RESEARCH ARTICLE

Allelic structure and distribution of 103 STR loci in a Southern Tunisian population

ABDELLATIF MAALEJ[1], AHMED REBAI[2], ADNEN AYADI[1], JOMAA JOUIDA[3], HAFEDH MAKNI[1] and HAMMADI AYADI[1]*

[1]*Laboratoire de Génétique Moléculaire Humaine, Faculté de Médecine, Avenue Majida Boulila, 3029 Sfax, Tunisia*
[2]*Unité de bio-informatique, Centre de Biotechnologie de Sfax, B.P. 'K', 3038 Sfax,Tunisia*
[3]*Dispensaire de Bir ElHfai, Sidi Bouzid, Tunisia*

Abstract

Genotypes of 103 short tandem repeat (STR) markers distributed at an average of 40 cM intervals throughout the genome were determined for 40 individuals from the village of BirEl Hfai (BEH). This village of approximately 31.000 individuals is localized in the south-west of Tunisia. The allele frequency distributions in BEH were compared with those obtained for individuals in the CEPH (Centre d'Etude du Polymorphisme Humain) data using a Kolmogorov–Smirnov two-sample test. Fourteen out of the 103 markers (13.2%) showed significant differences (P<0.05) in distribution between the two populations. Population heterogeneity in BEH was indicated by an excess of observed homozygosity deviations from Hardy–Weinberg equilibrium at 3 loci (P<0.0005). No evidence for genotypic disequilibrium was found for any of the marker pairs. This demonstrated that in spite of a high inbreeding level in the population, few markers showed evidence for a different pattern of allelic distribution compared to CEPH.

[Maalej A., Rebai A., Ayadi A., Jouida J., Makni H. and Ayadi H. 2004 Allelic structure and distribution of 103 STR loci in a Southern Tunisian population. *J. Genet*. **83**, 65–71]

Introduction

Short tandem repeat (STR) polymorphisms currently form the basis of the most extensive genetic maps of humans and are commonly used in the mapping of disease genes (Weber *et al*. 1989; Buetow *et al*.1994; Gyapay *et al*. 1994; Matise *et al*. 1994). More recently, they are being utilized in the construction of evolutionary trees (Bowcock *et al*. 1994) and in forensic identification (Fregeau *et al*. 1993; Gill *et al*. 1994). In all of these applications, the interpretation of the results would be affected by, and in some cases depends upon, differences in the allele frequency distributions between populations. In particular, the use of DNA typing systems in forensic identification has been criticized, in part, on the grounds that the calculations used for assessing likelihood of matches were allele frequency dependent, which raised the question of what constituted a proper reference population for determining allele frequencies (Lander *et al*. 1989; Chakraborty and Kidd 1991; Lewontin and Hartl 1991; Risch and Devlin 1992). Variation in allele frequency distributions between populations, and the possibility of undetected subpopulations of the reference population, could alter the probability of a match calculation. Most studies addressing the issue of genetic sub-structuring in human populations have employed protein electrophoretic polymorphisms (Chakraborty and Ghosh 1981; Chakraborty *et al*. 1988; Aguirre *et al*. 1991), or VNTR (variable number of tandem repeats) marker systems (Deka *et al*. 1991, 1992). Only a few studies of human populations using STRs have been published (Edwards *et al*. 1992; Wall *et al*. 1993). Despite the lower informativeness of STR versus VNTR markers, their use has been likely to increase, due to the ease and reliability of scoring (Budowle *et al*. 1991; Edwards *et al*. 1992; Pena and Chakraborty 1994). It is important, therefore, to address the issue of

*For correspondence. E-mail: hammadi.ayadi@fmsf.rnu.tn.

Keywords. microsatellite markers; polymorphism; allele frequency; population studies; disequilibrium.

population substructuring with STR markers. In the present paper, we evaluated the allele spectrum of a large number of STR loci evenly distributed on 22 chromosomes and compared their polymorphism to that in the European population.

The allele frequencies of STR marker loci were determined from genotypes of a representative sample from the village of BirEl Hfai in Tunisia. These data were collected as part of a project to map the gene(s) of susceptibility to thyroid autoimmune disorders (Maalej *et al*. 2001). Association as well as linkage studies were employed which required the assessment of marker allele frequencies. The analyses presented in this paper served as a resource to instruct both inquiries, as well as to provide pertinent information to the wider issue of genetic diversity of STRs in the human population.

