Animal Genetics,
1997, **28,** 103–115

Genetic diversity of Asian water buffalo (*Bubalus bubalis*): microsatellite variation and a comparison with protein-coding loci

J S F Barker, S S Moore, D J S Hetzel, D Evans, S G Tan, K Byrne

Summary

Twenty-one microsatellite loci in 11 populations of Asian water buffalo (eight swamp, three river type) were analysed and, within and among populations, genetic variability was compared with results from 25 polymorphic protein-coding loci. Within-population mean heterozygosity ranged from 0·380–0·615, approximately twice that estimated from the protein-coding loci (0·184–0·346). Only eight significant departures from Hardy–Weinberg equilibrium (involving four loci) were detected; global tests showed significant heterozygote deficiencies for these four loci. Non-amplifying alleles are likely to be segregating in some or all populations for one of these loci, and probably for the other three. There was significant differentiation between the swamp and river types of water buffalo, and among populations within each buffalo type. Estimates of θ (measure of population differentiation) for each locus for the eight swamp populations were all highly significant (mean θ = 0·168 ± 0·018). Mean θ for protein-coding loci was not significantly different (0·182 ± 0·041). The variance among protein-coding loci was significantly higher than among microsatellite loci, suggesting balancing selection affecting allele frequencies at some protein-coding loci. Genetic distances show clear separation of the swamp and river types, which were estimated to have diverged at least 10 000–15 000 years ago. The topology of the swamp populations' microsatellite tree is consistent with their geographical distribution and their presumed spread through south-east Asia. By contrast, the tree based on the protein-coding loci distances is quite different, being clearly distorted by a bottleneck effect in one population, and possibly in at least two others. As many domestic livestock breeds are possibly descended from small numbers of founders, microsatellite-based trees are to be preferred in assessing breed genetic relationships.

Keywords: allozymes, Asian water buffalo, *Bubalus bubalis*, genetic diversity, microsatellites, molecular phylogenetics

Correspondence: J S F Barker.

Accepted 29 November 1996

J S F Barker
Department of Animal Science, University of New England, Armidale NSW 2351, Australia
S S Moore
D J S Hetzel
D Evans
K Byrne
CSIRO Division of Tropical Animal Production, Molecular Animal Genetics Centre, University of Queensland, Brisbane, Australia
S G Tan
Department of Biology, Universiti Pertanian Malaysia, 43400 UPM, Serdang, Malaysia

Introduction

In domestic livestock species, subdivision or structuring is generally more obvious than in wild species, in that separate breeds are recognized. In the broadest sense, the term breed includes strains and populations, the members of which are distinguished from other such groups in local, national or regional usage; i.e. a breed is a cultural entity, recognized as such by the community where it is found. Early studies of livestock breeds (Rendel 1967; Kidd 1974) emphasized breed relationships, not just the magnitude of differentiation, and the objectives of these and later studies have been to aid in understanding domestication, breed origins and their history and evolution, to identify genetically unique breeds, to provide an objective basis for conservation decisions and to aid the formulation of breeding plans (Barker 1994).

Until very recently, the many studies of genetic structure have used allele frequency data at protein coding (primarily allozyme) loci (e.g. Ward *et al.* 1992 summarize results for more than 300 animal species). Since the late 1970s, molecular methods have provided new markers for the study of genetic variation, even to the level of analysis at the DNA sequence itself (Hillis & Moritz 1990; Avise 1994).

Among these molecular markers, simple-sequence repeat (microsatellite) loci have been found to be common in all eukaryotic genomes so far examined, with frequencies as high as one every 6 kb (Beckmann & Weber 1992). As microsatellites are also highly polymorphic, they provide extremely useful markers for comparative studies of genetic variation, parentage assessment and studies of gene flow and hybridization (Bruford & Wayne 1993; Roy *et al.* 1994), and could well be the markers of choice for analyses of population structure in both wild (Scribner *et al.* 1994) and domesticated species (MacHugh *et al.* 1994; van Zeveren *et al.* 1995).

In the study of livestock breeds where the primary focus is on estimating breed relationships, it is assumed that the markers are neutral to

selection and the populations are in equilibrium under drift and migration, so that the breed similarities reflect common ancestry. This may not be true for all classes of markers, or for particular markers within a class, so our understanding of population structure and breed relationships will benefit from comparative studies utilizing different classes of markers. Where results are concordant, support for the estimated relationships will be increased, where they are not, investigating reasons for the difference will increase our understanding of factors affecting structure and relationships.

Barker *et al.* (1996) presented an analysis of 17 water buffalo populations of south-east Asia, using data on 53 protein-coding (primarily allozyme) loci (25 polymorphic). We present here the results from an analysis of 21 microsatellite loci, assayed in a subset of the same animals in 11 of these populations. Relationships between the swamp and river types and among geographically distinct swamp populations are considered, and measures of genetic structure and relationships, derived from the allozyme and microsatellite data, are compared.

Materials and methods

Sample collection

Blood collection and treatment was as described by Barker *et al.* (1996). DNA from the white blood cell samples was extracted by an organic solvent method. The white cells were washed in phosphate-buffered saline by alternate centrifugation and resuspension, then lysed with SDS detergent in the present of 10 mM Tris buffer (pH 7·8), 25 mM EDTA and 100 µg/ml proteinase-K. The samples were then incubated at 55°C for 60 min, following which incubation was continued overnight in a shaking waterbath at 37°C. The sample was then extracted with an equal volume mix of saturated NaCl and chloroform. The aqueous phase of this extraction was combined with twice its volume of ethanol to precipitate the DNA, which was washed in 70% ethanol and, finally, resuspended in 10 mM Tris/25 mM EDTA buffer.

The populations and numbers of animals (in brackets) studied were as follows. Swamp buffalo: Thailand - Surin (25); Malaysia -Trengganu (25), Sabah (25), Sarawak (25); Indonesia - Bogor (25), Sulawesi (25); Philippines - Musuan (26); and Australia (23). Lankan buffalo: South Sri Lanka (23). River buffalo: Sri Lanka - Murrah (25); and Malaysia - Murrah (14). For each population, the animals assayed were a subset of those used by Barker *et al.* (1996) (see their Fig. 1 for the location of each population).

Microsatellite isolation and genotype assay

Forty-six of a panel of 80 bovine microsatellites were polymorphic in a sample of five swamp buffalo; 21 of these microsatellites were chosen for this study based solely on the 'robustness' (yield and reproducibility) of the amplification (Moore *et al.* 1995). Fifty micrograms of buffalo DNA was used as template for polymerase chain

