these populations had expanded in size in the last few thousand generations, or that there had been a recent influx of rare alleles from other populations. The latter seems most unlikely, given the magnitude of genetic differentiation among populations and the geographic distances separating most populations. The BOTTLENECK method assumes that the loci are independent, or at least not tightly linked (recombination fraction  $> 0.1$ ). In all, 16 of the loci have been assigned to three linkage groups (Wilson *et al*, unpublished), but limited data do not indicate any tight linkages. Further, there was no evidence for linkage disequilibrium, even for the juxtaposed pair, so that any effect of linkage is likely to be very small. The BOTTLENECK method also assumes that all loci are in Hardy–Weinberg equilibrium. Although only those loci that did not show deviations from Hardy–Weinberg equilibrium were used in these tests, undetected null alleles would bias the results towards heterozygosity deficiency. This could account for some of the significant heterozygosity deficiency, but

most likely not all, thus implying population size expansion for *S. hibisci*, but not *S. acclinata*. The host *Hibiscus* species for *S. hibisci* are known as colonizers of disturbed habitats such as forest clearings and edges, and roadsides. In the last few thousand generations of *S. hibisci*, encompassing the period of European settlement in Australia since 1788, individual *Hibiscus* patches may well have increased in size (number of plants), but at the same time become more isolated through forest clearing and agricultural development. The *Hibiscus* populations on which *S. acclinata* breeds in the Northern Territory, in contrast, are located in areas that have not been subject to agricultural development, and have likely been at their present size for a very long time.

These differences in population history between the two species could also contribute to the lower number of alleles in *S. acclinata*. Allele numbers for X-linked loci are markedly reduced, as are private allele frequencies – 5.44 per population for *S. hibisci* and 3.2 per population for *S. acclinata*. Both these results are consistent with greater drift effects (smaller effective size) for *S. acclinata* populations. The flowering periods for the *S. acclinata* host species are generally shorter than for the *S. hibisci* host species, which is likely to produce an annual bottleneck that is smaller for *S. acclinata*.

Overall estimates of both  $\theta$  and genotypic differentiation of populations were significant for both species, although much higher for *S. hibisci* than for *S. acclinata*. It might be assumed that this species difference is a function of the much broader geographical distribution of the *S. hibisci* populations, that is, greater isolation. In fact, it does appear to be an isolation effect, but for only the two populations CAR-S and CHM. Mean  $\theta$  values for each of CAR-S and CHM with the other populations are 0.018 and 0.028 respectively, against an overall mean of 0.014. When CAR-S and CHM are excluded, the overall mean  $\theta$  is 0.008, essentially the same as for the overall mean for *S. acclinata* (0.006 – Table 5). Further, both populations show strong evidence of drift or selection effects on allele frequency distributions. For example, at Sh8ii the allele at highest frequency in CAR-S (0.321) has frequencies of 0.012–0.105 in the other populations, while for Sh34d the third most frequent allele in CHM (0.135) is the highest frequency allele in all other populations (0.486–0.833).

The CAR-S population is located in Carnarvon Gorge (Figure 1), a mesic habitat in contrast to the predominant open forest and grassland of central Queensland. The nearest known *Hibiscus* populations (*H. divaricatus* and *H. splendens*) are in similar habitat on the Blackdown Tableland some 160–180 km to the northeast (records of the Queensland Herbarium). The CAR-S population thus

is likely a relict population that has been isolated at least since the cool, dry periods of the Pleistocene. CHM, from the northern part of the *S. hibisci* distribution (Figure 1), is separated from populations to the south by the dry Burdekin gap (Kelemen and Moritz, 1999), which is to the north of the BCS and BCY populations, and this gap also has isolated the northern populations since the Pleistocene.

For the *S. hibisci* populations (other than CAR-S and CHM) and the *S. acclinata* populations, migration cannot explain the history and current population structure. If migration were an important component, either no population differentiation or isolation by distance would be expected. Neither is true: population differentiation is significant, and no isolation by distance was detected. Indeed as noted earlier, both species show patterns of pairwise population differentiation that are inconsistent with migration.

#### Host-plant specialization

Although based on only two pairs of populations breeding on *H. heterophyllus* and *H. splendens*, the early stages of host-race formation possibly have been detected (Table 7). Clearly, this evidence is stronger for BCY/BCS, where also the results showing that two loci are significantly differentiated for the nearby but different *Hibiscus* species comparisons, but not significant for the distant but same *Hibiscus* species comparisons, suggest selectively driven differences. Of course, these microsatellite loci are assumed to be neutral, but they may themselves be, or they may mark regions under selection.