Materials and methods

Genomic DNA was extracted from 40 healthy subjects originating from the region of the village of BirEl Hfai (BEH) in Tunisia. The female / male ratio was of 42.5%. The subjects were genotyped for 103 microsatellite markers distributed on twentytwo chromosomes. Polymorphic microsatellites were amplified by the polymerase chain reaction using a radioactive-detection system (see Maalej *et al*. 2001 for details).

Statistical analyses

Expected heterozygosity at each locus were calculated using the program HET, which is part of the LINKAGE Utilities (v.5.2) package (Ott 1999). The average (over all loci) observed heterozygosity for the CEPH (Centre d'Etude du Polymorphisme Humain) and Tunisian population were compared with a standard *t* test at 5% level of significance. The allele frequency distributions in the two populations were compared using a Kolmogorov–Smirnov two-sample test. The statistic test, D_{KS}, was calculated as the absolute value of the maximum difference between the frequency of allele *i* in the first and in the second population, $D_{KS} = \max [|p_{i1} - p_{i2}|]$. The significance threshold for a level α was calculated as: $D_{KS}^{crit} = C_\alpha[(n_1 - n_2)/n_1 n_2]^{1/2}$ (valid for n_1 and $n_2 > 40$) where $C_\alpha = 1.36$ and 1.95 for $\alpha = 5\%$ and 0.1%, respectively.

Exact tests for HWE (Hardy–Weinberg equilibrium) (Guo and Thompson 1992) at each locus were performed with the program GenePop version 3.3 (Raymond and Rousset 1995) using as parameters for the Markov chain: 100 batches and 1000 iterations per batch for each locus.

Linkage disequilibrium among unlinked markers showing a different pattern of allele distribution from that in CEPH sample was assessed using the program GDA (Lewis and Zaykin 2001). This program calculates the composite coefficient of gametic linkage disequilibrium between two loci (noted D_{AB} for loci A and B), from genotypic data (Weir 1996). We also assessed, using GDA, linkage disequilibrium among a region of six STR markers spanning a 4.3 cM interval on chromosome 2. This was the region where the highest linkage evidence with autoimmune thyroid diseases was found in a large family from the BEH region (Maalej *et al*. 2001). The markers used and the distance between them are given below (according to the last updated Marshfield map and based on the physical distance between D2S2168 and D2S171):

D2S2221–1.2 cM–D2S2168–0.4 cM–D2S171–1.1 cM–D2S2144–0.5 cM–D2S174–1.1 cM–D2S365

Results

Allele frequency distributions in our sample were compared with those of the CEPH database. Alleles were standardized by typing markers on an individual with known genotype (mother of family 13047 from CEPH) to ensure the proper identity of alleles in the two populations. Results of the Kolmogorov–Smirnov test comparing the frequency distribution of alleles between BEH and CEPH samples are given in table 1. Fourteen among the 103 loci (13.6%) tested showed significant difference in distributions between these two populations at the 5% significance level. However, these differences were not evenly distributed on all chromosomes as could be seen from table 1. Particularly, chromosome 3 had two loci with significant differences at 5% level, but all its six loci had significant differences at 10% level with an average D_{KS} value of 0.24 (very close to the critical value at 5% which is 0.25). Chromosome 11 had also two markers (among four) which showed a high difference in frequency between BEH and CEPH populations. However, only three markers showed significant differences at the 0.1% level (D2S113, D10S535 and D11S903) but no evidence for genotypic or gametic disequilibrium between these three loci was found (results not shown).

Comparisons between levels of heterozygosity in BEH and CEPH datasets, and between observed and expected heterozygosities in the BEH population (tested as a departure from HWE) are given in table 2. The level of heterozygosity in BEH was significantly different from that of the CEPH. The *t* test between the two populations showed that the mean observed heterozygosity level in BEH (0.7535 ± 0.0021) was significantly lower ($P = 0.012$) than the observed mean for the CEPH population (0.7821 ± 0.0013). In addition, the histograms of observed *H* values showed different patterns of distribution between the two populations; the distribution in the BEH population was skewed with a long left-tail for small *H* values whereas in the CEPH sample the distribution was rather symmetric (figure 1). Sixtyfive among the 103 STR had observed *H* values larger in the CEPH sample than in the BEH sample and the difference between *H* values was >0.1 for 18 of these STRs.

