Genetic variability of the marine mussel *Mytilus galloprovincialis* assessed using two-dimensional electrophoresis

E Mosquera, JL López and G Alvarez

Departamento de Biología Fundamental, Facultad de Biología, Universidad de Santiago de Compostela, 15782 Santiago de Compostela, Spain

Two-dimensional electrophoresis (2-DE) has been used to measure the degree of genetic variability of the marine mussel *Mytilus galloprovincialis*. Genetic polymorphisms were detected in 33 of a total of 86 polypeptides scored among the most abundant proteins from foot samples in 38 individuals. Estimates of average heterozygosity were 0.101 ± 0.018 and 0.114 ± 0.021 in a natural and a cultured population, respectively, from the NW of the Iberian Peninsula. These are the highest estimates of average heterozygosity reported by 2-DE in an animal species to date. We consider that these data throw open the question of

the level of genetic variability detectable by two-dimensional electrophoresis. Multilocus genotype data were used to infer haplotypic frequencies by means of the EM algorithm in order to detect linkage disequilibrium between loci coding abundant proteins. Significant associations were found in 22.7% of the 406 two-locus pairs analysed. Also, clusters of loci in which all pairwise combinations exhibit statistically significant associations were detected and physical linkage between some of these loci is postulated from the linkage disequilibrium data.

Heredity (2003) **90**, 432–442. doi:10.1038/sj.hdy.6800266

Keywords: genetic variability; Hardy–Weinberg departures; linkage disequilibrium; *Mytilus galloprovincialis*; protein polymorphisms; two-dimensional electrophoresis

Introduction

Two-dimensional electrophoresis (2-DE) is a powerful technique for separating complex mixtures of denatured proteins according to two independent criteria: charge and molecular weight (O'Farrell, 1975). Combined with nonspecific protein staining, the technique permits the visualization on a single gel of a very large number of gene products that represent the more abundant proteins in a cell or tissue. Unlike one-dimensional electrophoresis (1-DE), which is restricted to the analysis of native soluble proteins, mainly enzymes, 2-DE allows the examination of a broad spectrum of proteins and, consequently, a substantially increased number of protein-encoding loci. This suggests that 2-DE could have a great potential for the study of genetic variability of populations, in that it allows a more representative sample of the genome to be analysed. However, studies of genetic variability in natural populations of animal species by means of 2-DE have been relatively scarce, because 2-DE is technically more difficult and time consuming than 1-DE and, furthermore, because the first results revealed substantially less genetic variation than had been estimated by 1-DE (Edwards and Hopkinson, 1980; Aquadro and Avise, 1981; Neel, 1990). Moreover,

most studies have been focused on a few species (particularly man and *Drosophila*) so that the available information is to a large extent redundant and biased as a means of getting an appropriate view of the levels of genetic variability detected by 2-DE. In man, 2-DE estimates of average heterozygosity (H) range from 0.000 to 0.040 for different tissues when more than 30 loci are scored (Walton *et al.*, 1979; Smith *et al.*, 1980; Hamaguchi *et al.*, 1981; Comings, 1982; Goldman and Merrill, 1983; Rosenblum *et al.*, 1983; Hanash *et al.*, 1986a,b; Takahashi *et al.*, 1986). Other studies in man yield higher values of heterozygosity (0.045–0.080) but they are based on less than 30 loci (Rosenblum *et al.*, 1984; Asakawa *et al.*, 1985). Apart from man, only two other species of mammals have been studied by 2-DE with respect to genetic variability: cheetah ($H = 0.013$, O'Brien *et al.*, 1983) and mouse ($H = 0.020$, Racine and Langley, 1980). In invertebrates, 2-DE estimates of genetic variability are restricted to two *Drosophila* species, in which estimated heterozygosities are rather similar to those reported for mammals (0.040 and 0.018 for whole body and male reproductive tract, respectively, in *D. melanogaster* and 0.000 and 0.028 for whole body and male reproductive tract, respectively, in *D. simulans*) (Leigh Brown and Langley, 1979; Ohnishi *et al.*, 1982; Coulthart and Singh, 1988). Therefore, additional estimates of genetic variability by 2-DE for different species of vertebrates and invertebrates are needed. In this article, we have applied 2-DE to determine the level of genetic variation for the most abundant proteins of the marine mussel *Mytilus*

Correspondence: G Alvarez, Departamento de Biología Fundamental, Facultad de Biología, Universidad de Santiago de Compostela, 15782 Santiago de Compostela, Spain. E-mail: bfjurado@usc.es
Received 28 March 2002; accepted 10 January 2003

galloprovincialis, remembering that molluscs are one of the taxonomic groups within the animal kingdom where the highest levels of genetic variability have been reported (the mean value of allozyme heterozygosity is 0.145 ± 0.010 for 105 molluscan species reviewed by Ward *et al*, 1992). In fact, *M. galloprovincialis* is a mollusc species where a particularly high value of heterozygosity for enzyme loci has been found ($H = 0.240$, Grant and Cherry, 1985).

The degree of nonrandom associations, or linkage disequilibrium, between loci coding abundant proteins in natural populations is unknown. Since a large number of gene products are revealed in a single 2-DE gel, the multilocus genotype array of each individual can be directly inferred. From these 2-DE multilocus genotype data, maximum likelihood estimates of haplotype frequencies can be obtained and, therefore, associations between genetic polymorphisms can be detected in a random mating population (Excoffier and Slatkin, 1995; Slatkin and Excoffier, 1996; Weir, 1996). Furthermore, the extent of linkage disequilibrium between gene markers produced as a result of various evolutionary forces such as mutation, drift, selection or admixture during the evolutionary history of a population, can be used indirectly to infer how strongly these markers are linked on the same chromosome. If the linkage disequilibrium of the markers was created a long time ago, a strong linkage disequilibrium detected now may suggest close physical linkage between the markers, because linkage disequilibria decay with time. This principle is the basis of linkage disequilibrium mapping, an approach with a great potential to obtain information on physical linkage, which is extensively used with molecular markers (Jorde, 1995; Kaplan *et al*, 1995; Weir, 1996; Wu and Zeng, 2001). In this article, the first attempt to detect linkage disequilibrium between loci coding abundant proteins as detected by 2-DE is attempted in the marine mussel *M. galloprovincialis*.

