 Human Molecular Genetics, 1998, Vol. 7, No. 3 399–406

The Δ*ccr5* mutation conferring protection against HIV-1 in Caucasian populations has a single and recent origin in Northeastern Europe

Frédérick Libert[1,+], Pascale Cochaux[2,+], Gunhild Beckman[3], Michel Samson[1], Marina Aksenova[4], Antonio Cao[5], Andrew Czeizel[6], Mireille Claustres[7], Concepción de la Rúa[8], Maurizio Ferrari[9], Claude Ferrec[10], Guillermo Glover[11], Bjorn Grinde[12], Sefik Güran[13], Vaidutis Kucinskas[14], Joao Lavinha[15], Bernard Mercier[10], Gönül Ogur[13], Leena Peltonen[16], Cristina Rosatelli[5], Marianne Schwartz[17], Victor Spitsyn[18], Laszlo Timar[6], Lars Beckman[3], Marc Parmentier[1,*] and Gilbert Vassart[1,2]

[1]IRIBHN and [2]Service de Génétique Médicale, Université Libre de Bruxelles, Campus Erasme, B-1070 Brussels, Belgium, [3]Department of Medical Genetics, Umea University, S-901 85 Umea, Sweden, [4]Department of Molecular Diagnostics and Genomic Fingerprintings, National Research Center 'GosNII Genetika', 113 545 Moscow, Russia, [5]Istituto di Clinica e Biologia dell'eta Evolutiva, Universita degli Studi di Cagliari, I-09121, Cagliari, Italy, [6]Department of Human Genetics and Teratology, 'Bela Johan' National Institute of Public Health, H-1966, Budapest, Hungary, [7]Laboratoire de Biochimie Génétique, Institut de Biologie, 34060 Montpellier, France, [8]Departamento Biologia Animal y Genetica, Facultad de Ciencias, Universidad del Pais Vasco, 48080 Bilbao, Spain, [9]Laboratorio Centrale, H. San Raffaele, 20132 Milano, Italy, [10]Laboratoire de Biogénétique, 29275 Brest, France, [11]Centro de Bioquimica y Genetica Clinica, Unidad de Genetica Molecular, 30100 Espinardo, Spain, [12]Department of Virology, National Institute of Public Health, N-0403 Oslo, Norway, [13]Department of Medical Genetics, Gata Medical Faculty, Ankara, Turkey, [14]Center of Human Genetics, Vilnius University, Vilnius, Lithuania, [15]Instituto Nacional de Saude, 1699 Lisboa, Portugal, [16]Department of Human Molecular Genetics, National Public Health Institute, FIN-00300 Helsinki, Finland, [17]Molecular Genetics Laboratory, Juliane Marie Centret, Rigshospitalet, DK-2100 Copenhagen, Denmark and [18]Research Center for Medical Genetics, Academy of Medical Sciences, Moscow, Russia

Received August 21, 1997; Revised and Accepted December 8, 1997

The chemokine receptor CCR5 is encoded by the *CMKBR5* gene located on the p21.3 region of human chromosome 3, and constitutes the major co-receptor for the macrophage-tropic strains of HIV-1. A mutant allele of the CCR5 gene, Δ*ccr5*, was shown to provide to homozygotes with a strong resistance against infection by HIV. The frequency of the Δ*ccr5* allele was investigated in 18 European populations. A North to South gradient was found, with the highest allele frequencies in Finnish and Mordvinian populations (16%), and the lowest in Sardinia (4%). Highly polymorphic microsatellites (IRI3.1, *D3S4579* and IRI3.2, *D3S4580*) located respectively 11 kb upstream and 68 kb downstream of the CCR5 gene deletion were used to determine the haplotype of the chromosomes carrying the Δ*ccr5* variant. A strong linkage disequilibrium was found between Δ*ccr5* and specific alleles of the IRI3.1 and IRI3.2 microsatellites: >95% of the Δ*ccr5* chromosomes carried the IRI3.1-0 allele, while 88% carried the IRI3.2-0 allele. These alleles were found respectively in only 2 or 1.5% of the chromosomes carrying a wild-type CCR5 gene. From these data, it was inferred that most, if not all Δ*ccr5* alleles originate from a single mutation event, and that this mutation event probably took place a few thousand years ago in Northeastern Europe. The high frequency of the Δ*ccr5* allele in Caucasian populations cannot be explained easily by random genetic drift, suggesting that a selection advantage is or has been associated with homo- or heterozygous carriers of the Δ*ccr5* allele.

INTRODUCTION

Members of the chemokine receptor gene family recently have been identified as the co-receptors which allow entry of human immunodeficiency virus type 1 (HIV-1) into human CD4-positive cells (1–6). Chemokine receptors belong to the superfamily of

*To whom correspondence should be addressed Tel: +32 2 555 4171; Fax: +32 2 555 4655; Email: mparment@ulb.ac.be
+These authors contributed equally to this work

serpentine G protein-coupled receptors; they are classified as CXC- (or α-) or CC- (or β-) chemokine receptors, according to the primary structure of their natural agonists (7). After the landmark identification of CXCR4 as the co-receptor for infection by T-tropic HIV-1 strains (those strains capable of infecting primary T lymphocytes and T cells adapted to continuous culture) (1), CCR5 (8) was shown to act as the major co-receptor for M-tropic HIV-1 strains (2–6,9). By contrast to T-tropic strains, which generally appear at late stages of the disease, M-tropic strains infect monocytes/macrophages and have been shown to be responsible for viral transmission (10–12). The gene encoding CCR5 (*CMKBR5*) is located in the p21.3 region of human chromosome 3, within a cluster including most of the other CC-chemokine receptor genes (13).