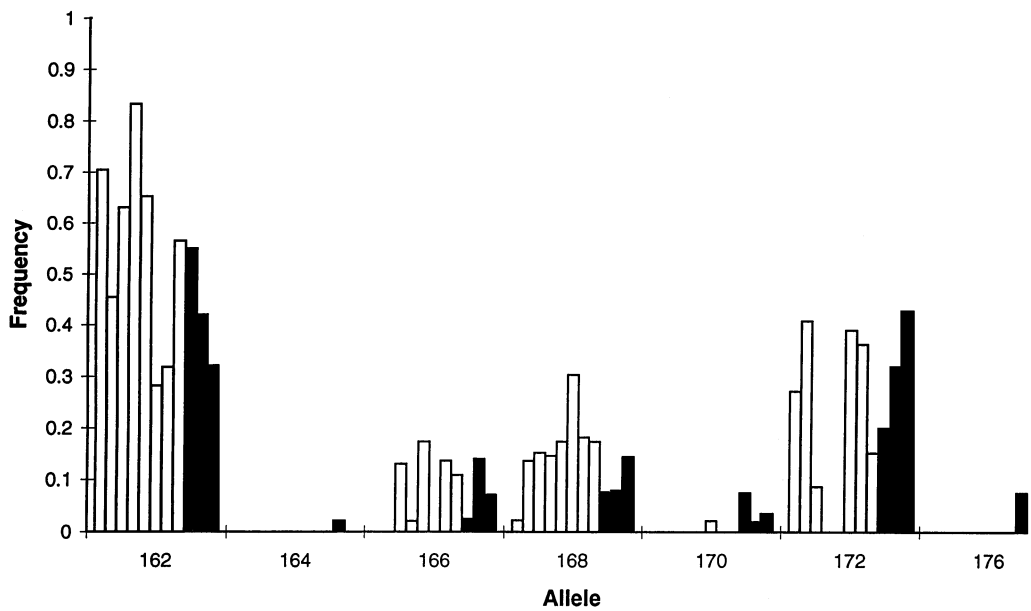

Fig. 1. Allele frequency distribution for microsatellite locus CSSM036, where the same allele was at the highest frequency in both swamp and river buffalo. Open bars, swamp populations; filled bars, river populations.

reaction (PCR) as described by Moore *et al.* (1995). Visualization was achieved by incorporation of [^{32}P-dCTP] during the PCR cycles and subsequent autoradiography of dried gels. Allele sizes were determined by comparison with pUC19 sequencing ladders. The number of animals genotyped per population–locus combination was generally less than the number sampled per population. For 11 animals, the amount of DNA available enabled genotyping of only 3–13 loci. A further 24 animals failed to amplify for three or more (up to 10) loci, presumably because of poor quality DNA. All cases, where amplification failed, were rerun at least once.

Statistical analyses

Allele frequencies and mean heterozygosity estimates for each population were obtained using the BIOSYS-1 computer program (Swofford & Selander 1989). Tests of genotype frequencies for deviations from Hardy–Weinberg equilibrium were carried out using the exact tests of the computer program GENEPOP (Raymond & Rousset 1995). Significance levels for each test were determined by applying, to the probability estimates calculated by GENEPOP, the sequential Bonferroni procedure (Hochberg 1988; Lessios 1992) over loci within each population.

The number of alleles expected under the infinite alleles neutral mutation model for each locus in each population was computed (Ewens 1972; Ely *et al.* 1992 – correcting equation 5 in the latter paper by replacing unity in the numerators with θ). Means of observed and expected numbers of alleles were calculated for each locus, separately for swamp and river types, and paired *t*-tests were performed to determine if there were significant differences.

Hierarchical *F*-statistics were computed by using the methods of Weir & Cockerham (1984), as implemented in the computer program GDA (P. Lewis, personal communication), and the significance of *F*-statistics estimates was determined using the permutation tests in the FSTAT program (Goudet 1995). The sequential Bonferroni procedure was applied over loci in deriving significance levels. Mean gene diversity for each locus and genetic distances among populations (standard genetic distance of Nei (1978) and the D_A distance of Nei *et al.* (1983)) were obtained using the DISPAN computer program (T. Ota, personal communication). The $(\delta\mu)^2$ distances (Goldstein *et al.* 1995) among populations were obtained using the MICROSAT computer program (Minch *et al.* 1995) and Reynold's distances (Reynolds *et al.* 1983) using PHYLIP 3·5c (Felsenstein 1993). The method of Slatkin (1993), as implemented in GENEPOP, was used to assess the genetically effective migration rate ($\hat{M} = Nm$, the number of migrants exchanged per generation), and to test for isolation by distance. Pairwise values of *N*m were calculated from θ (Weir & Cockerham 1984). The pairwise log (*N*m) values were then correlated with log (geographic distance) between each pair of populations, and the significance of the association estimated using Mantel's (1967) permutation test. A significant association indicates genetic structuring, and limited dispersal.

Estimates of Nei's standard genetic distance (*D*) and assumed mutation rates of microsatellite loci (α) were used to estimate the time of divergence (*t*, in generations) between the swamp and river buffalo types, between swamp and the Lankan buffalo, and among swamp populations, where $D = 2\alpha t$ (Nei 1976).

Results

In all presentation of results, the Lankan buffalo (South Sri Lanka population) is grouped with the two Murrah breed populations as river buffalo, although morphologically it resembles the swamp type. This classification is based on cytological evidence – both Lankan and river buffalo have the same chromosome number ($2n = 50$), as compared with swamp buffalo ($2n = 48$) (Scheurmann *et al.* 1974; Bongso *et al.* 1977), and on the phylogenetic relationships derived using protein-coding loci (Barker *et al.* 1996).

Genetic variability

A table of allele frequencies for each of the 21 microsatellite loci in each of the 11 populations is available from the senior author of this article (J S F Barker). The number of alleles observed at each locus in the swamp and river types, and the numbers shared between the two types, are given in Table 1. The number of alleles per locus varied from two (CSSM015) to 20 (CSSM047), but only for locus CSSM015 were all detected alleles found in both swamp and river types. The mean number of alleles per locus is similar in the two types, but there are marked differences for some loci, e.g. CSSM019 with 14 alleles in swamp buffalo and only five in river buffalo; CSSM047 with nine alleles in swamp buffalo and 18 in river buffalo. Other loci show similar numbers of alleles in the two buffalo types, but with very few common to both (e.g. CSSM041 and CSRM060). Representative allele frequency distributions (Figs 1 & 2) give examples where (1) the same allele is at the highest

frequency in both types and distributions are similar (CSSM036), and (2) alleles at the highest frequency in one type are absent from the other, and the overall distribution is bimodal (CSSM019). Average gene diversity (Nei 1973) over all loci is 0·514 (Table 1) while, for individual loci, average gene diversities range from 0·104 (HMH1R) to 0·711 (CSSM047). Across loci, average gene diversity increases with increasing number of alleles (regression coefficient = 0·021 ± 0·006, $P < 0·01$). Two loci, HMH1R and CSSM045, appear to be outliers with the lowest diversity estimates but, even when these are excluded, the regression coefficient of gene diversity on number of alleles remains significant ($b = 0·013 \pm 0·004$, $P < 0·01$).

Seventeen loci were polymorphic in all populations, while four were monomorphic in one to three populations (namely CSSM038 in Sabah and Sarawak; CSSM045 in Sabah, Sarawak and Surin; BRN in Sarawak; and HMH1R in Sabah and Sarawak). Measures of genetic variation for each population (mean number of alleles per locus and observed and expected heterozygosity) are given in Table 2. Observed heterozygosity is, on average, less than expected in both types, and heterozygosity is lower in swamp buffalo than in the river type. Among populations, heterozygosity is lower in Sabah, Sarawak and Australia.

Of the 231 locus–population combinations, only eight showed significant deviations of observed genotypic proportions from Hardy–Weinberg equilibrium. Six populations had a significant test for only one locus, and one population (Surin) had two, with all eight showing a deficiency of heterozygotes. These significant deviations involved four loci (CSSM022, Trengganu; CSSM033, Sarawak and Sri Lanka South; CSSM046, Australia, Philippines, Sabah and Surin; and CSRM060, Surin); the global tests across populations were significant only for these four loci. For locus CSSM046, all populations except Murrah Sri Lanka showed an observed deficiency of heterozygotes.