Materials and methods

Mussel sampling and two-dimensional gel electrophoresis

Adult mussels were sampled from two *M. galloprovincialis* populations in the NW of the Iberian Peninsula in November 1998: a midintertidal population from a wave-exposed rocky shore in Ribadeo and a raft-cultured population from an inner cultivation area in the Ría de Arosa (Vilagarcía, Figure 1). Mussels were brought alive to the laboratory, where they were dissected. The foot was removed from each individual and frozen, lyophilized and stored at -80°C . Proteins were extracted by suspending 30 mg of lyophilized tissue in 1 ml of O'Farrell lysis buffer (9.5 M urea, 2% NP-40 (v/w), 2% ampholytes pH 7–9 and 100 mM dithiothreitol (DTT)). The mixture was shaken for 4 h at 30°C and centrifuged at $12000 \times g$ for 30 min.

2-DE of mussel foot was performed as described by López *et al* (2001). Protein loads of 127 and 255 μg (in order to study the less abundant proteins) were applied to each gel. In this way, at least two gels were run for each of the individuals analysed. Isoelectric focusing using carrier ampholytes was carried out at 200 V for 2 h, 500 V for 2 h and 800 V for 16 h. Second dimension electrophoresis was carried out at 25 mA per gel for 1.5 h followed by 35 mA per gel for 5 h. 2-DE gels were stained by the silver technique of Morrissey (1981).

Data analysis

Genotype frequencies were determined by direct count on the gels. Goodness-of-fit of genotype frequencies to Hardy–Weinberg proportions at each locus was tested by means of the χ^2 and the exact test. The exact test is very useful when the sample size and/or some genotype frequencies are small because it does not rest on asymptotic approximations for large samples (Louis

Figure 1 Map of the NW coast of the Iberian Peninsula showing the sampling sites of *M. galloprovincialis*.

and Dempster, 1987; Guo and Thompson, 1992). Departures from Hardy–Weinberg proportions were measured by two F_{IS} statistics: the f_C of Weir and Cockerham (1984) and the f_T of Robertson and Hill (1984). The ratio of the estimate squared of f_T to its variance was used as a test for $F_{IS} = 0$ with one degree of freedom, which is optimal for detecting Hardy–Weinberg deviations in terms of statistical power (Robertson and Hill, 1984; Rousset and Raymond, 1995). Unbiased estimates of expected heterozygosity for single loci (h) and the population average heterozygosity (H), calculated as the average of h over all loci, were computed according to Nei (1978). The degree of genetic variation among populations was measured by means of the coefficient of gene differentiation (G_{ST}) as defined by Nei (1973, 1986).

Haplotypic frequencies for locus pairs were estimated from genotypic data using the expectation-maximization (EM) algorithm (Dempster *et al*, 1977; Excoffier and Slatkin, 1995). Computation of linkage or gametic disequilibrium from haplotypic frequency estimates was accomplished taking into account that many of the protein loci analysed are multiallelic (Weir and Cockerham, 1978; Weir, 1996). In this way, a separate disequilibrium coefficient (D_{ij}) is defined for each pair of alleles, A_i and B_j at loci A and B , respectively, as $D_{ij} = X_{ij} - p_i q_j$, where X_{ij} is the frequency of gamete $A_i B_j$ and p_i and q_j are the corresponding allele frequencies. A normalized measure of the extent of gametic disequilibrium for each pair of alleles is computed as $D'_{ij} = D_{ij}/D_{\max}$ (Lewontin, 1964), where $D_{\max} = \min[p_i q_j, (1 - p_i)(1 - q_j)]$ when $D_{ij} < 0$ or $D_{\max} = \min[p_i(1 - q_j), (1 - p_i)q_j]$, when $D_{ij} > 0$. The extent of overall disequilibrium between all the alleles at two loci is measured by $D' = \sum_{i=1}^k \sum_{j=1}^l p_i q_j |D'_{ij}|$, where k and l are the number of alleles in loci A and B , respectively (Hedrick, 1987; Zapata *et al*, 2001). The statistical significance of gametic disequilibrium for a particular allelic pair is performed by the χ^2 statistic $\chi^2_{ij} = 2ND_{ij}^2/p_i(1 - p_i)q_j(1 - q_j)$ with one degree of freedom, where N is the number of sampled individuals (Weir and Cockerham, 1978; Weir, 1996). The overall hypothesis that none of the D_{ij} is different from zero was tested by means of χ^2 tests (χ^2_{ij}) for each one of the allelic pairs with Bonferroni correction for multiple tests (Zapata *et al*, 2001). Also, a likelihood ratio test was computed as an overall test for gametic disequilibrium between loci (Slatkin and Excoffier, 1995).

Analysis of departures from Hardy–Weinberg proportions (exact test and F_{IS} statistics) was performed using GENET-2 and GENEPOP v. 3.1 programs (Quesada *et al*, 1992; Raymond and Rousset, 1995). Estimation of haplotypic frequencies from genotypic data by means of the EM algorithm and computation of the likelihood ratio test were performed by the ARLEQUIN program (Schneider *et al*, 2000). The EM algorithm was always started from 100 random initial haplotype frequencies, to avoid nonconvergence problems in the iterative process and to ensure finding the global maximum likelihood estimate (Excoffier and Slatkin, 1995). The significance of the likelihood ratio test was based on the empirical distribution of likelihood ratios under the null hypothesis of gametic equilibrium generated from 20 000 randomizations, that is, random permutations of alleles at each locus (Slatkin and Excoffier, 1996).

Results

The protein pattern revealed by two-dimensional gel electrophoresis of an individual foot sample of *M. galloprovincialis* is shown in Figure 2. In any given gel, approximately 1200 polypeptide spots could be routinely visualized by general protein staining. Our analysis was limited to a subset of 86 polypeptides selected on the basis of three criteria: (i) reproducibility, (ii) intensity and (iii) relative isolation on the gel (Rosenblum *et al*, 1983, 1984; Asakawa *et al*, 1985; Hanash *et al*, 1986a, b). The choice of spots was made by two investigators who had not previously scored any of them with respect to variability, that is, selected in an unbiased fashion. With respect to reproducibility, spots at the margins of the gel were discarded because of variable migration. The criterion for intensity was that if genetic variation were to result in two spots instead of one (ie a heterozygote), both spots would have a staining intensity above the background. With respect to isolation, spots in crowded regions or spots associated with streaking were not included in the analysis. We arbitrarily assign numbers to the spots to be analysed, working from the upper left to the lower right of the gel. All gels were scored independently by two investigators.

Genetic polymorphisms were detected as position mobility shifts due to charge alteration. All the putative polymorphisms exhibited gene dosage dependence in heterozygotes, which is consistent with a genetic basis. As an example, various genetic polymorphisms are illustrated in Figure 3. Allelic variants were named from the basic side of the gel to the acid one. Of the 86 protein spots examined, 32 and 26 exhibited genetic polymorphisms in the Ribadeo and Vilagarcía samples, respectively. The positions on the gel of the polypeptides scored in this study and the identity of five of them, determined by mass spectrometry (López *et al*, 2001), are indicated in Figure 4.