Resistance to infection by HIV-1 has been described in individuals who remain seronegative despite repeated exposure to the virus (14). The mechanism of resistance in these 'exposed-uninfected' individuals is certainly not unique, but one convincing genetic explanation has been provided by the identification in individuals of Caucasian origin of a null mutation of CCR5 (15–17). Homozygotes for a 32 bp deletion in the gene segment encoding the second extracellular loop of CCR5 (Δ*ccr5*) appear to resist infection. This conclusion is based on the observation (i) of a high frequency of these natural Δ*ccr5*/Δ*ccr5* knock-outs in cohorts of exposed-uninfected individuals (16–18), (ii) the virtual absence of Δ*ccr5* homozygotes in large cohorts of seropositive individuals (15,17,19) although rare exceptions have been published to date (20–22) and (iii) on the observation that white blood cells from Δ*ccr5* homozygotes are highly resistant to infection, *in vitro*, by M-tropic HIV-1 strains (15,16,18,20,23). Heterozygotes for the Δ*ccr5* allele exhibit slower progression into the clinical stages of AIDS (19,24–26), and appear partially protected against HIV infection (15,26). This mutant allele was not found in populations from Black Africa and the Far East (15,27), but other rare mutations affecting the CCR5 gene have been described in all populations (28).

CCR5 is expressed on monocytes, macrophages, memory T cells and dendritic cells, and is believed to play an important role in the chemokinesis of these cell populations *in vitro* and *in vivo*, in response to RANTES, MIP-1α and MIP-1β (29,30). CCR5 is also expressed in the microglial cells, in correlation with the ability of HIV-1 to infect the central nervous system (31), but no function has so far been attributed to the receptor in this location. Δ*ccr5* homozygotes are apparently healthy, suggesting that the normal function(s) of the receptor can be taken over efficiently by other chemokine receptors and their ligands. A recent report suggests, however, that Δ*ccr5* heterozygotes are more susceptible to opportunistic infections during the late stages of AIDS progression (32).

Table 1. Distribution of CCR5 genotypes and frequency of the Δ*ccr5* allele in European populations

Population	CCR5 genotype (*n*) CCR5/CCR5	CCR5/Δ*ccr5*	Δ*ccr5*/Δ*ccr5*	Δ*ccr5* allele (frequency)	χ^2HW	*n*
Mordvinians	58	28	0	0.163	3.250	86
Finns	67	31	0	0.158	3.470	98
Russians	61	21	1	0.139	0.290	83
Lithuanians	220	61	2	0.115	1.030	283
Swedish Saamis	101	18	1	0.083	0.030	120
Swedes	152	46	6	0.142	1.160	204
Norwegians	79	21	0	0.105	1.380	100
Danes	78	22	0	0.110	1.530	100
French (Brittany)	79	20	1	0.110	0.050	100
French (Montpellier)	79	20	0	0.101	1.250	99
Belgians	582	114	8	0.092	0.810	704
Spaniards (Murcia)	81	19	0	0.095	1.100	100
Basques	79	9	1	0.062	1.470	89
Portugese	88	13	0	0.064	0.460	101
Hungarians	84	13	2	0.086	2.640	99
Italians (Milano)	82	15	1	0.087	0.110	98
Sardinians	92	8	0	0.040	0.170	100
Turks	91	13	0	0.063	0.450	104
Total	2153	492	23	0.101		2668

The CCR5 genotype was determined by PCR for at least 80 individuals originating from each of 18 different locations spread all over Europe, and the frequency of the Δ*ccr5* allele was calculated. The concordance of the observed genotype frequencies with the Hardy–Weinberg equilibrium was determined by a chi square test (χ^2HW).

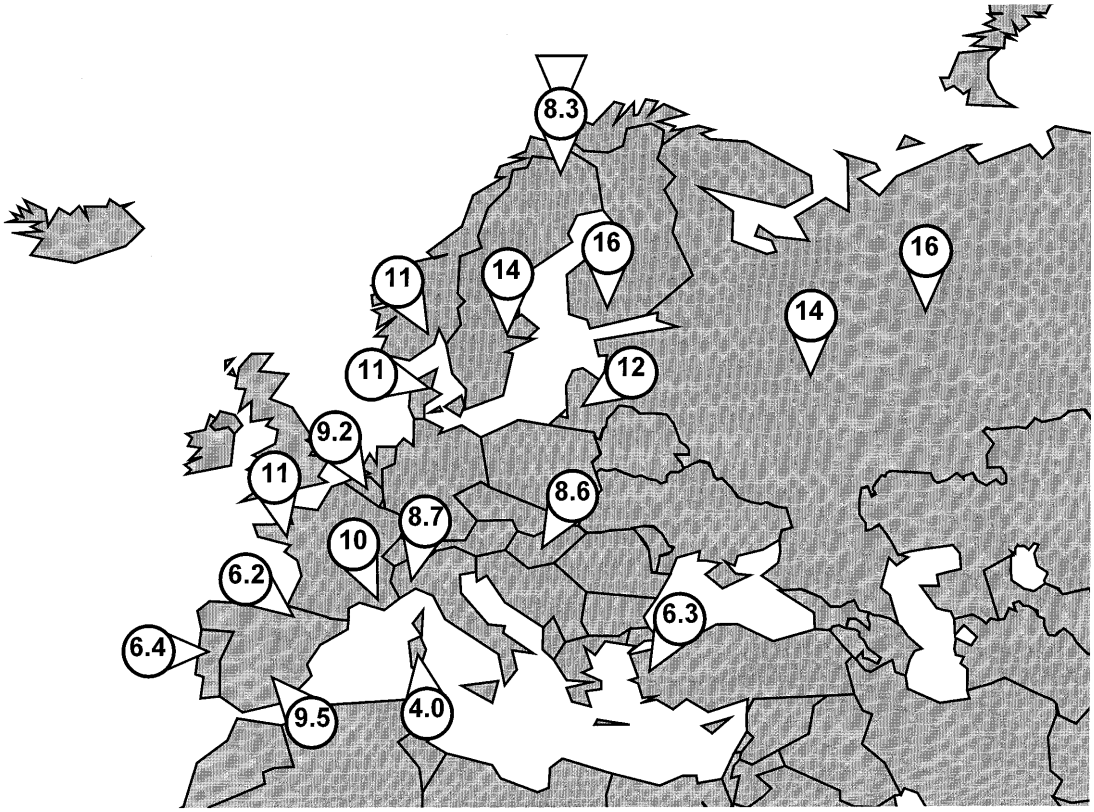

Figure 1. Frequency of the Δ*ccr5* allele in European populations. Schematic map of Europe indicating the frequency (in percentage) of the CCR5 mutant allele Δ*ccr5* in the various populations tested in this study. The samples originated from the regions of Oslo (Norway), Umea and Lapland (Sweden), Helsinki (Finland), Saransk (Mordvinia), Moscow (Russia), Copenhagen (Denmark), Vilnius (Lithuania), Brest and Montpellier (France), Brussels (Belgium), Milan and Cagliari (Italy), Budapest (Hungary), Lisbon (Portugal), Espinardo and Bilbao (Spain) and Ankara (Turkey).