Table 1. Number of microsatellite alleles at each locus in swamp and river buffalo, number shared between the two types and average gene diversity within populations

Locus	Numbers of alleles				Average gene diversity
	Total	Swamp	River	Shared	
CSSM008	7	5	7	5	0·555
CSSM013*	5	5	2	2	0·499
CSSM015	2	2	2	2	0·406
CSSM019	15	14	5	4	0·640
CSSM022	6	5	6	5	0·431
CSSM029	11	10	5	4	0·532
CSSM032	9	7	6	4	0·549
CSSM033	11	9	7	5	0·601
CSSM036*	7	5	7	5	0·583
CSSM038	13	8	9	4	0·486
CSSM041	11	7	5	1	0·514
CSSM043	12	9	9	6	0·683
CSSM045	7	6	4	3	0·249
CSSM046*	6	5	4	3	0·506
CSSM047	20	9	18	7	0·711
CSSM057	11	10	7	6	0·652
CSRM060	15	8	9	2	0·508
CSSM061	12	9	10	7	0·594
CSSM062	8	5	6	3	0·541
BRN	10	9	6	5	0·443
HMH1R*	3	2	2	1	0·104
Mean/locus	9·57	7·10	6·48	4·00	0·514
(SE)	(0·94)	(0·62)	(0·76)	(0·39)	

*Same allele is at highest frequency in both types.

Expected numbers of alleles

The number of alleles expected, under the infinite alleles model of neutral mutation, was calculated separately for the swamp and river types and compared with the numbers observed (Table 3). In paired *t*-tests for each of the 21 loci, 15 were significant for swamp, and four for river, with the observed number less than expected in every case. With only three populations of the river type, the test is not sensitive. There is no statistical test yet available of the number of alleles expected under the stepwise mutation model, but the high proportion of significant differences from the infinite alleles model indicate a likely closer fit to the stepwise mutation than to the infinite alleles model.

Analyses of F*-statistics*

Although our data suggest a closer fit to the stepwise mutation model, Weir & Cockerham's (1984) θ statistic is used here, rather than the R_{ST} statistic of Slatkin (1995), which was derived specifically under the assumptions of the generalized stepwise model. R_{ST} and θ are not expected to be greatly different for short-term differentiation of populations within a species (Slatkin 1995), and we wish to compare these microsatellite results with those from protein-coding loci (Barker *et al.* 1996). Mean estimates from jackknifing over loci for θ_T (differentiation between types, i.e. swamp vs. river) and θ_S (differentiation between populations within types) were 0·178 ± 0·037 and 0·295 ± 0·030, respectively. However, separate analyses of swamp and river populations showed that dif-

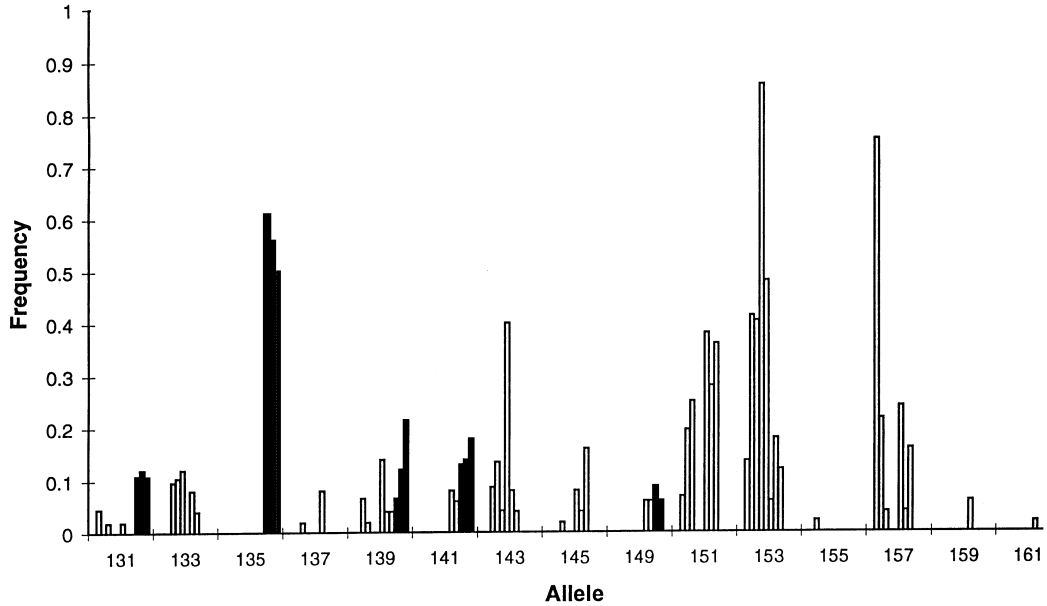

Fig. 2. Allele frequency distribution for microsatellite locus CSSM019, where the distributions are partially overlapping in swamp and river buffalo. Open bars, swamp populations; filled bars, river populations.

Table 2. Sample size, number of alleles per locus and heterozygosity (standard errors in parenthesis) averaged over 21 microsatellite loci in 11 water buffalo populations

Buffalo type and population	Mean sample size	Mean no. of alleles per locus	Mean heterozygosity	
			Observed	Expected*
Swamp				
Surin	23·1 (0·5)	5·2 (0·5)	0·589 (0·057)	0·615 (0·050)
Trengganu	23·4 (0·3)	4·5 (0·4)	0·500 (0·053)	0·578 (0·050)
Sabah	22·9 (0·4)	2·7 (0·3)	0·400 (0·056)	0·380 (0·049)
Sarawak	23·9 (0·3)	2·6 (0·3)	0·417 (0·055)	0·405 (0·052)
Bogor	21·5 (0·4)	4·0 (0·3)	0·516 (0·052)	0·540 (0·048)
Sulawesi	22·9 (0·6)	3·9 (0·3)	0·537 (0·047)	0·564 (0·041)
Philippines	24·7 (0·4)	4·7 (0·4)	0·499 (0·045)	0·543 (0·042)
Australia	21·6 (0·4)	3·0 (0·2)	0·409 (0·047)	0·425 (0·038)
Average			0·483	0·506
River				
Sri Lanka South	21·3 (0·4)	5·3 (0·5)	0·531 (0·040)	0·565 (0·045)
Murrah - Sri Lanka	24·1 (0·3)	5·0 (0·4)	0·613 (0·040)	0·607 (0·033)
Murrah - Malaysia	13·7 (0·1)	4·2 (0·5)	0·531 (0·046)	0·564 (0·047)
Average			0·558	0·579

*Unbiased estimate (Nei 1978).

ferentiation among swamp populations was much greater than that among river populations (Table 4). For the swamp populations, estimates of θ for each locus all show significant population differentiation. None of the within-population inbreeding estimates (*f*) for each locus are significantly different from zero, except for CSSM046, but the mean *f* estimate over all the loci is significantly greater than zero. When CSSM046 is excluded, the mean estimate of *f* is reduced to 0·024, but is still significant ($P = 0·024$). In the river populations, none of the *f* estimates is significantly greater than zero, and the mean is barely significant ($P = 0·048$). In the river populations, only six loci found only in the river populations had θ estimates significantly greater than zero (CSSM008, CSSM045, CSSM046, CSSM047, CSSM060 and CSSM061).