Genotype frequencies and analysis of Hardy–Weinberg deviations for the polymorphic protein loci in the Ribadeo and Vilagarcía samples are shown in Tables 1 and 2, respectively. Only three of a total of 32 polymorphic loci showed statistically significant departures from Hardy–Weinberg expectations in the Ribadeo sample. Locus 2 showed a significant heterozygote deficiency ($f_T = +0.490$, $P < 0.01$; $X^2 = 6.47$, $P < 0.01$; exact test, $P = 0.10$), while a significant heterozygote excess ($f_T = -0.430$, $P < 0.05$; $X^2 = 5.81$, $P < 0.05$; exact test, $P = 0.03$) was found for locus 48. For locus 46, statistically significant deviations were not detected by the f_T and the χ^2 tests ($f_T = +0.171$, $P > 0.05$; $X^2 = 5.74$, $P > 0.05$) but the probability of the exact test was 0.0497 and a statistically significant heterozygote deficiency was detected for allele 20 at this locus ($f_{ii} = +0.424$, $P < 0.05$). In the Vilagarcía sample, only two loci (6 and 76) present statistically significant deviations from Hardy–Weinberg proportions ($f_T = +0.703$, $P < 0.05$, for both loci). When a Bonferroni correction is applied in each sample to control for potential Type I errors due to the use of multiple tests, statistical significance of Hardy–Weinberg deviations is not found for any of these five loci. The deviations from Hardy–Weinberg expectations expressed by means of the two F_{IS} statistics, f_C and f_T , displayed similar values in both samples. In the Ribadeo sample, f_C ranged from -0.426 to 0.478 and f_T varied between -0.430 and 0.490

Figure 2 Two-dimensional protein pattern obtained from the foot of *M. galloprovincialis*. Numbers identify the 86 polypeptides analysed and letters (a = acidic, b = basic spot) indicate polypeptides for which variants were observed. The pI is indicated on the horizontal axis and the molecular weight on the vertical axis.

Figure 3 Examples of three genetic polymorphisms. From left to right, the acidic homozygote, the heterozygote and the basic homozygote are displayed. Arrows point to the location of absent and present allelic variants.

across loci. The mean values over loci were -0.004 ± 0.018 and -0.011 ± 0.016 for f_C and f_T respectively. In the Vilagarcía sample, f_C ranged from -0.581 to 0.632 and f_T varied between -0.571 and 0.703 , the average values being -0.065 ± 0.032 and -0.033 ± 0.032 , respectively. In both samples, the averages of f_T and f_C

estimated from 2-DE polymorphisms are very close to the heterozygote excess expected by sampling from a population in Hardy-Weinberg equilibrium, which is $F_{IS} = -(1/(2N - 1))$ (Kirby, 1975; Robertson and Hill, 1984). Thus, the F_{IS} value expected by sampling is -0.017 for Ribadeo ($N=30$) and -0.066 for Vilagarcía

Figure 4 Schematic representation of the 86 polypeptides analysed. Polymorphic variants are joined with a line. The most frequent allelic variants are in black and the rarer are in grey. Monomorphic proteins are outlined.

($N=8$). It appears, therefore, that genotype data are generally consistent with their Hardy–Weinberg proportions for the Ribadeo and Vilagarcía populations, although significant deviations could be occurring in some specific loci, particularly towards a deficiency of heterozygote genotypes. Heterozygote deficiency is a common observation from several species of marine molluscs and has been described in *M. galloprovincialis* for some allozyme loci (Sanjuan *et al*, 1990, 1994; Raymond *et al*, 1997). Although no satisfactory explanation for this observation has yet been identified, the possibility that locus-specific factors such as selection or Wahlund effect are involved has been suggested (Raymond *et al*, 1997).

Indexes of genetic variability (expected heterozygosity, degree of polymorphism and number of alleles per locus) for the Ribadeo and Vilagarcía population samples are given in Table 3. Patterns of polymorphism characterized by the number and the frequency of the protein variants were similar in the two population samples. The Ribadeo sample showed 54 monomorphic and 32 polymorphic loci (16 loci with two, 12 loci with three and four loci with 4 alleles) while the Vilagarcía sample presented 60 monomorphic and 26 polymorphic loci (17 loci with two, 5 loci with three and 4 loci with four alleles). The estimates of genetic variation were also very similar for both populations. Average heterozygosity was 0.101 ± 0.018 and 0.114 ± 0.021 in Ribadeo and Vilagarcía, respectively. The degree of gene differentiation (G_{ST}) between the two populations for single loci ranged from

0 (locus16) to 0.544 (locus 23) and the G_{ST} estimated for the whole loci set was 0.067. These estimates are higher than those reported for natural mussel populations of the Atlantic coast of the Iberian Peninsula based on allozyme data (Quesada *et al*, 1995). This could be due to the small sample size of the Vilagarcía sample as well as to the sampling process associated with mussel culture in rafts from individuals collected in natural populations.

A total of 406 two-locus associations were evaluated from 29 polymorphic loci of the Ribadeo population sample. Loci 2, 46 and 48 were not included in this analysis since Hardy–Weinberg departures were found at these three loci, and the maximum likelihood estimation of haplotype frequencies is based on the assumption of Hardy–Weinberg proportions (Hill, 1974; Excoffier and Slatkin, 1995). Statistical significance for gametic association between loci was detected in 18 pairs (4.43%) by the likelihood ratio test and in 81 pairs (19.95%) by χ^2 tests (with Bonferroni correction) for each pair of alleles. When using a single χ^2 test for the most frequent alleles at a given locus pair, significant associations were detected in 53 locus pairs (13.05%). Overall, 92 (22.66%) of 406 pairs of loci show statistically significant associations as detected by some of the different tests. Our results show that the χ^2 tests detect more significant gametic associations than the likelihood ratio test. However, it must be emphasized that the statistical power for detecting significant gametic disequilibrium by the χ^2 test is not large, particularly when allelic

Table 1 Genotypic frequencies for 32 polymorphic protein loci in the Ribadeo population sample. Deviations from Hardy–Weinberg proportions are measured by the F_{IS} statistics, f_C of Weir and Cockerham (1984) and f_T of Robertson and Hill (1984). Statistical significance of these deviations is evaluated by the test based on f_T estimator (Robertson and Hill, 1984), the χ^2 test and the exact probability test