The high frequency of the Δ*ccr5* allele in some populations raises the question of whether it is the result of random genetic drift, or rather the consequence of selective pressures possibly driven by an increased resistance to an infectious agent related or not to HIV. The aim of the present study was to determine more precisely the distribution of the Δ*ccr5* allele in populations of 'Caucasian' descent and to explore the possibility that the mutation has emerged more than once. By genotyping >2500 individuals from 18 European populations, we found the highest Δ*ccr5* allele frequency in Northeastern Europe. Analysis of tightly linked polymorphic microsatellites demonstrated strong linkage disequilibrium. Our observations are compatible with a single and relatively recent origin of the Δ*ccr5* allele, coming most probably from the Finno-Ugrian population, and could suggest a selective pressure in favor of the mutant allele.

RESULTS

Geographic distribution of the Δ*ccr5* allele

The CCR5 genotype was determined by polymerase chain reaction (PCR) in population samples from 18 regions covering most of Europe. The mutant Δ*ccr5* allele was found in all investigated populations (Table 1 and Fig. 1), but a significant heterogeneity was observed among the population samples ($\chi^2 = 47.8$; 17 d.f.; $P = 0.00013$). A general North to South downhill gradient was observed, with only the Lapps (Saamis) constituting a notable exception. The highest frequencies were found in Northern and Northeastern Europe (16% in Finnish and Mordvinian populations). The Swedish Saamis had a significantly lower frequency (8.3%) than Swedes ($P = 0.024$). We found lower frequencies in the Mediterranean area, with the lowest observed figures in Sardinians (4%). No significant deviation from the Hardy–Weinberg equilibrium was observed, when considering the whole material, or the individual populations (Table 1).

Characterization of the IRI3.1 (*D3S4579*) microsatellite

A TG repeat microsatellite located 11 063 bp upstream from the Δ*ccr5* deletion site was identified by sequencing the 17 kb separating the genes encoding CCR2 and CCR5 (8) and in the 140 kb genomic sequence of a bacterial artificial chromosome (BAC), which was released recently in the databases (33). The polymorphism of this microsatellite was analyzed in 290 individuals belonging to different populations (Tables 1 , 2 and 3). Most of the homozygotes for the Δ*ccr5* allele were tested, as well as a large number of heterozygotes. Homozygotes for the wild-type CCR5 gene were selected mainly from four populations from Northern (Sweden and Finland) and Southern (Sardinia and Turkey) Europe, in order to investigate potential variations in allele distributions. The IRI3.1 microsatellite displayed extensive polymorphism (Fig. 2 and Table 2): 13 alleles were observed, with an average heterozygosity of 85%. Among the chromosomes bearing the wild-type CCR5 gene, the major alleles were IRI3.1+3

(average frequency of 0.21), IRI3.1+4 (0.22), IRI3.1+5 (0.17) and IRI3.1+6 (0.22). The allele distribution (Table 2) was found to be significantly different between Northern (Sweden and Finland) and Southern (Sardinia and Turkey) populations (χ^2 = 37.7, 7 d.f., P <0.00001). This difference was essentially due to the lower frequency of the IRI3.1+4 allele (χ^2 = 21.9, 1 d.f., P <0.00001) and the somewhat higher frequency of long repeats (IRI3.1+8 and up, χ^2 = 6.5, 1 d.f., P = 0.01) in Southern European populations.

Haplotype determination

Out of the 16 Δ*ccr5* homozygotes that were tested for the microsatellite, 15 displayed the same genotype, being homozygous for the IRI3.1-0 allele (Fig. 2 and Table 3). The other alleles were named relative to this allele. The single exception harbors the IRI3.1-0 allele together with an IRI3.1+4 allele. From the data collected from CCR5/CCR5 and Δ*ccr5*/Δ*ccr5* individuals (Table 3), the linkage disequilibrium between CCR5 and IRI3.1 was extremely significant (χ^2 = 286, 1 d.b., P <<0.00001). The IRI3.1-0 allele was seldom found in association with the wild-type CCR5 allele, being one of the rarer alleles among CCR5/CCR5 individuals (average frequency of 0.02). We then investigated a total of 98 Δ*ccr5* heterozygotes, originating mostly from the same cohorts as the CCR5/CCR5 homozygotes. Out of these, only five did not display an IRI3.1-0 allele (Table 3). Four of these exceptions had in common an IRI3.1+1 allele, differing by a single doublet from the IRI3.1-0 allele. It should be noted that the IRI3.1+1 allele is also a rare allele on chromosomes bearing a wild-type CCR5 gene, with an average frequency of 0.006 (Table 2). The last exception was heterozygous for the IRI3.1+5 and IRI3.1+8 alleles. Among heterozygotes for Δ*ccr5* and IRI3.1-0, the probability of having the IRI3.1-0 allele associated with CCR5 and another allele with Δ*ccr5* is low (<0.001) and can be neglected. We can therefore consider that among the 98 Δ*ccr5* heterozygotes, the IRI3.1-0 allele, when present, is associated with Δ*ccr5*. Taking this into account, we found that, out of 130 chromosomes carrying Δ*ccr5*, 124 are associated with IRI3.1-0 (95.4%), while six Δ*ccr5* alleles [95% confidence interval (c.i.): 2.20–13.6 considering a Poisson distribution] belong to other haplotypes (4.4%, 1.7–10.5%).