For the swamp populations, the regression of Slatkin's log $\hat{M}$ on log (geographic distance) was negative but not significant. Although isolation by distance was not detected, the estimated $\hat{M}$ values (Table 5) show substantial variation (average = 1·96 ± 1·59). Average $\hat{M}$ among the three river populations was 6·42, and the average for all swamp–river pairs of populations was 0·51.

Genetic distances

All correlation coefficients among the four measures of genetic distance that were computed were highly significant ($P < 0·001$), with the $(\delta\mu)^2$ distance showing the lowest correlation with other measures. Correlations between microsatellite based and protein-coding loci-based distances were lower than those among different distance measures within each type of data, but again were highly significant ($P < 0·001$).

Takezaki & Nei (1996) showed that the D_A distance is generally best for inferring the correct topology for both the infinite allele model (protein-coding loci) and for the stepwise mutation

model (microsatellites). Therefore, to allow direct comparison with results from protein-coding loci (Barker *et al.* 1996), the D_A distance has been used for detailed presentation of genetic distances (Table 6), and a dendrogram of genetic relationships among the populations (Fig. 3A) has been constructed as a neighbour-joining tree (Saitou & Nei 1987). Most nodes are strongly supported, and the overall topology of the swamp populations is consistent with their geography. The low bootstrap value (34%) and short branch length for the Bogor/Australia node is as expected, as the Australian population descends from some 80 animals imported from Timor (and possibly Java) between 1826 and 1843 (Letts 1962; Tulloch 1974). Neighbour-joining trees, using $(\delta\mu)^2$ and Reynold's distances showed some differences in topology from each other, and from the D_A distance tree, while the tree based on Nei's standard genetic distance (*D*) had the same topology as that using the Reynold's distance. All microsatellite-based trees have a clear separation of the swamp and river (including the Lankan buffalo) types, fully supporting the cytological and protein-coding loci differentiation.

Table 3. Number of microsatellite alleles observed and expected in swamp and river buffalo

	Swamp		River	
Locus	Mean no. observed per population	Mean no. expected per population†	Mean no. observed per population	Mean no. expected per population†
CSSM008	3·6	5·5*	4·7	6·8
CSSM013	3·4	5·7***	2·0	2·7
CSSM015	2·0	3·8**	2·0	3·4
CSSM019	6·5	8·3*	4·7	6·2
CSSM022	2·9	3·2	5·0	7·1
CSSM029	3·9	5·1**	3·3	4·5*
CSSM032	4·8	5·6*	4·7	4·1
CSSM033	4·8	6·1*	5·7	6·8
CSSM036	3·5	5·5**	5·7	7·0
CSSM038	3·6	5·0	6·3	6·0
CSSM041	3·6	5·5**	3·7	3·7
CSSM043	5·3	7·7***	6·3	7·4*
CSSM045	2·1	1·6	4·0	6·4
CSSM046	3·8	5·5*	3·7	4·6
CSSM047	5·0	7·7**	10·7	13·2**
CSSM057	5·9	7·9*	4·7	6·0
CSRM060	2·9	3·8	6·7	8·1
CSSM061	4·6	5·9*	7·3	8·0
CSSM062	3·3	4·7**	5·3	7·6*
BRN	3·8	5·3	3·7	3·3
HMH1R	1·8	1·5	2·0	1·6

†Significance of difference between observed and expected numbers, $*P<0{\cdot}05$, $**P<0{\cdot}01$, $***P<0{\cdot}001$.

However, the D_A distance tree derived from protein-coding loci (Fig. 3B) shows major differences in branch lengths and in the topology for the swamp populations, as compared with the D_A distance microsatellite tree (Fig. 3A). Mean genetic distances among swamp populations were 0·228 from microsatellites and 0·098 from protein-coding loci and, between the swamp and river types, 0·897 and 0·173, respectively (using Nei's standard genetic distance, as this measure has been used most commonly in previous allozyme studies of population differentiation and relationships).

Estimation of divergence times

Rates of mutation will vary at different microsatellite loci, and we have used the average for 28 human loci ($1{\cdot}2 \times 10^{-3}$) from Weber & Wong (1993) in calculating divergence times (Table 7). As noted earlier, historical evidence is available for the time of origin of the Australian population (1826–1843) and for its source population (Timor and possibly Java). Assuming a generation interval of 8–10 years, the Australian population separated 16–20 generations ago. If we take the Bogor (Java) population as representing the ancestral population, the estimate of this separation time is too long at 72 generations. However, the Australian population descends from only about 80 animals and, allowing for this bottleneck (Nei 1987; equation 9·64), the corrected *D* is 0·0602 and the estimated divergence time is 25 generations – in very good agreement with the actual time.

Discussion

Swamp and river buffalo are recognized as two types of a single species that are differentiated on morphological, behavioural and cytological criteria, and whose endemic distributions are parapatric. The swamp type exists throughout south-east Asia, from Assam and Nepal in the west to the Yangtze Valley of China, while the river type is native to the Indian subcontinent, but has spread west to the Balkans, Italy and Egypt within historical times (Cockrill 1974). Furthermore, the two types are interfertile. Although Garza *et al.* (1995) and Moore *et al.* (1995) have cautioned that alleles of the same size in different species may be non-identical by descent, we assume that the swamp and river types are sufficiently closely related that such non-identity by descent would be, at least, very unlikely. More alleles were detected in the swamp type for most loci (see Table 1; CSSM047 was exceptional, with twice as many alleles in

river buffalo as in swamp buffalo), and an average of 42% of the alleles detected were shared between the two types. As only 62 river-type animals in three populations were assayed, with two of these populations being the same breed, the numbers of alleles in the river type and the proportion of shared alleles are minimal estimates. Even so, marked differences in the allele frequency distributions (e.g. CSSM019; see Fig. 2) are common. For eight loci, the most frequent allele in one type is absent or at low frequency in the other type. Three further loci

Table 4. *F*-statistics analyses for (1) each of 21 microsatellite loci in eight swamp buffalo and three river buffalo populations, and (2) mean estimates for these populations