Locus	Genotypes															N	f_C	f_T	χ^2	Exact test (probability)
	10/10	10/20	10/30	10/40	10/50	20/20	20/30	20/40	20/50	30/30	30/40	30/50	40/40	40/50	50/50					
2	27	2	—	—	—	1	—	—	—	—	—	—	—	—	—	30	+0.478	+0.490**	6.47**	0.10
3	0	8	0	—	—	21	1	—	—	0	—	—	—	—	—	30	−0.140	−0.073	0.92	1
4	0	1	—	—	—	29	—	—	—	—	—	—	—	—	—	30	0.000	0.000	0.01	—
6	0	1	—	—	—	29	—	—	—	—	—	—	—	—	—	30	0.000	0.000	0.01	—
8	0	6	—	0	0	19	—	3	1	—	—	—	1	0	0	30	+0.045	+0.082	0.55	0.45
14	0	1	0	—	—	17	10	—	—	2	—	—	—	—	—	30	+0.059	+0.038	0.11	0.73
16	13	12	—	—	—	5	—	—	—	—	—	—	—	—	—	30	+0.155	+0.158	0.58	0.45
17	0	2	0	—	—	25	3	—	—	0	—	—	—	—	—	30	−0.051	−0.029	0.25	1
19	24	5	—	—	—	1	—	—	—	—	—	—	—	—	—	30	+0.208	+0.212	1.10	0.33
23	0	4	0	—	—	20	5	—	—	0	—	—	—	—	—	29	−0.115	−0.076	0.98	1
24	—	—	—	—	—	23	5	2	—	0	0	—	0	—	—	30	−0.086	−0.050	0.52	1
26	9	3	7	6	—	2	0	1	—	0	1	—	1	—	—	30	+0.042	+0.095	3.79	0.32
29	3	10	—	—	—	17	—	—	—	—	—	—	—	—	—	30	+0.164	+0.168	0.66	0.38
39	0	1	—	—	—	28	—	—	—	—	—	—	—	—	—	29	0.000	0.000	0.01	—
40	0	3	0	0	—	22	3	0	—	1	1	—	0	—	—	30	+0.219	+0.090	1.07	0.08
41	16	12	—	—	—	0	—	—	—	—	—	—	—	—	—	28	−0.256	−0.259	2.08	0.30
46	0	0	2	—	—	11	9	—	—	8	—	—	—	—	—	30	+0.323	+0.171	5.74	0.05
48	2	21	—	—	—	7	—	—	—	—	—	—	—	—	—	30	−0.426	−0.430*	5.81*	0.03
50	15	13	—	—	—	2	—	—	—	—	—	—	—	—	—	30	−0.050	−0.051	0.14	1
51	0	1	0	—	—	26	3	—	—	0	—	—	—	—	—	30	−0.040	−0.019	0.15	1
53	3	13	—	—	—	14	—	—	—	—	—	—	—	—	—	30	+0.016	+0.016	0.00	1
57	—	—	—	—	—	0	5	0	—	12	12	—	0	—	—	29	−0.286	−0.193	4.99	0.32
58	0	1	0	—	—	28	1	—	—	0	—	—	—	—	—	30	−0.009	0.000	0.02	1
59	4	15	—	—	—	11	—	—	—	—	—	—	—	—	—	30	−0.041	−0.041	0.10	1
64	0	2	0	—	—	23	3	—	—	0	—	—	—	—	—	28	−0.055	−0.031	0.27	1
66	0	3	—	—	—	27	—	—	—	—	—	—	—	—	—	30	−0.036	−0.036	0.08	1
67	0	0	1	—	—	0	1	—	—	28	—	—	—	—	—	30	−0.009	0.000	0.02	1
68	0	3	—	—	—	27	—	—	—	—	—	—	—	—	—	30	−0.036	−0.036	0.08	1
76	25	5	—	—	—	0	—	—	—	—	—	—	—	—	—	30	−0.074	−0.075	0.25	1
79	27	3	—	—	—	0	—	—	—	—	—	—	—	—	—	30	−0.036	−0.036	0.08	1
80	0	3	0	—	—	24	2	—	—	0	—	—	—	—	—	29	−0.053	−0.030	0.26	1
86	0	0	5	1	—	1	5	1	—	11	5	—	0	—	—	29	−0.051	−0.035	0.69	0.91
Mean values $\pm$ SE																	−0.004 $\pm$ 0.018		−0.011 $\pm$ 0.016	

* $P < 0.05$. ** $P < 0.01$.

Table 2 Genotypic frequencies for 26 polymorphic protein loci in the Vilagarcía population sample. Deviations from Hardy–Weinberg proportions are measured by the F_{IS} statistics, f_C of Weir and Cockerham (1984) and f_T of Robertson and Hill (1984). Statistical significance of these deviations is evaluated by the test based on f_T estimator (Robertson and Hill, 1984), the χ^2 test and the exact probability test

Locus	Genotypes															N	f_C	f_T	χ^2	Exact test (probability)
	10/10	10/20	10/30	10/40	10/50	20/20	20/30	20/40	20/50	30/30	30/40	30/50	40/40	40/50	50/50					
3	0	2	—	—	—	6	—	—	—	—	—	—	—	—	—	8	−0.077	−0.082	0.16	1
6	1	1	—	—	—	6	—	—	—	—	—	—	—	—	—	8	+0.632	+0.703*	2.78	0.20
8	—	—	—	—	—	5	2	—	1	0	—	0	—	—	0	8	−0.105	−0.052	0.26	1
14	0	1	0	0	—	2	4	1	—	0	0	—	0	—	—	8	−0.333	−0.143	2.18	1
16	2	6	—	—	—	0	—	—	—	—	—	—	—	—	—	8	−0.556	−0.571	2.88	0.44
19	5	3	—	—	—	0	—	—	—	—	—	—	—	—	—	8	−0.167	−0.176	0.43	1
23	6	2	—	—	—	0	—	—	—	—	—	—	—	—	—	8	−0.077	−0.082	0.16	1
24	1	0	0	—	—	3	4	—	—	0	—	—	—	—	—	8	+0.125	+0.393	2.17	0.09
26	2	2	1	1	—	1	1	0	—	0	0	—	0	—	—	8	+0.079	+0.036	0.36	1
29	0	4	—	—	—	4	—	—	—	—	—	—	—	—	—	8	−0.273	−0.286	0.89	1
40	0	0	0	1	—	5	1	1	—	0	0	—	0	—	—	8	+0.160	+0.012	0.60	0.39
46	—	—	—	—	—	3	5	—	—	0	—	—	—	—	—	8	−0.400	−0.416	1.65	0.49
48	2	4	—	—	—	2	—	—	—	—	—	—	—	—	—	8	+0.067	+0.071	0.00	1
50	3	5	—	—	—	0	—	—	—	—	—	—	—	—	—	8	−0.400	−0.416	1.65	0.49
51	—	—	—	—	—	7	1	—	—	0	—	—	—	—	—	8	0.000	0.000	0.04	—
53	0	3	—	—	—	5	—	—	—	—	—	—	—	—	—	8	−0.167	−0.176	0.43	1
57	0	0	1	0	—	0	3	0	—	3	1	—	0	—	—	8	−0.228	−0.091	0.71	1
58	0	1	—	0	—	6	—	1	—	—	—	—	0	—	—	8	−0.037	−0.005	0.08	1
59	3	3	—	—	—	2	—	—	—	—	—	—	—	—	—	8	+0.300	+0.327	0.45	0.53
64	0	1	0	—	—	5	1	—	—	0	—	—	—	—	—	7	−0.044	−0.007	0.10	1
66	—	—	—	—	—	7	1	—	—	0	—	—	—	—	—	8	0.000	0.000	0.04	—
67	0	—	1	—	—	—	—	—	—	7	—	—	—	—	—	8	0.000	0.000	0.04	—
75	5	3	—	—	—	0	—	—	—	—	—	—	—	—	—	8	−0.167	−0.176	0.43	1
76	6	1	—	—	—	1	—	—	—	—	—	—	—	—	—	8	+0.632	+0.703*	2.78	0.20
79	6	2	—	—	—	0	—	—	—	—	—	—	—	—	—	8	−0.077	−0.082	0.16	1
86	0	—	6	0	—	—	—	—	—	1	1	—	0	—	—	8	−0.581	−0.345	4.09	0.14
Mean values ± SE																	−0.065 ± 0.032		−0.033 ± 0.032	