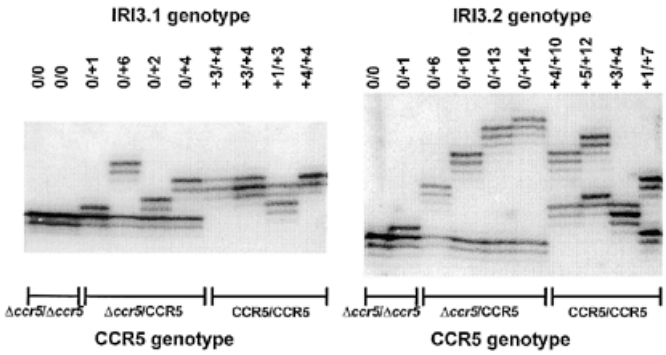

Figure 2. PCR amplification of the IRI3.1 and IRI3.2 microsatellites. Genomic fragments of ~150 bp containing the IRI3.1 or IRI3.2 microsatellites were amplified by PCR and separated on a denaturing polyacrylamide gel. A set of samples representative of the three CCR5 genotypes (CCR5/CCR5, CCR5/Δ*ccr5* and Δ*ccr5*/Δ*ccr5*) are shown. The microsatellite alleles are identified for each individual by the number of dinucleotide repeats, relative to IRI3.1-0 and IRI3.2-0.

Table 2. Observed frequency of IRI3.1 alleles in CCR5/CCR5 homozygotes from different populations

IRI3.1	N. Eur.		Sweden		Finland		S. Eur.		Sardinia		Turkey	
allele	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%
–2	0	0.0	0	0.0	0	0.0	1	0.6	0	0.0	1	1.3
0	3	1.9	0	0.0	3	3.8	0	0.0	0	0.0	0	0.0
+1	1	0.6	0	0.0	1	1.3	1	0.6	1	1.3	0	0.0
+2	10	6.3	3	3.8	7	8.8	2	1.3	0	0.0	2	2.6
+3	30	19.0	16	20.5	14	17.5	39	25.3	24	31.6	15	19.2
+4	49	31.0	27	34.6	22	27.5	14	9.1	5	6.6	9	11.5
+5	20	12.7	10	12.8	10	12.5	33	21.4	15	19.7	18	23.1
+6	34	21.5	16	20.5	18	22.5	39	25.3	20	26.3	19	24.4
+7	8	5.1	4	5.1	4	5.0	11	7.1	4	5.3	7	9.0
+8	2	1.3	1	1.3	1	1.3	7	4.5	1	1.3	6	7.7
+9	1	0.6	1	1.3	0	0.0	4	2.6	3	3.9	1	1.3
+10	0	0.0	0	0.0	0	0.0	1	0.6	1	1.3	0	0.0
+11	0	0.0	0	0.0	0	0.0	2	1.3	2	2.6	0	0.0
Total	158	100	78	100	80	100	154	100	76	100	78	100

The alleles of the IRI3.1 microsatellite located 11 kb upstream of the CCR5 gene deletion were determined by PCR and gel electrophoresis. Alleles are identified following the relative number of dinucleotide repeats. Allele distributions and frequencies are given for individuals homozygous for the wild-type CCR5 allele, in two regions of Northern Europe (N. Eur., Sweden and Finland) and two regions of Southern Europe (S. Eur., Sardinia and Turkey). A total of 75–80 alleles were analyzed in each region.

Table 3. Frequency of IRI3.1 alleles in individuals with different CCR5 genotypes

IRI3.1 Allele	CCR5/CCR5		CCR5/Δ*ccr5*		Δ*ccr5*/Δ*ccr5*		Overall
	n	%	*n*	%	*n*	%	%
–2	1	0.3	0	0.0	0	0.0	0.3
0	7	2.0	96	49.0	31	96.9	11.5
+1	2	0.6	7	3.6	0	0.0	0.5
+2	16	4.5	2	1.0	0	0.0	4.1
+3	75	21.3	21	10.7	0	0.0	19.2
+4	79	22.4	27	13.8	1	3.1	20.5
+5	58	16.5	14	7.1	0	0.0	14.8
+6	78	22.2	18	9.2	0	0.0	19.9
+7	19	5.4	4	2.0	0	0.0	4.9
+8	9	2.6	5	2.6	0	0.0	2.3
+9	5	1.4	1	0.5	0	0.0	1.3
+10	1	0.3	1	0.5	0	0.0	0.3
+11	2	0.6	0	0.0	0	0.0	0.5
Total	352	100	196	100	32	100	100

The overall frequencies were estimated considering an average frequency of 10% for the Δ*ccr5* allele.

These other haplotypes could originate from three main causes: (i) independent mutation events occurring on a different genetic background; (ii) crossing-over disrupting the pre-existing association between Δ*ccr5* and IRI3.1-0; and (iii) modification of the number of microsatellite repeats by replicative slippage. Microsatellite mutation can most probably be held responsible for the four IRI3.1+1 alleles found in heterozygotes, this allele being rare among CCR5/CCR5 homozygotes. This is in agreement with the reported observation that most of the mutations affecting microsatellites involve the gain or loss of a single repeat unit (34), and that mutations are biased in favor of longer allele lengths (35). If we neglect independent mutations (they can, however, not be ruled out), we end up with four (1.09–10.24; c.i. 95%) chromosomes resulting from microsatellite mutation events and two (0.24–7.22; c.i. 95%) chromosomes resulting from recombination events out of 130 chromosomes.