	Swamp			River		
	f	θ	*F*	*f*	θ	*F*
Locus[†]						
CSSM008	–0·034 (0·073)	0·251 (0·130)***	0·223 (0·134)**	0·033 (0·084)	0·117 (0·081)**	0·150 (0·140)
CSSM013	–0·011 (0·085)	0·033 (0·018)**	0·023 (0·098)	0·001 (0·098)	0·028 (0·033)	0·032 (0·132)
CSSM015	–0·040 (0·087)	0·133 (0·082)***	0·097 (0·098)	–0·192 (0·117)	0·078 (0·125)	–0·105 (0·150)
CSSM019	0·082 (0·046)	0·197 (0·080)***	0·262 (0·069)**	–0·055 (0·049)	–0·012 (0·016)	–0·069 (0·032)
CSSM022	0·162 (0·158)	0·189 (0·068)***	0·314 (0·104)**	–0·077 (0·132)	0·040 (0·032)	–0·035 (0·126)
CSSM029	–0·039 (0·062)	0·239 (0·066)***	0·208 (0·070)**	–0·294 (0·100)	–0·007 (0·006)	–0·301 (0·108)
CSSM032	–0·001 (0·073)	0·030 (0·011)***	0·029 (0·067)	–0·008 (0·051)	0·032 (0·040)	0·026 (0·089)
CSSM033	0·120 (0·078)	0·179 (0·086)***	0·273 (0·058)**	0·206 (0·035)	0·047 (0·006)	0·243 (0·035)**
CSSM036	–0·001 (0·068)	0·105 (0·034)***	0·104 (0·068)	0·095 (0·181)	0·012 (0·030)	0·111 (0·201)
CSSM038	–0·003 (0·069)	0·201 (0·069)***	0·199 (0·085)**	0·010 (0·009)	0·011 (0·028)	0·021 (0·028)
CSSM041	–0·037 (0·068)	0·061 (0·020)***	0·027 (0·075)	0·048 (0·101)	0·028 (0·060)	0·074 (0·118)
CSSM043	0·029 (0·054)	0·064 (0·023)***	0·090 (0·036)	0·177 (0·051)	–0·002 (0·007)	0·175 (0·047)
CSSM045	0·063 (0·057)	0·222 (0·138)***	0·281 (0·189)	0·177 (0·063)	0·078 (0·082)*	0·237 (0·042)
CSSM046	0·487 (0·069)**	0·283 (0·085)***	0·631 (0·059)**	0·016 (0·041)	0·146 (0·095)*	0·155 (0·052)
CSSM047	0·024 (0·049)	0·113 (0·042)***	0·133 (0·051)**	0·087 (0·062)	0·045 (0·021)*	0·129 (0·078)
CSSM057	–0·065 (0·043)	0·195 (0·055)***	0·141 (0·050)**	–0·089 (0·083)	0·008 (0·013)	–0·081 (0·071)
CSRM060	0·146 (0·119)	0·199 (0·041)***	0·313 (0·081)**	–0·083 (0·084)	0·060 (0·073)*	–0·012 (0·151)
CSSM061	0·073 (0·057)	0·217 (0·074)***	0·273 (0·077)**	0·041 (0·090)	0·071 (0·059)*	0·104 (0·042)
CSSM062	0·007 (0·044)	0·202 (0·081)***	0·206 (0·083)**	0·147 (0·084)	0·037 (0·045)	0·181 (0·106)
BRN	0·063 (0·049)	0·273 (0·083)***	0·319 (0·087)**	0·195 (0·147)	0·010 (0·029)	0·201 (0·135)
HMH1R	0·130 (0·189)	0·101 (0·075)***	0·204 (0·126)*	0·234 (0·149)	0·022 (0·026)	0·256 (0·173)
Mean estimates [b]	0·047 (0·027)***	0·168 (0·018)***	0·207 (0·034)***	0·031 (0·028)*	0·038 (0·008)***	0·068 (0·028)***

f, Within-population inbreeding estimate; *F*, total inbreeding estimate; θ, measure of population differentiation.
†Standard deviations in parentheses – estimate from jackknife over populations.
‡Standard deviations in parentheses – estimate from jackknife over loci.
$*P<0·05$, $**P<0·01$, $***P<0·001$, from permutation tests in FSTAT program.

Table 5. Gene flow between pairs of swamp populations (Slatkin's (1993) $\hat{M}^* = Nm^\dagger$) estimated from microsatellite data (below the diagonal) and from 18 polymorphic protein-coding loci (above the diagonal)

	Surin	Trengganu	Sabah	Sarawak	Bogor	Sulawesi	Philippines	Australia
Surin	–	6·64	1·27	3·15	3·02	6·32	2·24	1·21
Trengganu	6·39	–	1·17	1·75	2·27	3·10	1·41	0·70
Sabah	0·93	0·79	–	0·87	0·56	0·79	0·65	0·38
Sarawak	1·00	0·81	5·10	–	0·97	1·22	0·99	0·83
Bogor	3·54	2·72	0·80	0·80	–	9·59	1·06	0·65
Sulawesi	2·68	2·27	0·65	0·69	5·64	–	1·75	0·94
Philippines	3·19	1·97	2·41	2·60	1·93	1·52	–	0·57
Australia	1·07	0·89	0·44	0·45	1·41	1·35	0·79	–

*$\hat{M}$, genetically effective migration rate.
†*N*m, number of migrants exchanged per generation.

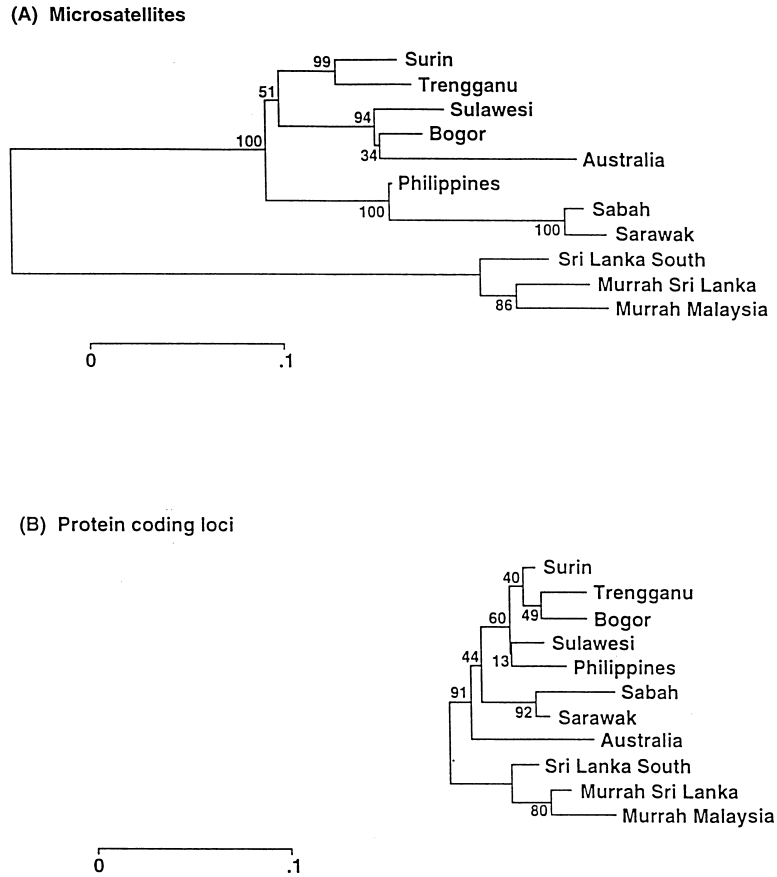

Fig. 3. Dendrograms of relationships among 11 water buffalo populations, using D_A genetic distances and the neighbour-joining method of clustering, which were based on (A) 21 polymorphic microsatellite loci, and (B) 25 polymorphic protein-coding loci. Numbers on the nodes are percentage bootstrap values from 1000 replications of resampled loci, and a scale bar for branch lengths is shown.

Table 6. Matrix of D_A genetic distances among 11 water buffalo populations

	Swamp							River		
	Trengganu	Sulawesi	Bogor	Australia	Philippines	Sabah	Sarawak	Sri Lanka South	Murrah SL*	Murrah M†
Surin	0·075	0·153	0·138	0·242	0·112	0·244	0·262	0·517	0·516	0·540
Trengganu		0·139	0·148	0·243	0·149	0·259	0·283	0·520	0·540	0·566
Sulawesi			0·061	0·148	0·178	0·278	0·285	0·545	0·580	0·565
Bogor				0·129	0·150	0·229	0·249	0·553	0·586	0·576
Australia					0·245	0·330	0·342	0·585	0·582	0·570
Philippines						0·111	0·112	0·511	0·515	0·543
Sabah							0·034	0·601	0·641	0·665
Sarawak								0·608	0·646	0·672
Sri Lanka South									0·102	0·101
Murrah SL										0·088

*Murrah – Sri Lanka.
†Murrah – Malaysia.

show the most common allele or alleles in the river type as rare in the swamp, and present only in the Philippines population. One animal was heterozygous for a 'river' allele at two loci, and four others were heterozygous for a 'river' allele at one locus. These five animals are certainly not F_1 crossbreds, as all animals sampled were clearly swamp type. However, river buffalo were first imported into the Philippines in 1917 (Villegas 1958) and, as crossbreeding has been

Table 7. Estimated divergence times among buffalo types and populations

Types/population	*D**	Divergence times (generations)
Swamp vs. Murrah	0·9182	383
Swamp vs. Lankan	0·8532	356
Among swamp		
Maximum *D*	0·4918	205
Average *D*	0·2276	95
Australia/Bogor	0·1717	72
Corrected *D*	0·0602	25

**D*, standard genetic distance, (Nei 1978).

practised extensively, these five animals must have some crossbred ancestry.