Table 1. Comparison of allele distributions between CEPH and South Tunisian population.

Chromosome	Marker	No. chromosomes Tun.	CEPH	D_{KS} test[a]	Chromosome	Marker	No. chromosomes Tun.	CEPH	D_{KS} test[a]
1	D1S228	74	52	0.233	12	D12S77	74	56	0.127
	D1S198	78	56	0.213		D12S87	76	56	0.118
	D1S305	70	56	0.205		D12S1653	66	25	0.192
	D1S249	70	56	0.105		D12S83	67	26	0.128
	D1S229	74	56	0.125		D12S86	72	26	0.111
	D1S235	74	56	0.216		D12S357	76	56	0.474
2	D2S281	70	50	0.150	13	D13S171	70	56	0.215
	D2S2221	66	48	0.223		D13S170	74	56	0.094
	D2S159	74	56	0.100		D13S286	74	56	0.240
	D2S151	76	56	0.142		D13S285	72	54	0.240
	D2S2378	68	28	0.147					
	D2S113	72	50	**0.360**					
	D2S2328	70	28	0.168					
3	D3S1307	74	56	0.230	14	D14S261	74	56	0.090
	D3S1266	62	56	0.199		D14S75	76	56	0.223
	D3S1285	74	56	**0.288**		D14S63	76	56	0.115
	D3S1292	64	56	**0.273**		D14S74	72	54	0.283
	D3S1282	72	56	0.230		D14S65	70	56	0.320
	D3S1262	74	54	0.243		D14S78	74	56	0.100
						D14S292	74	56	0.110
						D14S272	70	54	0.201
4	D4S2946	66	28	0.110	15	D15S205	60	54	0.118
	D4S392	76	48	**0.260**		D15S120	76	56	0.118
	D4S2983	70	28	0.110					
	D4S1548	72	56	0.110					
	D4S3033	74	27	0.082					
	D4S426	68	56	0.191					
5	D5S2095	66	28	0.173	16	D16S521	74	56	0.219
	D5S2114	66	28	0.112		D16S407	76	56	0.190
	D5S418	70	56	0.117		D16S411	76	56	0.111
	D5S424	74	128	0.128		D16S422	76	52	0.198
	D5S421	76	128	**0.211**					
	D5S400	76	56	0.129					
	D5S408	78	56	0.220					
	D5S393	76	56	0.140					
6	D6S344	74	56	0.137	17	D17S926	70	54	0.130
	D6S281	68	26	0.183		D17S786	76	56	**0.250**
	D6S290	74	56	0.080		D17S953	66	56	**0.325**
	D6S292	74	56	0.230		D17S1880	72	28	0.111
	D6S261	64	56	0.240		D17S787	74	56	0.121
	D6S466	76	56	0.105		D17S949	68	54	**0.254**
	D6S273	70	56	0.228		D17S784	76	54	0.091
7	D7S664	76	56	0.192	18	D18S53	74	56	0.109
	D7S484	76	56	0.130		D18S70	76	56	**0.290**
	D7S639	58	54	0.344					
	D7S669	76	56	0.110					
	D7S486	80	50	**0.320**					
	D7S495	72	54	0.200					
	D7S550	72	56	0.200					
8	D8S550	72	56	0.110	19	D19S413	76	56	0.220
	D8S514	72	56	0.188		D19S413	76	54	0.220
9	D9S288	72	56	0.142	20	D20S107	74	52	0.190
	D9S1845	70	28	0.123					
	D9S153	76	54	0.174					
	D9S164	64	56	0.190					
10	D10S249	74	56	0.090	22	D22S424	72	56	0.222
	D10S191	74	56	0.158					
	D10S193	68	54	0.170					
	D10S535	74	56	**0.508**					
	D10S212	74	56	0.081					
11	D11S922	60	56	0.147	X	DXS999	43	42	0.368
	D11S1349	76	56	0.136		DXS1038	49	42	0.299
	D11S903	74	56	**0.383**		DXS1275	45	28	0.311
	D11S1332	74	56	**0.309**		DXS1205	49	28	**0.428**

[a]Values in bold are significant at the 5% level and those in bold and underlined are significant at 0.1% level.

Table 2. Comparison of heterozygosity level between CEPH population and South Tunisian population and results of test for deviation from HWE.