Approximate dating of the Δ*ccr5* mutation

The rates of crossing-over between IRI3.1 and Δ*ccr5*, and of microsatellite mutation can both be used to estimate grossly how long this haplotype has been present in European population history, using the Luria and Delbruck's equations as described by Hästbacka *et al*. (36). Considering the distance of 11 063 bp between the CCR5 gene deletion and the IRI3.1 microsatellite, and an approximate recombination rate of 1% per megabase and generation for the human genome, the recombination frequency (θ) was estimated to be 0.0111% per generation. Based on the Luria–Delbrück equations (36), we obtained a number of ~140 generations (17–516; c.i. 95%) since the initial mutation leading to Δ*ccr5*. Considering an average of 25 years per generation, this places the mutation event some 3500 (400–13 000; c.i. 95%) years ago.

The same calculation was made, using the frequency of microsatellite mutations, and the reported average mutation rate (0.056%) observed for microsatellites in CEPH families (34). If we consider that the four IRI3.1+1 alleles found in Δ*ccr5* heterozygotes result from events of repeat amplifications (3.08; 0.84–7.9%; c.i. 95%), the number of generations since the Δ*ccr5* mutation was estimated to be 56 (15–147; c.i. 95%), corresponding to 1400 years (375–3675; c.i. 95%).

Table 4. Frequency of IRI3.2 alleles in individuals with different CCR5 genotypes

IRI3.2 Allele	CCR5/CCR5		CCR5/Δ*ccr5*		Δ*ccr5*/Δ*ccr5*		Overall
	n	%	*n*	%	n	%	%
–1	2	0.7	0	0.0	0	0.0	0.7
0	4	1.5	103	52.6	23	88.5	10.2
+1	14	5.1	8	4.1	1	3.8	5.0
+2	6	2.2	1	0.5	0	0.0	2.0
+3	4	1.5	6	3.1	0	0.0	1.3
+4	29	10.7	7	3.6	0	0.0	9.6
+5	64	23.5	18	9.2	2	7.7	21.9
+6	65	23.9	20	10.2	0	0.0	21.5
+7	20	7.4	8	4.1	0	0.0	6.6
+8	4	1.5	3	1.5	0	0.0	1.3
+9	5	1.8	5	2.6	0	0.0	1.7
+10	9	3.3	7	3.6	0	0.0	3.0
+11	10	3.7	0	0.0	0	0.0	3.3
+12	20	7.4	4	2.0	0	0.0	6.6
+13	10	3.7	2	1.0	0	0.0	3.3
+14	6	2.2	4	2.0	0	0.0	2.0
Total	272	100	196	100	26	100	100

The overall frequencies were estimated considering an average frequency of 10% for the Δ*ccr5* allele.

Analysis of a second microsatellite (IRI3.2, *D3S4580*)

In order to confirm the observations, we analyzed a second microsatellite (GA repeat adjacent to a GT repeat) included in the BAC covering the CCR5 gene (33), but located 68 kb downstream of the Δ*ccr5* deletion. A total of 247 individuals (most of the sampling tested for IRI3.1) from different European populations were tested, including 136 CCR5/CCR5 homozygotes, 98 CCR5/Δ*ccr5* heterozygotes and 13 Δ*ccr5*/Δ*ccr5* homozygotes. The IRI3.2 microsatellite was also found to be highly polymorphic, with 16 different alleles and 72% heterozygozity. Among CCR5/CCR5 homozygotes, the most frequent alleles were IRI3.2+4 to +7 and IRI3.2+12 (Table 4). No significant difference was found for allele distribution among European populations (not shown). When analyzing Δ*ccr5*/Δ*ccr5* homozygotes, a strong linkage disequilibrium was found, the Δ*ccr5* allele being associated with the IRI3.2-0 allele in 23 chromosomes out of 26 (88%), while the IRI3.2-0 allele only represents 1.5% of the alleles associated with wild-type CCR5 (Table 4, $P << 0.0001$). The three exceptions were an IRI3.2+1 allele (most likely resulting from microsatellite mutation), and two alleles of greater

size (IRI3.2+5), presumably resulting from cross-overs. All Δ*ccr5* alleles were associated with either IRI3.1-0 or IRI3.2-0. As expected, heterozygotes exhibited intermediate frequencies. Ten heterozygotes out of 98 did not bear an IRI3.2-0 allele, and half of these contained an IRI3.2+1 allele. Taking into account the corrections for the expected frequency of cross-overs giving rise to an association of Δ*ccr5* with IRI3.2+1 alleles, we estimated that out of 123 analyzed chromosomes bearing Δ*ccr5*, 5.6 (1.9–12.6, c.i. 95%) were the result of microsatellite mutation and 7.4 (2.9–15.1, c.i. 95%) were the result of cross-overs between CCR5 and IRI3.2. The number of generations since the origin of the Δ*ccr5* mutation was estimated as 83 (28–192, c.i. 95%) on the basis of the microsatellite mutations, corresponding to 2100 years (700–4800, c.i. 95%). Using the recombination events, figures of 90 generations (35–190, c.i. 95%) and 2250 years (900–4700, c.i. 95%) were obtained.

DISCUSSION

The Δ*ccr5* allele frequency is higher in Northern Europe

The Δ*ccr5* allele was found all across Europe, but a difference in allele frequencies was found among regions, with a North to South downhill gradient. This gradient is largely in agreement with the results of Martinson *et al*. (27) who studied the Δ*ccr5* allele in European and non-European countries, but a larger number of regions was investigated in the present work. In discussing their results, Martinson *et al*. stated that the distribution of the Δ*ccr5* allele was similar to that of some other Europe-centered genes, namely the ΔF508 cystic fibrosis mutation and the hemochromatosis gene. We find, however, a clear difference in the distribution of the Δ*ccr5* allele on one hand and of the ΔF508 or hemochromatosis genes on the other. The latter two genes have a West European 'gene center' and low frequencies in eastern Europe (37,38), whereas high frequencies of the Δ*ccr5* allele are found in eastern Europe. The fact that the highest frequencies of Δ*ccr5* were found in Mordvinians (16.3%) and Finns (15.8%) raises the question of whether the Δ*ccr5* mutation originated somewhere among the Finno-Ugrian tribes of Russia. The relatively modest frequencies in two other Finno-Ugrian populations, Saamis and Hungarians, would seem to be in disagreement with such a hypothesis. However, Saamis are known to differ genetically from other Finno-Ugrian people (39), and present-day Hungarians should have received a considerable admixture from other populations, which is expected to lower their original Δ*ccr5* frequencies. The location of the Δ*ccr5* gene center in Northern European countries could explain the high frequency found by Martinson *et al*. (27) for Ashkenazi Jews, who have lived for centuries in these regions, and for whom a founder effect has been postulated for a number of genetic diseases (40).