* $P < 0.05$.

Table 3 Estimates of genetic variability from 2-DE in two populations of *Mytilus galloprovincialis*

	Ribadeo	Vilagarcía
Sample size	30	8
Proportion of polymorphic loci	0.372	0.302
Average number of alleles per locus	1.61 ± 0.10	1.45 ± 0.09
Average heterozygosity	0.101 ± 0.018	0.114 ± 0.021

frequencies are extreme and disequilibrium is not relatively intense (Brown, 1975; Zapata and Alvarez, 1992). This must be particularly considered in our case, given that the sample size is not large ($2N=60$). In general, two broad patterns of gametic disequilibrium can be observed in those pairs of loci where statistically significant associations are detected by tests. In 53 locus pairs, the statistical association is mainly due to the most frequent alleles, while in 39 locus pairs gametic disequilibrium is associated only with rare alleles (alleles at a low or intermediate frequency). In the first type of locus pair, the magnitude of gametic disequilibrium detected is high for the overall disequilibrium ($D' = 0.687 \pm 0.036$) and for the disequilibrium of the most frequent alleles ($|D'_{ij}| = 0.699 \pm 0.037$), while in those locus pairs where statistically significant disequilibrium is due to rare alleles, the extent of disequilibrium observed is clearly lower for both the overall disequilibrium ($D' = 0.411 \pm 0.040$) and the disequilibrium exhibited by the most frequent alleles ($|D'_{ij}| = 0.334 \pm 0.045$).

From the disequilibrium analysis, groups of loci are detected in which all pairwise combinations exhibit statistically significant gametic disequilibrium. One of these clusters involving six loci (8, 17, 24, 39, 79 and 80) is shown in Table 4. Overall D' for locus pairs in this cluster ranges from 0.224 to 1 with a mean value of 0.561 ± 0.088 . In addition, three loci of this cluster (39, 79 and 80) show significant associations between the most frequent alleles for all pairwise combinations (D'_{ij} for the most frequent alleles ranging from 0.621 to 1). Also, a cluster involving

Table 5 Cluster of four loci (58, 76, 79 and 80) showing statistically significant associations between all pairs. For each locus pair, the overall D' and the number of statistically significant $\chi^2_{ij}/$ number of total χ^2 (in parentheses) are indicated in the upper line; the $|D'_{ij}|$ and the χ^2_{ij} values for the most frequent alleles (in parentheses) are shown in the lower line

	58	76	79	80
58	—			
76	0.457 (1/3) 0.438 (4.34*)	—		
79	0.477 (2/3) 0.458 (8.15**)	0.607 (1/1) 0.607 (12.87***)	—	
80	0.492 (3/9) 0.453 (4.45*)	0.365 (2/3) 0.326 (6.15*)	0.636 (2/3) 0.621 (13.05***)	—

* $P < 0.05$. ** $P < 0.01$. *** $P < 0.001$.

four loci (58, 76, 79 and 80) has been detected (Table 5). Overall D' for locus pairs in this cluster varies from 0.365 to 0.636 with an average value of 0.506 ± 0.041 . In this cluster, all pairwise combinations show significant associations between the most frequent alleles (D'_{ij} range from 0.326 to 0.621). Other clusters involving three or four loci have also been detected.

Discussion

2-DE has been used to measure the degree of genetic variability for loci coding abundant proteins in the marine mussel *M. galloprovincialis*. Among the 86 polypeptides selected in an unbiased fashion for scoring for genetic variability, the natural population sample from Ribadeo exhibited such variation in 32 (37.2%) polypeptides with an average heterozygosity of 0.101 ± 0.018 . Similar results were obtained in a cultured population from Vilagarcía ($P = 30.2\%$, $H = 0.114 \pm 0.021$). These estimates of heterozygosity are clearly higher than those

Table 4 Cluster of six loci (8, 17, 24, 39, 79 and 80) showing statistically significant associations between all pairs. For each locus pair, the overall D' and the number of statistically significant $\chi^2_{ij}/$ over the number of total χ^2 (in parentheses) are indicated in the upper line; the $|D'_{ij}|$ and the χ^2_{ij} values for the most frequent alleles (in parentheses) are shown in the lower line

	8	17	24	39	79	80
8	—					
17	0.287 (2/12) 0.202 (0.89)	—				
24	0.273 (1/12) 0.188 (1.12)	0.224 (1/9) 0.086 (0.30)	—			
39	1 (1/4) 1 (3.84)	1 (2/3) 1 (13.53***)	1 (2/3) 1 (7.29**)	—		
79	0.598 (1/4) 0.575 (4.18*)	0.302 (1/3) 0.258 (2.33)	0.301 (1/3) 0.230 (1.26)	1 (1/1) 1 (18.29***)	—	
80	0.282 (2/12) 0.208 (0.91)	0.233 (1/9) 0.160 (1.16)	0.282 (1/9) 0.286 (3.24)	1 (2/3) 1 (10.66**)	0.636 (2/3) 0.621 (13.05***)	—

* $P < 0.05$. ** $P < 0.01$. *** $P < 0.001$.