Marker locus	No. of alleles		Observed *H*		Theroretical *H*		Marker locus	No. of alleles		Observed *H*		Theroretical *H*	
	Tun.	CEPH	Tun.	CEPH	HWE	Test[a]		Tun.	CEPH	Tun.	CEPH	HWE	Test
D1S228	7	7	0.567	0.760	0.678	0.006	D12S77	14	13	0.837	0.870	0.862	0.057
D1S198	7	8	0.897	0.800	0.804	0.223	D12S87	7	9	0.815	0.790	0.780	0.055
D1S305	7	9	0.857	0.820	0.815	0.793	D12S1653	8	8	0.727	0.810	0.726	0.076
D1S249	11	15	0.914	0.870	0.847	0.903	D12S83	9	8	0.735	0.800	0.787	0.449
D1S229	6	8	0.675	0.770	0.690	0.203	D12S86	18	18	0.888	0.880	0.924	0.492
D1S235	10	7	0.756	0.680	0.832	0.042	D12S357	9	11	0.6310	0.850	0.682	0.147
D2S281	7	9	0.743	0.792	0.749	0.908	D13S171	8	6	0.657	0.720	0.650	0.505
D2S2221	12	10	0.941	0.830	0.849	0.649	D13S170	13	13	0.783	0.890	0.852	0.013
D2S159	7	6	0.756	0.760	0.811	**0.0004**	D13S286	7	8	0.810	0.760	0.746	0.873
D2S151	6	8	0.631	0.820	0.768	0.309	D13S285	12	8	0.833	0.810	0.830	0.056
D2S2378	10	13	0.787	0.850	0.815	0.271							
D2S113	10	10	0.777	0.777	0.842	0.246							
D2S2328	13	13	0.794	0.890	0.836	0.047							
D3S1307	10	8	0.710	0.800	0.820	0.033	D14S261	8	7	0.756	0.760	0.752	0.802
D3S1266	6	5	0.645	0.720	0.692	0.246	D14S75	8	9	0.657	0.760	0.760	0.121
D3S1285	5	6	0.729	0.720	0.734	0.153	D14S63	8	9	0.657	0.535	0.670	0.037
D3S1292	10	10	0.812	0.840	0.818	0.402	D14S74	9	9	0.666	0.790	0.774	0.009
D3S1282	6	7	0.583	0.780	0.653	0.464	D14S65	11	9	0.828	0.790	0.990	0.710
D3S1262	9	8	0.783	0.790	0.795	0.061	D14S78	6	6	0.594	0.660	0.741	0.087
							D14S292	5	5	0.756	0.740	0.774	0.498
							D14S272	5	5	0.457	0.470	0.551	0.037
D4S2946	13	12	1	0.890	0.874	1	D15S205	12	14	1.00	0.880	0.888	0.472
D4S392	9	7	0.763	0.820	0.809	0.102	D15S120	10	8	0.684	0.730	0.738	0.392
D4S2983	7	13	0.657	0.870	0.803	**0.0002**							
D4S1548	10	9	0.833	0.812	0.812	0.582							
D4S3033	9	9	0.675	0.780	0.790	0.017							
D4S426	9	6	0.794	0.760	0.796	0.644							
D5S2095	12	13	0.848	0.890	0.855	0.875	D16S521	11	7	0.756	0.710	0.827	0.477
D5S2114	11	15	0.818	0.880	0.853	0.880	D16S407	11	10	0.789	0.850	0.827	0.102