A laboratory experiment (McEvey and Barker, 2001) found no evidence for host-plant specialization in *S. hibisci*, but only for *H. heterophyllus* and *H. diversifolius* and recording only progeny number. Other fitness components such as adult longevity and lifetime progeny numbers may be important, and specialization may not have evolved for all *Hibiscus* species. Thus, further laboratory studies and analyses of additional nearby population pairs are necessary to test the hypothesis of host-race formation.

All of the northern populations of *S. hibisci* (Figure 1) are breeding on *H. meraukensis*, a species that also is used by *S. acclinata* in the Northern Territory. Direct comparisons of *S. hibisci* and *S. acclinata* populations from *H. meraukensis* would help to elucidate the phylogeography and evolution of these two species. Further, genetic analyses of the northern *S. hibisci* populations breeding on *H. meraukensis* would test whether the genetic differentiation of CHM is entirely a function of isolation and drift, or whether there is some degree of host-plant specialization. *Hibiscus meraukensis* is known (records of the Queensland Herbarium) from a number of localities in northwest Queensland – the region between the known distributions of *S. hibisci* and *S. acclinata*. This region (the Carpentaria gap), recognized as a phylogeographic break for many species (primarily birds), was fairly arid at the glacial maximum of 20 000–15 000 years BP (Ford, 1987). It is not known if either *S. hibisci* or *S. acclinata* are now present and breeding on *H. meraukensis* in this region, but it is hoped that this will be investigated.

#### Divergence time of the two species

Can any conclusion be reached on the evolutionary divergence time of *S. acclinata* and *S. hibisci*? Clearly, they

are closely related. McEvey and Barker (2001) showed them to be partially interfertile, with reduced mating frequency and progeny numbers in  $F_1$  *inter se* matings as compared with parentals, and no progeny from  $F_2$  *inter se* matings. They have the same common allele at seven of 17 microsatellite loci, and share a high proportion of the alleles detected (82 of 253). Of the remaining alleles, 160 were present only in *S. hibisci*, but only 11 were present only in *S. acclinata*. Given the close relationship between the species, and the lower genetic diversity in *S. acclinata*, the allele numbers in the two species suggest that *S. acclinata* was derived from *S. hibisci*, following a strong bottleneck at the time of separation.

Genetic distance estimates ( $D_S$  and  $D_{TL}$ ) were similar in magnitude, with estimates from the average of the 45 pairwise estimates higher than those from the pooled population data, as expected (Goldstein *et al*, 1995). Estimated divergence times range from 18 250 ( $D_{TL}$ , pooled populations, 10 generations per year) to 62 000 years ( $D_S$ , average of pair-wise distances, five generations per year), with an average of the eight estimates of 39 169 years. A divergence time of 40 000 years would put it prior to the glacial maximum at 20 000–15 000 years BP when the Carpentaria gap was less arid, so that movement across this area was likely to be possible. However, given uncertainty regarding the appropriate mutation model, the mutation rate, constraints on allele size distribution and particularly size homoplasy, the true divergence time may be much longer.

#### Acknowledgements

This research was part of a larger collaborative program on the reproductive biology and evolution of *S. hibisci* funded by a grant from the NSF (INT-9512767) to WT Starmer and LL Wolf. Scientists from Australia (JSF Barker), Canada (MA Lachance, J Bowles) and the USA (M Polak, WT Starmer, LL Wolf) took part in field observations and collections of the populations. The laboratory work reported here was supported by an Australian Research Council grant and a University of New England research grant. I am indebted to Yean Wang and Pam Flynn for technical assistance, to L Craven and B Pfeil for advice on *Hibiscus* taxonomy, and to V Loeschcke, WT Starmer and P Sunnucks for helpful comments on the manuscript.