As expected, the microsatellite loci show very high levels of genetic diversity, with an average within-population gene diversity of 0·514 (per locus values ranging from 0·104–0·711), and expected heterozygosity for each population ranging from 0·380–0·615. These estimates are all approximately twice those derived from assay of 25 polymorphic protein-coding loci in the same 11 populations, which show an average within-population gene diversity of 0·287 (range 0·033–0·482), and expected heterozygosity for each population ranging from 0·184–0·346. Only six of the microsatellite loci had average within-population gene diversities less than the maximum estimate observed for the protein-coding loci.

There was a significant positive relationship between average within-population gene diversity for each locus and the number of alleles detected (regression coefficient = 0·021 ± 0·006, $P < 0·01$). However, the number of alleles detected increases with sample size, and the sample size for the river type was much less than that for swamp type (62 vs. 174). The average number of alleles per locus was less for the river type, and the proportion of shared alleles was only 0·42 (Table 1), so that averaging results over both types may bias the relationship between gene diversity and number of alleles. The regression of gene diversity on number of alleles therefore was calculated for the swamp populations only, resulting in a stronger relationship than for the pooled data (regression coefficient = 0·032 ± 0·010, $P < 0·01$). This positive relationship is expected under both the stepwise mutation and infinite allele models (Shriver *et al.* 1993). Our limited testing of the expected numbers of alleles under both models suggested a closer fit to the former model, but we cannot exclude the hypothesis of Shriver *et al.* (1993) that the mutational mechanism of microsatellite loci is close to, but not exactly, the one-step stepwise mutation model.

Hardy–Weinberg

Eight significant departures from Hardy–Weinberg were detected but, with 231 locus–population combinations, this number could be expected by chance. However, locus CSSM046 appears exceptional, with significant deviations in four populations, an observed deficiency of heterozygotes in all but one population and a high and significant *f* estimate in the swamp populations (Table 4, 0·487, $P < 0·01$). The three other loci (CSSM022, CSSM033 and CSSM060), for which global tests of departures from Hardy–Weinberg equilibrium were significant, also had observed deficiencies of heterozygotes in the majority of populations (seven, eight and six, respectively), with some high but not significant *f* estimates (Table 4). For these four loci then, it is possible that non-amplifying alleles are segregating in at least some populations (Pemberton *et al.* 1995). However, as noted earlier, samples from a number of animals failed to amplify for up to 10 loci. As only 12 samples failed to amplify for CSSM046, while, for the 17 loci other than the above four, an average of 8–9 samples failed to amplify, it was not feasible to specify any sample as homozygous for a non-amplifying allele. Given this possibility of non-amplifying alleles, however, the significant mean *f* estimates for both the swamp and river types (Table 4) are likely to be overestimates. Therefore, the microsatellite data show little or no evidence of inbreeding in these populations, in agreement with results for the protein-coding loci, where mean *f* estimates for swamp and river populations did not differ significantly from zero. By contrast to the microsatellites, the protein-coding loci showed many significant deviations from Hardy–Weinberg expectations (Barker *et al.* 1996). However, these deviations for each locus across populations were consistently either positive or negative, indicating locus-specific effects that suggest selection affecting some of these loci (Barker *et al.* 1996).

Population differentiation

All microsatellite loci show significant differentiation among the swamp populations, but only six loci are significant for the river populations and the mean θ estimate for the river populations is only about one-quarter of that for the

swamp (Table 4). However, with only three river populations, two of which are different populations of the one breed, the estimated θ is probably an underestimate of between-population diversity in the river type. The mean θ estimate for 12 swamp populations, based on 25 protein-coding loci, was 0·157 ± 0·034 (Barker *et al.* 1996) while, for the eight populations considered here, mean θ was 0·182 ± 0·041. Therefore, the differentiation among these eight swamp populations was greater when estimated from protein-coding loci than from microsatellites (0·168 ± 0·018), but the former also showed much greater variation, among loci, in estimates of θ. For microsatellites, the highest θ estimate was 0·283 for CSSM046, but five protein-coding loci had higher estimates with a maximum of 0·939 for catalase.

Detecting natural selection

Comparisons of geographic variation in allele frequencies for different classes of polymorphic loci is a powerful method for detecting effects of natural selection (McDonald 1994). As the effects of drift and migration are the same for all alleles at all neutral polymorphic loci, all such alleles have the same expected F_{ST}, and there should be no significant heterogeneity of F_{ST} estimates. Lewontin & Krakauer (1973) derived the sampling distribution of F_{ST}, and proposed a test of heterogeneity by a comparison of the observed variance with the theoretical expected variance. Problems with estimating the expected variance (Robertson 1975) preclude general application of this test, but if we assume the microsatellite loci to be neutral to selection, the observed variance of F_{ST} estimates from these loci can be used instead of an expected theoretical variance.

An alternative modified Lewontin–Krakauer test, used by Pogson *et al.* (1995), was to compare the mean F_{ST} estimates from two classes of loci, calculating:

$$\chi^2_{(n-1)} = (n\text{-}1)(F_{ST}(\text{RFLPs})/F_{ST}\,(\text{allozymes})$$

where n = number of populations and the F_{ST} for each of the two classes are weighted means.

In fact, both tests may be necessary, as the mean F_{ST} estimates may not differ, while variances do, or vice versa. McDonald (1994) has pointed out that comparing F_{ST} estimates for two classes of loci that differ in the number of alleles could be quite misleading and recommended treating all polymorphisms as two-allele polymorphisms, by using the frequencies of the overall most common allele and pooling the others. Therefore, for the eight swamp populations, unweighted F_{ST} (as computed by BIOSYS-1) were estimated from the collapsed two-allele data for each of the microsatellite and protein-coding loci. Mean F_{ST} and variance were 0·1714 and 0·011043 for microsatellites and 0·1481 and 0·023780 for protein-coding loci. Applying the modified Lewontin–Krakauer tests (as above), the means are not significantly different, but the variances are ($\chi^2_{(7)} = 15{\cdot}074$, $P < 0{\cdot}05$). Thus, the 25 protein-coding loci are more heterogeneous than expected from the variation shown by the microsatellite loci, indicating (assuming the microsatellite loci are neutral) selection affecting some of the protein-coding loci. A general test for this selection is to treat the ratio of the F_{ST} for each protein-coding locus to the mean F_{ST} for microsatellites as a variance ratio F-test, and to perform a two-tailed test of significance (d.f. = 7,7). On this test, six protein-coding loci show a significantly lower F_{ST} estimate [Alanine aminotransferase (Aat), Alkaline phosphatase (Alp), NADPH diaphorase (Dia-2), Glucose 6-phosphate dehydrogenase (G6pd), Leucine aminopeptidase (Lap), Purine nucleoside phosphorylase (Np)], suggesting some form of balancing selection affecting genetic variation at these loci.