D5S418	8	9	0.900	0.800	0.797	0.034	D16S411	8	6	0.710	0.780	0.817	0.031
D5S424	9	7	0.702	0.668	0.725	0.018	D16S422	15	9	0.868	0.780	0.866	0.798
D5S421	6	8	0.736	0.825	0.672	0.304							
D5S400	11	10	0.868	0.810	0.835	0.692							
D5S408	7	7	0.435	0.730	0.702	**$<10^{-4}$**							
D5S393	11	10	0.894	0.820	0.869	0.458							
D6S344	9	9	0.864	0.720	0.805	0.071	D17S926	9	15	0.885	0.810	0.825	0.232
D6S281	10	8	0.823	0.670	0.784	0.734	D17S786	7	7	0.710	0.760	0.756	0.649
D6S290	7	6	0.675	0.700	0.676	0.026	D17S953	9	8	0.666	0.750	0.769	0.036
D6S292	9	10	0.729	0.820	0.781	0.472	D17S1880	10	8	0.805	0.830	0.846	0.199
D6S261	9	11	0.687	0.820	0.683	0.777	D17S787	10	9	0.783	0.800	0.810	0.646
D6S466	7	6	0.526	0.710	0.729	0.015	D17S949	8	8	0.735	0.800	0.796	0.865
D6S273	6	6	0.771	0.750	0.714	0.136	D17S784	8	7	0.815	0.770	0.798	0.290
D7S664	10	6	0.631	0.710	0.649	0.599	D18S53	10	10	0.702	0.790	0.786	0.014
D7S484	7	7	0.631	0.730	0.636	0.615	D18S70	10	10	0.763	0.830	0.833	0.058
D7S639	6	10	0.655	0.740	0.692	0.355							
D7S669	9	8	0.684	0.800	0.760	0.007							
D7S486	7	12	0.725	0.800	0.790	0.288							
D7S495	9	7	0.555	0.810	0.786	0.008							
D7S550	8	13	0.864	0.720	0.769	0.918							
D8S550	14	12	0.833	0.870	0.877	0.432	D19S210	6	6	0.578	0.730	0.772	0.163
D8S514	7	9	0.777	0.770	0.666	0.508	D19S413	10	8	0.789	0.780	0.803	0.697
D9S288	11	9	0.805	0.840	0.805	0.664	D20S107	8	10	0.729	0.780	0.714	0.471
D9S1845	11	15	0.714	0.830	0.735	0.262							
D9S153	6	7	0.605	0.760	0.717	0.012							
D9S164	7	7	0.750	0.800	0.783	0.784							
D10S249	10	8	0.656	0.740	0.788	0.174	D22S424	10	3	0.888	0.560	0.837	0.640
D10S191	12	12	0.864	0.800	0.874	0.789							
D10S193	9	9	0.837	0.800	0.767	0.980							
D10S535	5	6	0.676	0.780	0.497	0.225							
D10S212	12	12	0.621	0.700	0.750	0.221							
D11S922	12	19	0.933	0.920	0.862	0.902	DXS999	4	7	0.714	0.720	0.456	0.998
D11S1349	8	11	0.842	0.840	0.841	0.841	DXS1038	7	11	0.877	0.80	0.768	0.331
D11S903	9	6	0.729	0.740	0.783	0.094	DXS1275	10	10	0.91	0.81	0.841	0.400
D11S1332	9	8	0.810	0.760	0.734	0.977	DXS1205	9	9	0.95	0.73	0.748	0.202