Table 6 Estimates of genetic variability from two-dimensional and single-dimensional electrophoresis in animal species

Species/tissues	2-DE				1-DE		
	Sample size	No. of loci	Average heterozygosity	Ref.	No. of loci	Average heterozygosity	Ref.
Man (<i>Homo sapiens</i>)					71	0.067	(18)
Fibroblasts	5	200	0.006	(1)	121	0.143	(19)
Lymphocytes	28	186	0.024	(2)			
	3	250	0.005 ± 0.003	(3)			
	40	106	0.031 ± 0.002	(4)			
Erythrocytes	27	46	0.031 ± 0.005	(5)			
	100	55	0.040 ± 0.003	(6)			
Platelets	44	33	0.026	(7)			
Plasma	62	20	0.062 ± 0.007	(8)			
	107 (a)	23	0.045 ± 0.006	(9)			
	110 (b)	23	0.057 ± 0.007	(9)			
	62 (c)	23	0.080 ± 0.011				
Brain	145	176	0.0004	(10)			
Kidney	25	83	0.000	(11)			
					47	0.000	(12)
Cheetah (<i>Acinonyx jubatus</i>)							
Fibroblasts	6	155	0.013	(12)			
Mouse (<i>Mus musculus</i>)					41	0.085	(20)
Kidney	—	72	0.020	(13)			
Drosophila (<i>D. melanogaster</i>)					19	0.132	(20)
Whole-body	20 (e)	54	0.040	(14)			
MRT(d)	20 (e)	307	0.018	(15)			
Drosophila (<i>D. simulans</i>)					79	0.118	(21)
Whole-body	19 (e)	70	0.000	(16)			
MRT(d)	20 (e)	295	0.028	(15)			
Mussel (<i>Mytilus galloprovincialis</i>)					23	0.240	(22)
Foot	30	86	0.101 ± 0.018	(17)			
	8	86	0.114 ± 0.021				

Note: (1) Walton *et al* (1979); (2) Goldman and Merrill (1983); (3) Hamaguchi *et al* (1981); (4) Hanash *et al* (1986a); (5) Rosenblum *et al* (1984); (6) Takahashi *et al* (1986); (7) Hanash *et al* (1986b); (8) Rosenblum *et al* (1983); (9) Asakawa *et al* (1985); (10) Comings (1982); (11) Smith *et al* (1980); (12) O'Brien *et al* (1983); (13) Racine and Langley (1980); (14) Leigh Brown and Langley (1979); (15) Coulthart and Singh (1988); (16) Ohnishi *et al* (1982); (17) This work; (18) Harris and Hopkinson (1972); (19) Nei and Graur (1984); (20) Nevo (1978); (21) Singh *et al* (1987); (22) Grant and Cherry (1985). (a) Amerindians; (b) Japanese; (c) Caucasoids; (d) MRT: Male reproductive tract; (e) isofemale lines.

previously reported by 2-DE for animal populations (see Table 6) and put *M. galloprovincialis* as the animal species with the highest genetic variability detected by 2-DE to date. *M. galloprovincialis* heterozygosity is nearly six-fold higher than the heterozygosity values of *Drosophila* (average of heterozygosity estimates from different studies weighted by the number of loci scored is 0.021 and 0.023 for *D. melanogaster* and *D. simulans*, respectively), mouse (0.020), cheetah (0.013) and man (weighted average from different tissues is 0.017). Our findings clearly show a need to extend the 2-DE analysis to an increased number of species in order to obtain a wider view on genetic variability for loci responsible for abundant proteins in natural populations.

When comparing 2-DE and 1-DE estimates of heterozygosity for different animal species, two aspects deserve consideration (see Table 6). Firstly, although the 2-DE data set is still not large, the range of variation of heterozygosity estimates for different animal species seems to be clearly smaller than that reported from allozymic data. In mammals, for example, the allozyme heterozygosity varies from 0 up to 22% in 321 species reviewed by Makarieva (2001). It has been shown that

these variations can be mainly explained by differences in the number of loci studied (Makarieva, 2001). From this perspective, the larger number of loci analysed by 2-DE with respect to 1-DE could be one of the reasons explaining the small variation in 2-DE estimates of heterozygosity among animal species. Second, although 2-DE variation is generally reduced, it parallels the allozymic data in that more variable species, as indicated by 1-DE, are also shown to be more variable from 2-DE analysis (see Table 6). In fact, there is a statistically significant correlation ($r = 0.838$, d.f. 4, $P < 0.05$) between heterozygosity estimates yielded by both techniques among the species included in Table 6. Two alternative explanations for this discrepancy, not mutually exclusive, have been widely discussed in the literature (Edwards and Hopkinson, 1980; Aquadro and Avise, 1981; Wanner *et al*, 1982; Mc Lellan *et al*, 1983; De Vienne *et al*, 1996). For one, the group of proteins analysed on 2-D gels is intrinsically less variable than the soluble enzymes and proteins previously studied. For another, the technique of 2-DE is not capable of resolving many variants that are detectable by the conventional electrophoretic methods, since proteins are denatured and some differences of

conformational nature could remain undetectable. In this respect, Wanner *et al* (1982) tested the ability of 2-DE to resolve the allelic variants of five loci previously studied by 1-DE. 2-DE could resolve more than 90% of the variants originally detected by 1-DE (16 of the 17). The unique allelic variant that remained undetectable for 2-DE was attributed to a conformational variation. Furthermore, DNA sequence analysis for different electrophoretic classes at several enzyme loci (*Adh*, *Gpdh*, *Sod*, *Est-5*, *Est-6* and *Xdh*) in *Drosophila* has revealed that most differences (75–100%) in protein mobility detected by single electrophoresis are due to variations in the net charge of the protein (Riley *et al*, 1992; Veuille and King, 1995; Barbadilla *et al*, 1996). Therefore, because the conformational variations in protein structure seems to play a minor role in the electrophoretic mobility, the remarkable reductions in the level of variability detected by 2-DE with respect to 1-DE, ranging from 58 (*M. galloprovincialis*) to 85.2% (man) from data in Table 6, can hardly be attributed to conformational changes undetected by 2-DE. This suggests that the more feasible explanation for the lower levels of genetic variability detected by 2-DE is that the group of proteins analysed by this technique is intrinsically less variable than the enzymes assayed by 1-DE.