Δ*ccr5* alleles originate from a single and recent mutation event

The existence of a 10 bp direct repeat, which flanks the 32 bp deletion of Δ*ccr5* (15), could suggest that the Δ*ccr5* mutation has possibly occurred more than once in Europe. This possibility is made extremely unlikely by the strong linkage disequilibrium which the Δ*ccr5* allele displays with the IRI3.1-0 and IRI3.2-0 alleles of microsatellites flanking the CCR5 gene on both sides, combined with the fact that both IRI3.1-0 and IRI3.2-0 alleles are rare (0.02 and 0.015 respectively) in the absence of CCR5 deletion (Table 2). It appears, therefore, that most if not all Δ*ccr5* chromosomes originate from a single mutation event. The fact that all Δ*ccr5* alleles were associated either with IRI3.1-0 or IRI3.2-0 further suggests that events of crossing-over or microsatellite mutation occurred independently on an IRI3.1-0/Δ*ccr5*/IRI3.2-0 original haplotype. It cannot be excluded that other mutagenesis events contributed to a very small proportion of the Δ*ccr5* alleles, but this would not change the conclusions of the work.

Considering the observed frequencies for recombination between the Δ*ccr5* allele and the microsatellites, and the observed frequencies for microsatellite mutations, we estimated the number of generations required since the original CCR5 mutation in order to obtain the present day frequencies, using the formula of Luria and Delbruck, as described (36). From the recombination frequency between CCR5 and IRI3.1, the Δ*ccr5* mutation appears to be relatively recent (some 3400 years old), while from the microsatellite mutation rate, it appears even more recent (some 1400 years ago). Similar analysis for the IRI3.2 microsatellite gave ~2000–2200 years, starting from mutation or crossing-over events. These calculations make a number of assumptions that are certainly not met (such as uniform exponential growth of the population and panmixis), and are prone to high error rates (due, among other factors, to the relatively low number of observations in the case of the IRI3.1 marker). Also, microsatellites have different evolution rates, and we have considered an average rate for our calculation. Nevertheless, it is likely that the Δ*ccr5* mutation is relatively recent, and took place within the Neolithic period, a few thousand years ago. This is consistent with the fact that the Δ*ccr5* mutation is not found in African populations nor in most Asian populations. European and Asian populations are considered to have separated ~35 000 years ago, while their common Middle East ancestors separated from African populations some 120 000 years ago (41).

Possible existence of a selection pressure in favor of the Δ*ccr5* allele

The high frequency of the Δ*ccr5* allele in European populations, starting from a single mutation event, could in theory result from random genetic drift, in the absence of selection against or in favor of the mutation. However, given the relatively recent origin of the mutation, and the size of the population in Europe during the Neolithic period (steady increase from a few hundred thousand to 400 million), random genetic drift is unlikely to account for the rapid increase in Δ*ccr5* allele frequency. A positive selection is therefore postulated, and selections acting primarily on heterozygotes are more likely to drive a new mutation to an average level of 10% allele frequency within a few hundred generations or less. Potential selection advantages could be associated with resistance to infectious diseases or their consequences, in accordance with the involvement of CCR5 in the recruitment of monocytes, macrophages and helper T cells. The precise nature of the putative selective factor(s), however, remains to be identified.

Conclusions and perspectives

Together, the geographical distribution of the Δ*ccr5* allele and the large excess of the IRI3.1-0/Δ*ccr5*/IRI3.2-0 haplotype strongly suggests a single emergence of the Δ*ccr5* mutation, taking place most probably in a Finno-Ugrian tribe of Russia. The high frequencies of the Δ*ccr5* allele in some populations could affect the epidemiology of the disease and the mean progression rate of

HIV-positive individuals in these populations. It could also favor in these populations the emergence of strains that can bypass the CCR5 defect by using other co-receptors efficiently (2,6,42). The possibility that a selection pressure has allowed the rapid expansion of the Δ*ccr5* allele in European populations will require further work in order to confirm the hypothesis, and to investigate the nature of the potential selection factors.

MATERIALS AND METHODS

DNA samples

Anonymous DNA samples were collected from a total of 2668 unrelated individuals of 18 different European populations representing most major ethno-linguistic groups (Baltic, Finno-Ugrian, Germanic, Romanic, Slavic) and including Basques, Saamis and Sardinians. In previous population studies, these groups have been found to show distinct and peculiar gene frequencies. The DNA was prepared from blood leukocytes by standard methods (43) and analyzed in Brussels.

Detection of the Δ*ccr5* allele

The 32 bp deletion in the CCR5 gene was detected by electrophoresis through a 3% agarose gel of a 735 bp fragment amplified by PCR from genomic DNA (forward primer, 5′CCTGGCTGTCGTCCATGCTG3′; reverse primer, 5′AGCCATGTGCACAACTCT3′; 30 cycles; annealing 60°C for 1 min; elongation 72°C for 1 min; denaturation 93°C for 1 min). To increase separation of the wild-type and deleted alleles, the 735 bp fragment was cleaved with *Eco*RI into a 332 bp constant fragment and a 403 or 371 bp fragment corresponding to the wild-type or Δ*ccr5* fragment, respectively.