[a]Values in bold are significant at the 5% level and those in bold and underlined are significant at 0.1% level.

Only three loci (D2S159, D4S2983, D5S408) showed significant deviations from HWE expectations (table 2) after correcting the *P* values for multiple loci testing using a Bonferroni correction (critical *P* value of 0.0005). Such deviations were associated in the three cases with a heterozygote deficiency as it could be seen from table 2 and as it had been checked with the appropriate test in GenePop. To further assess departures from HWE, we tested for association between all pairs of loci at the genotypic level using GenePop. No evidence for genotypic association was found at the 5% level (Bonferroni-corrected critical *P* value of 10^{-5}). Moreover, all of the pairwise tests had individual *P* values larger than 0.05. Regarding the chromosome 2 region, no evidence for linkage disequilibrium among the markers used was found (results not shown).

Discussion

Our study concerned the population of a small village from southern Tunisia which have typical features of traditional Arabo-islamic societies, characterized by a high rate of endogamy and consanguinity.

In our survey of BEH, we observed significant differences in the allele frequency distributions for 14 out of

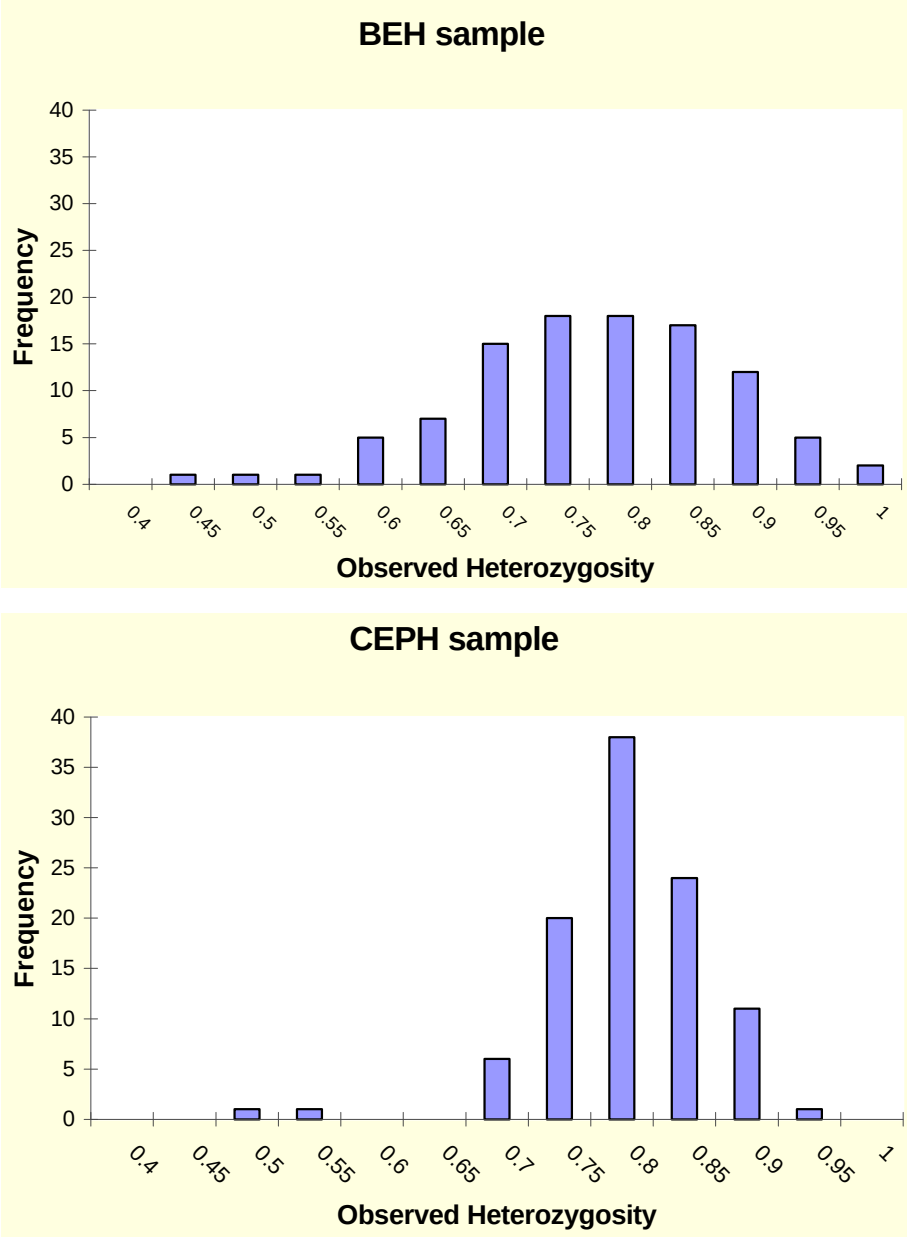

Figure 1. Histograms of distribution of observed heterozygosity of 103 STR markers in BirElHfai (BEH) sample from Tunisian population and CEPH sample from European population. Same scale was used for both histograms.

103 microsatellite markers (when compared with those of the CEPH database). We also observed an overall reduction in heterozygosity in the BEH population, as well as a significantly reduced observed versus expected heterozygosity at three of the loci (table 2). These observations could be simply explained by a high rate of consanguinity in BEH region estimated at 30% (N. Bougacha, personal communication). These data support our previous results which showed a lower polymorphisms of immunoglobulin variable heavy chain and T cell receptor beta chain genes in Tunisian population (Fakhfakh *et al*. 1997, 1999) and in accordance with those found by Morell *et al*. (1995) in Balinese population who showed a significant difference in 28 microsatellite markers out of 53 investigated ones. As shown in figure 1 there is a higher proportion of loci with smaller *H* values in our sample comparatively to CEPH (longer left-tail in the histogram). This reduction in heterozygosity in our study sample could be explained by the higher rate of inbreeding in BEH population.