Among a total of 406 two-locus pairs analysed for detecting linkage disequilibrium in the Ribadeo population sample, 92 show statistically significant associations. This is a quite large proportion (22.7%), taking into account that the size of the studied sample is not very large ($2N=60$) and the statistical power of tests for detecting gametic disequilibrium is low when allelic frequencies are extreme or the disequilibrium is not relatively intense (Brown, 1975; Zapata and Alvarez, 1992). Furthermore, information on physical linkage between some protein loci can be extracted from the disequilibrium analysis. Thus, clusters of loci showing statistically significant disequilibrium between all the locus pairs have been found (Tables 4 and 5). In some cases, the significant disequilibrium is associated with rare alleles. In these cases, a strong linkage disequilibrium detected between two genetic loci may be due to the recent occurrence of disequilibrium rather than a close physical linkage of the two loci, since rare alleles are most likely to have arisen recently in the population by the introduction of new mutations. Most interesting are those cases where the significant disequilibria are associated with the most frequent alleles which are old alleles. Although there are many evolutionary forces that can cause an association between the alleles at different loci, the decay of genetic disequilibrium along generations is expected to be smaller for increasing physical linkage and it provides the rationale for inferring physical linkage from linkage disequilibrium (Jorde, 1995; Kaplan *et al*, 1995; Weir, 1996; Wu and Zeng, 2001). On this basis, a cluster of four loci (58, 76, 79 and 80, Table 5) and several clusters of three loci (39, 79 and 80, Table 4 and data not shown) where statistically significant disequilibria are occurring between all pairwise combinations for common alleles have been found. It is very likely that some physical linkage must be occurring between those protein loci included in clusters. This information will be very useful for planning future experiments of linkage analysis from single-pair matings. The previous knowledge of sets of protein loci where

physical linkage is present can be very useful in selecting informative families for traditional linkage analysis (Weir, 1996). To date, the available information on linkage of genetic markers is very scarce in marine mussels (Beaumont, 1994), so that investigations on this issue will be of great interest.

Acknowledgements

This research was supported by grant PGIDT00-MAR20001PR from the Xunta de Galicia (Spain) to GA.

References

- Aquadro CF, Avise JC (1981). Genetic divergence between rodent species assessed by using two-dimensional electrophoresis. *Proc Natl Acad Sci USA* **78**: 3784–3788.
- Asakawa J, Takahashi N, Rosenblum BB, Neel JV (1985). Two-dimensional gel studies of genetic variation in the plasma proteins of Amerindians and Japanese. *Hum Genet* **70**: 222–230.
- Barbadilla A, King LM, Lewontin RC (1996). What does electrophoretic variation tell us about protein variation? *Mol Biol Evol* **13**: 427–432.
- Beaumont, AR (1994). Linkage studies in *Mytilus edulis*, the mussel. *Heredity* **72**: 557–562.
- Brown AHD (1975). Sample sizes required to detect linkage disequilibrium between two of three loci. *Theor Popul Biol* **8**: 184–201.
- Comings DE (1982). Two-dimensional gel electrophoresis of human brain proteins. III. Genetic and non-genetic variations in 145 brains. *Clin Chem* **28**: 798–804.
- Coulthart MB, Singh RS (1988). Low genic variation in male-reproductive-tract proteins of *Drosophila melanogaster* and *D. simulans*. *Mol Biol Evol* **5**: 167–181.
- Dempster AP, Laird NM, Rubin DB (1977). Maximum likelihood from incomplete data via the EM algorithm. *J R Stat Soc B* **39**: 1–38.
- De Vienne D, Burstin J, Gerber S, Leonardi A, Le Guilloux M, Murigneux A *et al* (1996). Two-dimensional electrophoresis of proteins as a source of monogenic and codominant markers for population genetics and mapping the expressed genome. *Heredity* **76**: 166–177.
- Edwards Y, Hopkinson DA (1980). Are abundant proteins less variable? *Nature* **284**: 511–512.
- Excoffier L, Slatkin M (1995). Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. *Mol Biol Evol* **12**: 921–927.
- Goldman D, Merril CR (1983). Human lymphocyte polymorphisms detected by quantitative two-dimensional electrophoresis. *Am J Hum Genet* **35**: 827–837.
- Grant WS, Cherry MI (1985). *Mytilus galloprovincialis* Lmk. in southern Africa. *J Exp Mar Biol Ecol* **90**: 179–191.
- Guo SW, Thompson EA (1992). Performing the exact test of Hardy–Weinberg proportion for multiple alleles. *Biometrics* **48**: 361–372.
- Hamaguchi H, Ohta A, Mukai R, Yabe T, Yamada M (1981). Genetic analysis of human lymphocyte proteins by two-dimensional electrophoresis. 1. Detection of genetic variant polypeptides in PHA-stimulated peripheral blood lymphocytes. *Hum Genet* **59**: 215–220.
- Hanash SM, Baier LJ, Welch D, Kuick R, Galteau M (1986a). Genetic variants detected among 106 lymphocyte polypeptides observed in two-dimensional gels. *Am J Hum Genet* **39**: 317–328.
- Hanash SM, Neel JV, Baier LJ, Rosenblum BB, Niezgoda W, Markel D (1986b). Genetic analysis of thirty-three platelet polypeptides detected in two-dimensional polyacrylamide gels. *Am J Hum Genet* **38**: 352–360.