Microsatellite analysis

A TG dinucleotide microsatellite was identified 11 063 bp upstream from the CCR5 gene deletion (33). It was analyzed by PCR, using AAGAGATTGGTTCCAGGCATG as forward primer, CCGGACCTTGCATTACAGGAC as reverse primer, and the following conditions: annealing 63°C for 1 min; elongation 72°C for 1 min; denaturation 94°C for 1 min; 0.75 mM $MgCl_2$; 125 ng each primers; 200 µM dCTP, dGTP and dTTP, 100 µM dATP and 1 µCi [α-^{32}P]dATP. Alternatively, PCR was performed using an α-^{32}P-labeled forward primer, in the absence of labeled nucleotides. A second microsatellite (GA repeat followed by GT repeat) was analyzed similarly (labeled forward primer, CCTTCTGGAGCAGCACTTCCA, reverse primer, GTAAATCTCCTAACAACATGC; annealing 58°C for 1 min; elongation 72°C for 1 min; denaturation 94°C for 1 min; 1.5 mM $MgCl_2$; 125 ng each primers; 200 µM dATP, dCTP, dGTP and dTTP). PCR products were analyzed by electrophoresis through a 6% sequencing polyacrylamide gel containing 35% formamide and 5.6 M urea and radioautography.

Statistics

Allele frequencies were calculated, and the fit to the Hardy–Weinberg equilibrium and the heterogeneity between population samples were evaluated by means of the χ^2 test, using Instat software. In several instances, the data from the tables were pooled for different alleles, in order to avoid numbers <5, and/or to reduce the number of degrees of freedom. As an example, when comparing IRI3.1 alleles between populations from Northern and Southern Europe, alleles IRI3.1-2 to +1 were pooled as well as alleles IRI3.1+8 to +11.

ACKNOWLEDGEMENTS

We thank M. Georges for fruitful discussion. This study was supported by the French Agence Nationale de Recherche sur le SIDA and the Belgian Programme on University Poles of Attraction initiated by the Belgian State, Prime Minister's office, Service for Sciences, Technology and Culture. Also supported by grants from the Fonds de la Recherche Scientifique Médicale, the Fonds Médical Reine Elisabeth, Télévie, the European Union (Biomed) and the Association Recherche Biomédicale et Diagnostic. The scientific responsibility is assumed by the authors. F.L. is Chercheur Qualifié of the Fonds National de la Recherche Scientifique and M.S. is a post-doctoral fellow of the European Community.

REFERENCES

1. Feng, Y., Broder, C.C., Kennedy, P.E. and Berger, E.A. (1996) HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor. *Science*, **272**, 872–877.
2. Doranz, B.J., Rucker, J., Yi, Y., Smyth, R.J., Samson, M., Peiper, S.C., Parmentier, M., Collman, R.G. and Doms, R.W. (1996) A dual-tropic primary HIV-1 isolate that uses fusin and the beta-chemokine receptors CKR-5, CKR-3, and CKR-2b as fusion cofactors. *Cell*, **85**, 1149–1158.
3. Deng, H., Liu, R., Ellmeier, W., Choe, S., Unutmaz, D., Burkhart, M., Di Marzio, P., Marmon, S., Sutton, R.E., Hill, C.M., Davis, C.B., Peiper, S.C., Schall, T.J., Littman, D.R. and Landau, N.R. (1996) Identification of a major co-receptor for primary isolates of HIV-1. *Nature*, **381**, 661–666.
4. Alkhatib, G., Combadiere, C., Broder, C.C., Feng, Y., Kennedy, P.E., Murphy, P.M. and Berger, E.A. (1996) CC-CKR5: a RANTES, MIP-1α, MIP-1β receptor as a fusion cofactor for macrophage-tropic HIV-1. *Science*, **272**, 1955–1958.
5. Dragic, T., Litwin, V., Allaway, G.P., Martin, S.R., Huang, Y., Nagashima, K.A., Cayanan, C., Maddon, P.J., Koup, R.A., Moore, J.P. and Paxton, W.A. (1996) HIV-1 entry into CD4+ cells is mediated by the chemokine receptor CC-CKR-5. *Nature*, **381**, 667–673.
6. Choe, H., Farzan, M., Sun, Y., Sullivan, N., Rollins, B., Ponath, P.D., Wu, L., Mackay, C.R., LaRosa, G., Newman, W., Gerard, N., Gerard, C. and Sodroski, J. (1996) The β-chemokine receptors CCR3 and CCR5 facilitate infection by primary HIV-1 isolates. *Cell*, **85**, 1135–1148.
7. Baggiolini, M., Dewald, B. and Moser, B. (1994) Interleukin 8 and related chemotactic cytokines: CXC and CC chemokines. *Adv. Immunol.*, **55**, 97–179.
8. Samson, M., Labbé, O., Mollereau, C., Vassart, G. and Parmentier, M. (1996) Molecular cloning and functional expression of a new human CC-chemokine receptor gene. *Biochemistry*, **35**, 3362–3367.
9. Zhang, L., Huang, Y., He, T., Cao, Y. and Ho, D.D. (1996) HIV-1 subtype and second-receptor use. *Nature*, **383**, 768.
10. Roos, M.T., Lange, J.M., de Goede, R.E., Coutinho, R.A., Schellekens, P.T., Miedema, F. and Tersmette, M. (1992) Viral phenotype and immune response in primary human immunodeficiency virus type 1 infection. *J. Infect. Dis.*, **165**, 427–432.
11. Wolinsky, S.M., Korber, B.T., Neumann, A.U., Daniels, M., Kunstman, K.J., Whetsell, A.J., Furtado, M.R., Cao, Y., Ho, D.D. and Safrit, J.T. (1996) Adaptive evolution of human immunodeficiency virus-type 1 during the natural course of infection. *Science*, **272**, 537–542.
12. Zhu, T., Mo, H., Wang, N., Nam, D.S., Cao, Y., Koup, R.A. and Ho, D.D. (1993) Genotypic and phenotypic characterization of HIV-1 patients with primary infection. *Science*, **261**, 1179–1181.
13. Samson, M., Soularue, P., Vassart, G. and Parmentier, M. (1996) The genes encoding the human CC-chemokine receptors CC-CKR1 to 5 are clustered in the p21.3–p24 region of chromosome 3. *Genomics* **36**, 522–526.
14. Paxton, W.A., Martin, S.R., Tse, D., O'Brien, T.R., Skurnick, J., VanDevanter, N.L., Padian, N., Braun, J.F., Kotler, D.P., Wolinsky, S.M. and Koup, R.A. (1996) Relative resistance to HIV-1 infection of CD4 lymphocytes from persons who remain uninfected despite multiple high-risk sexual exposure. *Nature Med.*, **2**, 412–417.