Gene flow and population history

Estimates of gene flow between the swamp populations (Table 5) show reasonable agreement between those derived from microsatellite and protein-coding loci (correlation coefficient = 0·71, $P < 0{\cdot}01$) although, because of possible selection effects on the latter, the microsatellite estimates would be preferred. Nevertheless, even the microsatellite-based estimates should be interpreted with caution, as these populations are unlikely to be in equilibrium and the estimates reflect both past and current gene flow. The swamp buffalo was domesticated in China about 7000 years ago (Chen & Li 1989), probably following the development of rice cultivation in the lower Yangtze region about 8000 years ago (Bellwood 1992). The spread of this agricultural system, and with it presumably the swamp buffalo, from southern China to mainland south-east Asia dates from some 5000 years ago, and to Sumatra and Java some 3000–4000 years ago. Taking Surin as the nearest (geographically) to the ancestral population, the high $\hat{M}$ estimates for Surin/Philippines and Philippines/Sabah, Philippines/Sarawak suggest that the buffalo may also have spread from China to the Philippines and then to Sabah and Sarawak. The highest estimates probably represent current gene flow between geographically-near populations (Surin/Trengganu and Sabah/Sarawak) or between populations in the same

country (Bogor/Sulawesi). The generally low estimates for Australia with other populations, but highest with Bogor and Sulawesi, are in accordance with its known history.

Although the estimated divergence time between Australia and Bogor, using the corrected Nei distance (Table 7), accords well with its known history, this is probably fortuitous as the estimation method is not likely to be accurate over such short periods. Taking the average Nei distance among the swamp populations and an average divergence time of 3000–4000 years (say 300–400 generations), the mutation rate would be $3 \times 10^{-4} - 4 \times 10^{-4}$. If this mutation rate range is then used, the swamp–river divergence is estimated as 1000–1500 generations or 10 000–15 000 years ago, possibly still a minimum estimate.

Phylogenetic relationships

A primary motivation for this study was the comparison of between-population diversity and phylogenetic relationships estimated using different classes of markers, namely protein-coding and microsatellite loci. Scribner *et al.* (1994) compared allozymes and three classes of variable number of tandem repeat (VNTR) markers (single-locus microsatellite, minisatellite and multilocus minisatellite) in *Bufo bufo* populations. The F_{ST} estimate from allozymes was 3–6 times greater than from the VNTR loci, but genetic distances were concordant across all four marker classes. However, only three populations, six polymorphic allozyme loci, three single-locus minisatellites, one multilocus minisatellite and one microsatellite were analysed. By contrast, Pogson *et al.* (1995), comparing 10 allozyme and 17 nuclear RFLP loci in *Gadus morhua*, found significantly higher mean F_{ST} and much larger genetic distances for RFLPs than for allozymes. However, the topology of the phylogenetic relationships among the populations was very similar for the two marker classes. As noted previously, we found no difference in mean F_{ST} estimates for microsatellites and allozymes, yet genetic distances among populations were much greater when estimated from microsatellites (Fig. 3).

More significantly, there were marked differences in the topologies of the swamp population trees derived from the two classes of markers. The topology of the microsatellite tree for the swamp populations is consistent with their geography and their presumed spread through south-east Asia and it is concordant with the gene flow estimates. By contrast, the protein-coding tree separates Australia and Sabah/Sarawak from the other five populations. These differences between the two trees could be quite enigmatic, but we are fortunate in that the history of the Australian population is known, and the protein-coding loci show a strong bottleneck effect owing to a small effective size of the founding population (Barker *et al.* 1996) and a consequent increase in its genetic distance from ancestral populations (Chakraborty & Nei 1977). For microsatellites, the bottleneck effect is apparent in the lower mean number of alleles per locus (Table 2), but as this is primarily loss of low frequency alleles, there is less effect on heterozygosity and on genetic distance estimates. While the protein-coding loci did not show a clear bottleneck effect for Sabah and Sarawak, this is apparent from the microsatellite results, with mean number of alleles per locus and expected heterozygosity even less than for Australia (Table 2). Whereas these bottleneck effects show as long branch lengths for Australia and for Sabah/Sarawak in the microsatellite tree, they completely distort the protein-coding tree. As many domestic livestock breeds are possibly descended from small founding populations, microsatellite-based trees are clearly to be preferred in assessing genetic relationships among breeds. Furthermore, in natural populations where founding sizes will not be known, these results should serve as a caution in interpreting population structure from genetic distance estimates based on allozymes or other markers with relatively low levels of heterozygosity.

Acknowledgements

We gratefully acknowledge Associate Professor K Yusoff for provision of laboratory facilities and Mr C H Lau for technical assistance with DNA extraction; the co-operation and assistance during sample collection of Professor C Chantalakhana, Ms Ancharlee and the Thai Department of Livestock Development; Professor R L Anacleto, Central Mindanao University of the Philippines; Dr P Abeynayake, University of Peradeniya, Sri Lanka; Dr Sudarisman and Dr P Daniels, BALITVET, Indonesia; Dr S P Chia, Department of Animal Services, Sabah; Dr K Y Hsiung, Department of Agriculture, Sarawak; Dr M Hilmi, Universiti Pertanian Malaysia; Mr O S Selvaraj, University of Malaya; and Ms J Purdie, Department of Primary Industry and Fisheries, N T, Australia. This work was supported by the Australian Centre for International Agricultural Research, the Malaysian Program for Intensification of Research in Priority Areas (Project no. 1–07–05–057), and Universiti Pertanian Malaysia.

References

Avise J.C. (1994) *Molecular Markers, Natural History and Evolution*. Chapman and Hall, New York.

Barker J.S.F. (1994) Animal breeding and conservation genetics. In: *Conservation Genetics* (ed. by V. Loeschcke, J. Tomiuk & S.K. Jain), pp. 381–95. Birkhäuser Verlag, Basel.

Barker J.S.F., Tan S.G., Selvaraj O.S. & Mukherjee T.K. (1997) Genetic variation within and relationships among populations of Asian water buffalo (*Bubalus bubalis*). *Animal Genetics* **28,** 1–13.

Beckmann J.S. & Weber J.L. (1992) Survey of human and rat microsatellites. *Genomics* **12,** 627–31.

Bellwood P. (1992) Southeast Asia before history. In: *The Cambridge History of south-east Asia*, Vol. 1 (ed. by N. Tarling), pp. 55–136. Cambridge University Press, Cambridge.

Bongso T.A., Kumaratileke W.L.J.S. & Buvanendran V. (1977) The karyotype of the indigenous buffalo (*Bubalus bubalis*) of Sri Lanka. *Ceylon Veterinary Journal* **25,** 9–11.

Bruford M.W. & Wayne R.K. (1993) Microsatellites and their application to population genetic studies. *Current Opinion in Genetics and Development* **3,** 939–43.

Chakraborty R. & Nei M. (1977) Bottleneck effects on average heterozygosity and genetic distance with the stepwise mutation model. *Evolution* **31,** 347–56.

Chen Y.C. & Li X.H. (1989) New evidence of the origin and domestication of the Chinese swamp buffalo (*Bubalus bubalis*). *Buffalo Journal* **1,** 51–5.