The reduced heterozygosities were indicative of the deviation from HWE and were confirmed in a test of HWE at each locus. The deviation from HWE could be due to random drift in a small population and / or a high rate of endogamy, feature of isolated populations. The set of markers analysed here were used as a resource to map a gene predisposing to AITD diseases in BEH population. Indeed, we have performed a whole-genome scan in a large family affected with AITD originating from BEH village. We found a significant LOD score (LOD = 3.11 under recessive mode of inheritance) at marker D2S171 but with the LOD score droping dramatically around this marker (<1 at all surrounding STRs including D2S2168) (Maalej *et al*. 2001). Regarding linkage disequilibrium within the D2S2221–D2S365 region containing the linked marker, we found no evidence for LD between six markers spanning 4.3 cM. This was an indicative of the occurrence of many ancestral recombination events which would have reduced the size of LD blocks in the BEH population.

This results contrasted with the findings of Huttley GA *et al*. (1999) who identified the interval D2S2373–D2S2170 (4.3 cM), having three markers in common with our interval, as a high LD region in the European population. Moreover, the whole region considered spans 5.6 Mb and the smallest distance between markers is 431 kb which is a tenfold magnitude of the average LD block size reported in African population (about 30 kb, Gabriel *et al*. 2002). In fact, many studies have concluded that LD is extremely variable within and among loci and populations (Pritchard and Przeworski 2001). In order to assess the structure of LD within this region in BEH population, more markers (especially SNP) should be typed.

In addition, The STR allele frequencies could be taken as a basis for the STR genetic diversity in our population, useful in forensic identification and especially in paternity testing. In fact, the most polymorphic STRs markers could be investigated and used to improve the probability of paternity when compared to the results obtained with classical polymorphic genetic markers such as HLA to corroborate the anthropological study of the history and structure of the BEH village.

Acknowledgements

This work was supported by Grant from the Secretariat d'état à la recherche scientifique et de la Technologie from Tunisia and a Grant C2966 from International Foundation of Science. We would like to thank Dr Jose Osorio for his critical reading of the manuscript and Pr Rafik Bouaziz and his group for their help in creating the database containing all detailed information on the markers typed in study. This database is available online on the website: http://www.pmm.tn.refer.org/.

References

Aguirre A., Vicario A., Mazon L. I., Estomba A., Martinez de Pancorbo M., Arrieta Pico V., Perez Elortondo F., Lostao C. M. 1991 Are the Basques a single and a unique population?. *Am. J. Hum. Genet.* **49**(2), 450–458.

Bowcock A. M., Ruiz-Linares A., Tomfohrde J., Minch E., Kidd J. R., Cavalli-Sforza L. L. 1994 High resolution of human evolutionary trees with polymorphic microsatellites. *Nature*. **368**(6470), 455–457.

Budowle B., Giusti A. M., Waye J. S., Baechtel F. S., Fourney R. M., Adams D. E., Presley L. A., Deadman H. A., Monson K. L. 1991 Fixed-bin analysis for statistical evaluation of continuous distributions of allelic data from VNTR loci, for use in forensic comparisons. *Am. J. Hum. Genet.* **48**(5), 841–855.

Buetow K. H., Weber J. L., Ludwigsen S., Scherpbier-Heddema T., Duyk G. M., Sheffield V. C., Wang Z., Murray J. C. 1994 Integrated human genome-wide maps constructed using the CEPH reference panel. *Nat. Genet.* **6**(4), 391–393.

Chakraborty R., Ghosh A. K. 1981 Distribution of the number of heterozygous loci in the Kota population of Nilgiri hills of India and estimation of average heterozygosity. *Ann. Hum. Biol.* **8**(5), 453–459.

Chakraborty R., Smouse P. E., Neel J. V. 1988 Population amalgamation and genetic variation: observations on artificially agglomerated tribal populations of Central and South America. *Am. J. Hum. Genet.* **43**(5), 709–725.

Chakraborty R., Kidd K. K. 1991 The utility of DNA typing in forensic work. *Science*. **254**(5039), 1735–1739.

Deka R., Chakroborty R., Ferrell R. E. 1991 A population genetic study of six VNTR loci in three ethnically defined populations. *Genomics*. **11**(1), 83–92.

Deka R., De Croo S., Yu L. M. and Ferrell R. E. 1992 Variable number of tandem repeat (VNTR) polymorphism at locus D17S5 (YNZ22) in four ethnically defined human populations. *Hum. Genet.* **90** (1–2), 86–90.

Edwards A., Hammond H. A., Jin L., Caskey C. T., Chakraborty R. 1992 Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. *Genomics* **12**(2), 241–253.