#### References

- Barker JSF (2003). Ecology of two *Scaptodrosophila* flower breeding species. *Drosophila Inf Serv* **86**: 38–40.
- Berlacher SH, Feder JL (2002). Sympatric speciation in phytophagous insects: Moving beyond controversy? *Ann Rev Entomol* **47**: 773–815.
- Brookfield JFY (1996). A simple new method for estimating null allele frequency from heterozygote deficiency. *Mol Ecol* **5**: 453–455.
- Chambers GK, MacAvoy ES (2000). Microsatellites: consensus and controversy. *Comp Biochem Physiol Part B* **126**: 455–476.
- Cook RM, Parsons PA, Bock IR. (1977). Australian endemic *Drosophila* II. A new *Hibiscus*-breeding species with its description. *Aust J Zool* **25**: 755–763.
- Cornuet JM, Luikart G (1996). Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics* **144**: 2001–2014.
- Dermitzakis ET, Clark AG, Batargias C, Magoulas A, Zouros E (1998). Negative covariance suggests mutation bias in a

- two-locus microsatellite system in the fish *Sparus aurata*. *Genetics* **150**: 1567–1575.
- Ellegren H, Primmer CR, Sheldon BC (1995). Microsatellite 'evolution': directionality or bias? *Nature Genetics* **11**: 360–362.
- Etges WJ (2002). Divergence in mate choice systems: does evolution play by rules? *Genetica* **116**: 151–166.
- Feder JL, Berlocher SH, Opp SB (1998). Sympatric host-race formation and speciation in *Rhagoletis* (Diptera:Tephritidae): a tale of two species for Charles D. In: Mopper S, Strauss SY (eds) *Genetic Structure and Local Adaptation in Natural Insect Populations*. Chapman & Hall: New York. pp 408–441.
- Forbes SH, Hogg JT, Buchanan FC, Crawford AM, Allendorf FW (1995). Microsatellite evolution in congeneric mammals: domestic and bighorn sheep. *Mol Biol Evol* **12**: 1106–1113.
- Ford J (1987). Hybrid zones in Australian birds. *Emu* **87**: 158–178.
- Funk DJ, Filchak KE, Feder JL (2002). Herbivorous insects: model systems for the comparative study of speciation ecology. *Genetica* **116**: 251–267.
- Goldstein DB, Linares AR, Cavalli-Sforza LL, Feldman MW (1995). Genetic absolute dating based on microsatellites and the origin of modern humans. *Proc Natl Acad Sci USA* **92**: 6723–6727.
- Goudet J (2001). FSTAT, a program to estimate and test gene diversities and fixation indices Version 2.9.3. Available from <http://www.unil.ch/izea/software/fstat.html>.
- Hochberg Y (1988). A sharper Bonferroni procedure for multiple tests of significance. *Biometrika* **75**: 800–802.
- Hutter CM, Schug MD, Aquadro CF (1998). Microsatellite variation in *Drosophila melanogaster* and *Drosophila simulans*: a reciprocal test of the ascertainment bias hypothesis. *Mol Biol Evol* **15**: 1620–1636.
- Janz N, Nyblom K, Nylin S (2001). Evolutionary dynamics of host-plant specialization: a case study of the tribe Nymphalini. *Evolution* **55**: 783–796.
- Kelemen L, Moritz C (1999). Comparative phylogeography of a sibling pair of rainforest *Drosophila* species (*Drosophila serrata* and *D. birchii*). *Evolution* **53**: 1306–1311.
- Lessios HA (1992). Testing electrophoretic data for agreement with Hardy–Weinberg expectations. *Marine Biol* **112**: 517–523.
- Mantel N (1967). The detection of disease clustering and a generalized regression approach. *Cancer Res* **27**: 209–220.
- McEvey SF, Barker JSF (2001). *Scaptodrosophila acinata*: a new *Hibiscus* flower-breeding species related to *S. hibisci* (Diptera: Drosophilidae). *Records Austr Museum* **53**: 255–262.
- Mopper S, Strauss SY (eds) (1998). *Genetic Structure and Local Adaptation in Natural Insect Populations*. Chapman & Hall: New York.
- Nauta MJ, Weissing FJ (1996). Constraints on allele size at microsatellite loci: implications for genetic differentiation. *Genetics* **143**: 1021–1032.
- Nei M (1972). Genetic distance between populations. *Am Nat* **106**: 283–292.
- Nei M (1978). Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* **89**: 583–590.
- Nei M, Tajima F, Tatenno Y (1983). Accuracy of estimated phylogenetic trees from molecular data II. Gene frequency data. *J Mol Evol* **19**: 153–170.