Cockrill, W.R. (ed.) (1974) *The Husbandry and Health of the Domestic Buffalo*. FAO, Rome.

Ely J., Deka R., Chakraborty R. & Ferrell R.E. (1992) Comparison of five tandem repeat loci between humans and chimpanzees. *Genomics* **14,** 692–8.

Ewens W.J. (1972) The sampling theory of selectively neutral alleles. *Theoretical Population Biology* **3,** 87–112.

Felsenstein J. (1993) PHYLIP (Phylogeny Inference Package) version 3·5c. Distributed by the author. Department of Genetics, University of Washington, Seattle.

Garza J.C., Slatkin M. & Freimer N.B. (1995) Microsatellite allele frequencies in humans and chimpanzees, with implications for constraints on allele size. *Molecular Biology and Evolution* **12,** 594–603.

Goldstein D.B., Linares A.R., Cavalli-Sforza L.L. & Feldman M.W. (1995) Genetic absolute dating based on microsatellites and the origin of modern humans. *Proceedings of the National Academy of Science of the USA* **92,** 6723–7.

Goudet J. (1995) FSTAT (Version 1·2): A computer program to calculate F-statistics. *Journal of Heredity* **86,** 485–6.

Hillis D.M. & Moritz C (eds) (1990) *Molecular Systematics*. Sinauer, Sunderland, MA.

Hochberg Y. (1988) A sharper Bonferroni procedure for multiple tests of significance. *Biometrika* **75,** 800–2.

Kidd K.K. (1974) Biochemical polymorphisms, breed relationships, and germ plasm resources in domestic cattle. *Proceedings of the 1st World Congress on Genetics Applied to Livestock Production* **1,** 321–8.

Lessios H.A. (1992) Testing electrophoretic data for agreement with Hardy–Weinberg expectations. *Marine Biology* **112,** 517–23.

Letts G.A. (1962) Early livestock introductions to the 'top-end' of the Northern Territory. *Australian Veterinary Journal* **38,** 282–7.

Lewontin R.C. & Krakauer J. (1973) Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics* **74,** 175–95.

MacHugh D.E., Loftus R.T., Bradley D.G., Sharp P.M. & Cunningham P. (1994) Microsatellite DNA variation within and among European cattle breeds. *Proceedings of the Royal Society, London B* **256,** 25–31.

McDonald J.H. (1994) Detecting natural selection by comparing geographic variation in protein and DNA polymorphisms. In: *Non-neutral Evolution Theories and Molecular Data* (ed. by B. Golding), pp. 88–100. Chapman and Hall, New York.

Mantel N. (1967) The detection of disease clustering and a generalized regression approach. *Cancer Research* **27,** 209–20.

Minch E., Ruiz-Linares A., Goldstein D.B., Feldman M.W. & Cavalli-Sforza L.L. (1995) MICROSAT (version 1·4): a computer program for calculating various statistics on microsatellite allele data. WWW: http://lotka. stanford. edu/research/distance. html.

Moore S.S., Evans D., Byrne K. *et al.* (1995) A set of polymorphic DNA microsatellites useful in swamp and river buffalo (*Bubalus bubalis*). *Animal Genetics* **26,** 355–9.

Nei M. (1973) Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Science of the USA* **70,** 3321–3.

Nei M. (1976) Mathematical models of speciation and genetic distance. In: *Population Genetics and Ecology* (ed. by S. Karlin & E. Nevo), pp. 723–65. Academic Press, New York.

Nei M. (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* **89,** 583–90.

Nei M. (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Nei M., Tajima F. & Tateno Y. (1983) Accuracy of estimated phylogenetic trees from molecular data II. Gene frequency data. *Journal of Molecular Evolution* **19,** 153–70.

Pemberton J.M., Slate J., Bancroft D.R. & Barrett J.A. (1995) Nonamplifying alleles at microsatellite loci: a caution for parentage and population studies. *Molecular Ecology* **4,** 249–52.

Pogson G.H., Mesa K.A. & Boutilier R.G. (1995) Genetic population structure and gene flow in the Atlantic cod *Gadus morhua*: A comparison of allozyme and nuclear RFLP loci. *Genetics* **139,** 375–85.

Raymond M. & Rousset F. (1995) GENEPOP (Version 1·2): Population genetics software for exact tests and ecumenicism. *Journal of Heredity* **86,** 248–9.

Robertson A. (1975) Gene frequency distributions as a test of selective neutrality. *Genetics* **81,** 775–85.

Roy M.S., Geffen E., Smith D., Ostrander E.A. & Wayne R.K. (1994) Patterns of differentiation and hybridization in North American wolflike canids, revealed by analysis of microsatellite loci. *Molecular Biology and Evolution* **11,** 553–70.

Rendel J. (1967) Studies of blood groups and protein variants as a means of revealing similarities and differences between animal populations. *Animal Breeding Abstracts* **35,** 371–83.

Reynolds J., Weir B.S. & Cockerham C.C. (1983) Estimation of the coancestry coefficient: basis for a short term genetic distance. *Genetics* **105,** 767–79.

Saitou N. & Nei M. (1987) The neighbour-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4,** 406–25.

Scheurmann E., Wiesner R., Fischer H. & Jainudeen M.R. (1974) Karyotyp, C-Banden und Identifizierung der Geschlechtschromosomen des Ceylonesischen wasserbüffels (*Bubalus bubalis*). *Giessener Beiträge zur Erbpathologie und Zuchthygiene* **6,** 1–7.

Scribner K.T., Arntzen J.W. & Burke T. (1994) Comparative analysis of intra- and interpopulation genetic diversity in *Bufo bufo*, using allozyme, single-locus microsatellite, minisatellite, and multilocus minisatellite data. *Molecular Biology and Evolution* **11,** 737–48.

Shriver M.D., Jin L., Chakraborty R. & Boerwinkle E. (1993) VNTR allele frequency distributions under the stepwise mutation model: a computer simulation approach. *Genetics* **134,** 983–93.

Slatkin M. (1993) Isolation by distance in equilibrium and non-equilibrium populations. *Evolution* **47,** 264–79.

Slatkin M. (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics* **139,** 457–62.

Swofford D.L. & Selander R.B. (1989) BIOSYS–1 A computer program for the analysis of allelic variation in population genetics and biochemical systematics. Release 1·7, Illinois Natural History Survey, Champaign IL. pp. 43.

Takezaki N. & Nei M. (1996) Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. *Genetics* **144,** 389–99.

Tulloch D.G. (1974) The feral swamp buffaloes of Australia's Northern Territory. In: *The Husbandry and Health of the Domestic Buffalo* (ed. by W.R. Cockrill), pp. 493–505. FAO, Rome.

Villegas V. (1958) The Indian buffalo. *Philippine Journal of Animal Industry* **19,** 85–90.

Ward R.D., Skibinski D.O.F. & Woodwark M. (1992) Protein heterozygosity, protein structure, and taxonomic differentiation. *Evolutionary Biology* **26,** 73–159.

Weber J.L. & Wong C. (1993) Mutation of human short tandem repeats. *Human Molecular Genetics* **2,** 1123–8.

Weir B.S. & Cockerham C.C. (1984) Estimating *F*-statistics for the analysis of population structure. *Evolution* **38,** 1358–70.

van Zeveren A., Peelman L., van de Weghe A. & Bouquet Y. (1995) A genetic study of four Belgian pig populations by means of seven microsatellite loci. *Journal of Animal Breeding and Genetics* **112,** 191–204.