Fakhfakh F., Ayadi H., Maalej A., Bahloul Z., Jarraya A., Zouali M. 1997 Polymorphism of the human Immunoglobulin heavy chain locus in rheumatoid arthritis. *Autoimmunity* **25**, 109–116.

Fakhfakh F., Maalej A., Makni H., Abid M., Jouida J., Zouali M. and Ayadi H. 1999 Analysis of Ig VH and TCR C*b* poly-

morphisms in a large family with thyroid autoimmune disorder. *Exp. Clin. Immunogenet* **16**, 185–191.

Fregeau C. J. and Fourney R. M. 1993 DNA typing with fluorescently tagged short tandem repeats: a sensitive and accurate approach to human identification. *Biotechniques* **15**(1), 100–119.

Gabriel S. B., Schaffner S. F., Nguyen H., Moore J. M., Roy J., Blumenstiel B., Higgins J., DeFelice M., Lochner A., Faggart M., Liu-Cordero S. N., Rotimi C., Adeyemo A., Cooper R., Ward R., Lander E. S., Daly M. J. and Altshuler D. 2002 The structure of haplotype blocks in the human genome. *Science* **296**, 2225–2229.

Gill P., Ivanov P. L., Kimpton C., Piercy R., Benson N., Tully G., Evett I., Hagelberg E. and Sullivan K. 1994 Identification of the remains of the Romanov family by DNA analysis. *Nat. Genet.* **6**(2), 130–135.

Guo S. W. and Thompson B. A. 1992 Performing exact test of Hardy-Weinberg proportions for multiple alleles. *Biometrics*. **48**, 361–372.

Gyapay G., Morissette J., Vignal A., Dib C., Fizames C., Millasseau P., Marc S., Bernardi G., Lathrop M., Weissenbach J. 1994 The 1993–94 Genethon human genetic linkage map. *Nat. Genet.* **7**(2), 246–339.

Hutley G. A., Smith M. W., Criington M. and O'Brien S. J. 1999 A scan of linkage disequilibrium across the human genome. *Genetics*. **152**, 1711–1722.

Lander E. S. 1989 DNA fingerprinting on trial. *Nature* **339**(6225), 501–505.

Lewis P. O. and Zaykin D. 2001 Genetic Data Analysis: Computer program for the analysis of allelic data. Version 1.0 (d16c). Free program distributed by the authors over the internet from http://lewis.eeb.uconn.edu/lewsihome/software.html

Lewontin R. C. and Hartl D. L. 1991 Population genetics in forensic DNA typing. *Science* **254**(5039), 1745–1750.

Maalej A., Makni H., Ayadi F., Bellassoued M., Jouida M., Jouida J., Bouguacha N., Abid M. and Ayadi H. 2001 A full genome screening in a large tunisian family affected with thyroid autoimmune disorders. *Genes. Immun.* **2**(2), 71–75.

Matise T. C., Perlin M. and Chakravarti A. 1994 Automated construction of genetic linkage maps using an expert system (MultiMap): a human genome linkage map. *Nat. Genet.* **6**(4), 384–390.

Morell R., Liang Y., Asher J. H. Jr., Weber J. L., Hinnant J. T., Winata S., Arhya I. N. and Friedman T. B. 1995 Analysis of short tandem repeat (STR) allele frequency distributions in a Balinese population. *Hum. Mol. Genet.* **4**(1), 85–91.

Ott J. 1999 *Analysis of human Genetic linkage, Third Edition*. Johns Hopkins University Press, Baltimore, pp. 382.

Pena S. D. and Chakraborty R. 1994 Paternity testing in the DNA era. *Trends. Genet.* **10**(6), 204–209.

Pritchard J. K. and Przeworski M. 2001 Linkage disequilibrium in human: models and data. *Am. J. Hum. Genet.* **69**, 1–14.

Raymond M. and Rousset F. 1995 GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *J. Heredity*. **86**, 248–249.

Risch N. J. and Devlin B. 1992 On the probability of matching DNA fingerprints. *Science* **255**(5045), 717–720.

Wall W. J., Williamson R., Petrou M., Papaioannou D. and Parkin B. H. 1993 Variation of short tandem repeats within and between populations. *Hum. Mol. Genet.* **2**(7), 1023–1029.

Weber J. L. and May P. E. 1989 Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am. J. Hum. Genet.* **44**(3), 388–396.

Weir B. S. 1996 *Genetic Data Analysis* II. Sinauer Assoc. Sunderland, MA.

Received 9 June 2003