- Harris H, Hopkinson DA (1972). Average heterozygosity per locus in man: An estimate based on the incidence of enzyme polymorphisms. *Ann Hum Genet* **36**: 9–20.
- Hedrick PW (1987). Gametic disequilibrium measures proceed with caution. *Genetics* **117**: 331–341.
- Hill WG (1974). Estimation of linkage disequilibrium in randomly mating populations. *Heredity* **33**: 229–239.
- Jorde LB (1995). Linkage disequilibrium as a gene-mapping tool. *Am J Hum Genet* **56**: 11–14.
- Kaplan NL, Hill WG, Weir BS (1995). Likelihood methods for locating disease genes in nonequilibrium populations. *Am J Hum Genet* **56**: 18–32.
- Kirby GC (1975). Heterozygote frequencies in small subpopulations. *Theor Popul Biol* **8**: 31–48.
- Leigh Brown AJ, Langley CH (1979). Reevaluation of level of genic heterozygosity in natural populations of *Drosophila melanogaster* by two-dimensional electrophoresis. *Proc Natl Acad Sci USA* **76**: 2381–2384.
- Lewontin RC (1964). The interaction of selection and linkage. I. General considerations; heterotic models. *Genetics* **49**: 49–67.
- López JL, Mosquera E, Fuentes J, Marina A, Vázquez J, Alvarez G. (2001). Two-dimensional gel electrophoresis of *Mytilus galloprovincialis*: differences in protein expression between intertidal and cultured mussels. *Mar Ecol Prog Ser* **224**: 149–156.
- Louis EJ, Dempster ER (1987). An exact test for Hardy-Weinberg and multiple alleles. *Biometrics* **43**: 805–811.
- Makarieva AM. (2001). Variance of protein heterozygosity in different species of mammals with respect to the number of loci studied. *Heredity* **87**: 41–51.
- McLellan T, Ames GF, Nikaido K (1983). Genetic variation in proteins: Comparison of one-dimensional and two-dimensional gel Electrophoresis. *Genetics* **104**: 381–390.
- Morrissey JH (1981). Silver stain proteins in polyacrylamide gels: a modified procedure with enhanced uniform sensitivity. *Anal Biochem* **117**: 307–310.
- Neel JV (1990). Average locus differences in mutability related to protein 'class': a hypothesis. *Proc Natl Acad Sci USA* **87**: 2062–2066.
- Nei M (1973). Analysis of gene diversity in subdivided populations. *Proc Natl Acad Sci USA* **70**: 3321–3323.
- Nei M (1978). Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* **89**: 583–590.
- Nei M (1986). Definition and estimation of fixation indices. *Evolution* **40**: 643–645.
- Nei M, Graur D (1984). Extent of protein polymorphism and the neutral mutation theory. *Evol Biol* **17**: 73–118.
- Nevo E (1978). Genetic variation in natural populations: patterns and theory. *Theor Popul Biol* **13**: 121–177.
- O'Brien SJ, Wildt DE, Goldman D, Merrill CR, Bush M (1983). The cheetah is depauperate in genetic variation. *Science* **221**: 459–462.
- O'Farrell PH (1975). High resolution two-dimensional electrophoresis of proteins. *J Biol Chem* **250**: 4007–4021.
- Ohnishi S, Leigh Brown AJ, Voelker RA, Langley CH (1982). Estimation of genetic variability in natural populations of *Drosophila simulans* by two-dimensional and starch gel electrophoresis. *Genetics* **100**: 127–136.
- Quesada H, Sanjuan J, Sanjuan A (1992). GENET-2 a BASIC program for the analysis of deviations from Hardy-Weinberg law. *J Hered* **83**: 460–461.
- Quesada H, Zapata C, Alvarez G (1995). A multilocus allozyme discontinuity in the mussel *Mytilus galloprovincialis*: the interaction of ecological and life-history factors. *Mar Ecol Prog Ser* **116**: 99–115.
- Racine RR, Langley DH (1980). Genetic heterozygosity in a natural population of *Mus musculus* assessed using two-dimensional electrophoresis. *Nature* **283**: 855–857.
- Raymond M, Rousset F (1995). GENEPOP (version 1.2): population genetics software for exact test and ecumenicism. *J Hered* **86**: 248–249.
- Raymond M, Väntö RL, Thomas F, Rousset F, Meelis T, Renaud F (1997). Heterozygote deficiency in the mussel *Mytilus edulis* species complex revisited. *Mar Ecol Prog Ser* **156**: 225–237.
- Riley MA, Kaplan SR, Veuille M (1992). Nucleotide polymorphism at the xanthine dehydrogenase locus in *Drosophila pseudoobscura*. *Mol Biol Evol* **9**: 56–69.
- Robertson A, Hill WG (1984). Deviations from Hardy-Weinberg proportions sampling variances and use in estimation of inbreeding coefficients. *Genetics* **107**: 703–718.
- Rosenblum BB, Neel JV, Hanash SM (1983). Two-dimensional electrophoresis of plasma polypeptides reveals 'high' heterozygosity indices. *Proc Natl Acad Sci USA* **80**: 5002–5006.
- Rosenblum BB, Neel JV, Hanash SM, Joseph JL, Yew N (1984). Identification of genetic variants in erythrocyte lysate by two-dimensional gel electrophoresis. *Am J Human Genet* **36**: 601–612.
- Rousset F, Raymond M (1995). Testing heterozygote excess and deficiency. *Genetics* **140**: 1413–1419.
- Sanjuan A, Quesada H, Zapata C, Alvarez G (1990). On the occurrence of *Mytilus galloprovincialis* Lmk. on the N.W. coast of the Iberian Peninsula. *J Exp Mar Biol Ecol* **143**: 1–14.
- Sanjuan A, Zapata C, Alvarez G (1994). *Mytilus galloprovincialis* and *M. edulis* on the coasts of the Iberian Peninsula. *Mar Ecol Prog Ser* **113**: 131–146.
- Schneider S, Roessli D, Excoffier L (2000). *Arlequin: a software for population genetics data analysis*. Ver. 2.000. Genetics and Biometry Laboratory, Department of Anthropology, University of Geneva.
- Singh RS, Choudhary M, David JR (1987). Contrasting patterns of geographic variation in cosmopolitan sibling species *Drosophila melanogaster* and *Drosophila simulans*. *Biochem Genet* **25**: 27–40.
- Slatkin M, Excoffier L (1996). Testing for linkage disequilibrium in genotypic data using the Expectation-Maximization algorithm. *Heredity* **76**: 377–383.
- Smith SC, Racine RR, Langley CH (1980). Lack of genic variation in the abundant proteins of human kidney. *Genetics* **96**: 967–974.
- Takahashi N, Neel JV, Nagahata-Shimoichi, Asakawa J, Tanaka Y, Satoh C. (1986). Inherited electrophoretic variants detected in a Japanese population with two-dimensional gels of erythrocyte lysates. *Ann Hum Genet* **50**: 313–325.
- Veuille M, King M (1995). Molecular basis of polymorphism at the esterase 5B locus in *Drosophila pseudoobscura*. *Genetics* **141**: 255–262.
- Walton KE, Styer D, Gruenstein EI (1979). Genetic polymorphism in normal human fibroblasts as analysed by two-dimensional polyacrylamide gel electrophoresis. *J Biol Chem* **254**: 7951–7960.
- Wanner LA, Neel JV, Meisler MH (1982). Separation of allelic variants by two-dimensional electrophoresis. *Am J Hum Genet* **34**: 209–215.
- Ward RD, Skibinski DOF, Woodward M (1992). Protein heterozygosity, protein structure, and taxonomic differentiation. *Evol Biol* **26**: 73–159.
- Weir BS (1996). *Genetic Data Analysis II*. Sinauer: Sunderland, MA.
- Weir BS, Cockerham CC (1978). Testing hypotheses about linkage disequilibrium with multiple alleles. *Genetics* **88**: 633–642.
- Weir BS, Cockerham CC (1984). Estimating *f*-statistics for the analysis of population structure. *Evolution* **38**: 1358–1370.
- Wu R, Zeng Z-B (2001). Joint linkage and linkage disequilibrium mapping in natural populations. *Genetics* **157**: 899–909.
- Zapata C, Alvarez G (1992). The detection of gametic disequilibrium between allozyme loci in natural populations of *Drosophila*. *Evolution* **46**: 1900–1917.
- Zapata C, Rodríguez S, Visedo G, Sacristán F (2001). Spectrum of nonrandom associations between microsatellite loci on human chromosome 11p15. *Genetics* **158**: 1235–1251.