- Pena SDJ, de Souza KT, de Andrade M, Chakraborty R (1994). Allelic associations of two polymorphic microsatellites in intron 40 of the human von Willebrand factor gene. *Proc Natl Acad Sci USA* **91**: 723–727.
- Piry S, Luikart G, Cornuet JM (1999). BOTTLENECK: a computer program for detecting recent reductions in the effective population size using allele frequency data. *J Hered* **90**: 502–503.
- Polak M, Starmer WT, Barker JSF (1998). A mating plug and male mate choice in *Drosophila hibisci* Bock. *Anim Behav* **56**: 919–926.
- Polak M, Wolf LL, Starmer WT, Barker JSF (2001). Function of the mating plug in *Drosophila hibisci* Bock. *Behav Ecol Sociobiol* **49**: 196–205.
- Pritchard JK, Feldman MW (1996). Statistics for microsatellite variation based on coalescence. *Theor Popul Biol* **50**: 325–344.
- Raymond M, Rousset F (1995). GENEPOP (Version 1.2): population genetics software for exact tests and ecumenicism. *J Hered* **86**: 248–249.
- Rousset F (1997). Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics* **145**: 1219–1228.
- Saitou N, Nei M (1987). The neighbour-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**: 406–425.
- Schug MD, Hutter CM, Wetterstrand KA, Gaudette MS, Mackay TF, Aquadro CF (1998). The mutation rates of di-, tri- and tetranucleotide repeats in *Drosophila melanogaster*. *Mol Biol Evol* **15**: 1751–1760.
- Sokal RR, Rohlf FJ (1981). *Biometry*, 2nd edn. W.H. Freeman and Co.: San Francisco.
- Starmer WT, Polak M, Wolf LL, Barker JSF (1998). Reproductive characteristics of the flower breeding *Drosophila hibisci* Bock (Drosophilidae) in eastern Australia: genetic and environmental determinants of ovariole number. *Evolution* **52**: 806–815.
- Starmer WT, Polak M, Wolf LL, Barker JSF (2000). Reproductive characteristics of the flower-breeding *Drosophila hibisci* Bock (Drosophilidae) in eastern Australia: within-population genetic determinants of ovariole number. *Heredity* **84**: 90–96.
- Starmer WT, Wolf LL, Barker JSF, Bowles JM, Lachance M-A (1997). Reproductive characteristics of the flower breeding *Drosophila hibisci* Bock (Drosophilidae) along a latitudinal gradient in eastern Australia: relation to flower and habitat features. *Biol J Linn Soc* **62**: 459–473.
- Stephens M, Smith NJ, Donnelly P (2001). A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* **68**: 978–989.
- Takezaki N, Nei M (1996). Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. *Genetics* **144**: 389–399.
- Thomas Y, Bethenod M-T, Pelozuelo L, Frérot B, Bourguet D (2003). Genetic isolation between two sympatric host-plant races of the European corn borer, *Ostrinia nubilalis* Hübner. I. Sex pheromone, moth emergence timing and parasitism. *Evolution* **57**: 261–273.
- Tomiuk J, Guldbandsen G, Loeschcke V (1998). Population differentiation through mutation and drift – a comparison of genetic identity measures. *Genetica* **102/103**: 545–558.
- Tomiuk J, Loeschcke V (1991). A new measure of genetic identity between populations of sexual and asexual species. *Evolution* **45**: 1685–1694.
- Tomiuk J, Loeschcke V (1995). Genetic identity combining mutation and drift. *Heredity* **74**: 607–615.
- Via S (2001). Sympatric speciation in animals: the ugly duckling grows up. *Trends Ecol Evol* **16**: 381–390.
- Weir BS (1990). *Genetic Data Analysis*. Sinauer: Sunderland.
- Weir BS, Cockerham CC (1984). Estimating F-statistics for the analysis of population structure. *Evolution* **38**: 1358–1370.
- Wilson ACC, Sunnucks P, Barker JSF (2002). Isolation and characterization of 20 polymorphic microsatellite loci for *Scaptodrosophila hibisci*. *Mol Ecol Notes* **2**: 242–244.
- Wolf LL, Polak M, Barker JSF, Bowles J, Starmer WT (2000a). Reproductive characteristics of *Drosophila hibisci* in the Northern Territory, Australia. *Biol J Linn Soc* **71**: 549–562.
- Wolf LL, Starmer WT, Polak M, Barker JSF (2000b). Genetic architecture of a wing size measure in *Drosophila hibisci* from two populations in eastern Australia. *Heredity* **85**: 521–529.
- Zhu Y, Queller DC, Strassmann JE (2000). A phylogenetic perspective on sequence evolution in microsatellite loci. *J Mol Evol* **50**: 324–338.