15. Samson, M., Libert, F., Doranz, B.J., Rucker, J., Liesnard, C., Farber, C.M., Saragosti, S., Lapoumeroulie, C., Cognaux, J., Forceille, C., Muyldermans, G., Verhofstede, C., Burtonboy, G., Georges, M., Imai, T., Rana, S., Yi, Y., Smyth, R.J., Collman, R.G., Doms, R.W., Vassart, G. and Parmentier, M. (1996) Resistance to HIV-1 infection in caucasian individuals bearing mutant alleles of the CCR5 chemokine receptor gene. *Nature*, **382**, 722–725.
16. Liu, R., Paxton, W.A., Choe, S., Ceradini, D., Martin, S.R., Horuk, R., MacDonald, M.E., Stuhlmann, H., Koup, R.A. and Landau, N.R. (1996) Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. *Cell*, **86**, 367–377.
17. Dean, M., Carrington, M., Winkler, C., Huttley, G.A., Smith, M.W., Allikmets, R., Goedert, J.J., Buchbinder, S.P., Vittinghoff, E., Gomperts, E., Donfield, S., Vlahov, D., Kaslow, R., Saah, A., Rinaldo, C., Detels, R., Hemophilia Growth and Development Study, Multicenter AIDS Cohort Study, Multicenter Hemophilia Cohort Study, San Francisco City Cohort, ALIVE Study and O'Brien, S.J. (1996) Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the CKR5 structural gene. *Science*, **273**, 1856–1862.
18. Aarons, E., Fernandez, M., Rees, A., McClure, M. and Weber, J. (1997) CC-chemokine receptor 5 genotype and *in vitro* susceptibility to HIV-1 of a cohort of British HIV-exposed uninfected homosexual men. *AIDS*, **11**, 688–689.
19. Huang, Y., Paxton, W.A., Wolinsky, S.M., Neumann, A.U., Zhang, L., He, T., Kang, S., Ceradini, D., Jin, Z., Yazdanbakhsh, K., Kunstman, K., Erickson, D., Dragon, E., Landau, N.R., Phair, J., Ho, D.D. and Koup, R.A. (1996) The role of a mutant CCR5 allele in HIV-1 transmission and disease progression. *Nature Med.*, **2**, 1240–1243.
20. Biti, R., French, R., Young, J., Bennet, B. and Steward, G. (1997) HIV-1 infection in an individual homozygous for the CCR5 deletion allele. *Nature Med.*, **3**, 252–253.
21. O'Brien, T.M., Winkler, C., Dean, M., Nelson, J., Carrington, M., Michael, N. and White, G.C. (1997) HIV-1 infection in a man homozygous for CCR5Δ32. *Lancet*, **349**, 1219.
22. Theodorou, I., Meyer, L., Magierowska, M., Katlama, C. and Rouzioux, C. (1997) HIV-1 infection in an individual homozygous for CCR5 delta 32. *Lancet*, **349**, 1219–1220
23. Rana, S., Besson, G., Cook, D.G., Rucker, J., Smyth, R.J., Yi, Y., Turner, J., Guo, H., Du, J.-G., Peiper, S.C., Lavi, E., Samson, M., Libert, F., Liesnard, C., Vassart, G., Doms, R.W., Parmentier, M., and Collman, R.G. (1997) Role of CCR5 in infection of primary macrophages and lymphocytes by M-tropic strains of HIV: resistance to patient-derived and prototype isolates resulting from the Δ*ccr5* mutation. *J. Virol.*, **71**, 3219–3227.
24. Rappaport, J., Cho, Y.-Y., Hendel, H., Schwartz, E.J., Schachter, F. and Zagury, J.-F. (1997) 32 bp CCR-5 gene deletion and resistance to fast progression in HIV-1 infected heterozygotes. *Lancet*, **349**, 922–923.
25. Eugen-Olsen, J., Iversen, A.K.N., Garred, P., Koppelhus, U., Pedersen, C., Benfield, T.L., Sorensen, A.M., Katzenstein, T., Dickmeiss, E., Gerstoft, J., Skinhoj, P., Svejgaard, A., Nielsen, J.O. and Hofmann, B. (1997) Heterozygosity for a deletion in the CKR-5 gene leads to prolonged AIDS-free survival and slower CD4 T-cell decline in a cohort of HIV-seropositive individuals. *AIDS*, **11**, 305–310.
26. Michael, N.L., Chang, G., Louie, L.G., Mascola, J.R., Dondero, D., Birx, D.L. and Sheppard, H.W. (1997) The role of viral phenotype and CCR-5 gene defects in HIV-1 transmission and disease progression. *Nature Med.*, **3**, 338–340.
27. Martinson, J.J., Chapman, N.H., Rees, D.C., Liu, Y.T. and Clegg, J.B. (1997) Global distribution of the CCR5 gene 32-basepair deletion. *Nature Genet.*, **16**, 100–103.
28. Ansari-Lari, M.A., Liu, X.M., Metzker, M.L., Rut, A.R. and Gibbs, R.A. (1997) The extent of genetic variation in the CCR5 gene. *Nature Genet.*, **16**, 221–222.
29. He, J., Chen, Y., Farzan, M., Choe, H., Ohagen, A., Gartner, S., Busciglio, J., Yang, X., Hofmann, W., Newman, W., Mackay, C.R., Sodroski, J. and Gabuzda, D. (1997) CCR3 and CCR5 are co-receptors for HIV-1 infection of microglia. *Nature*, **385**, 645–649.
30. Wu, L., Paxton, W.A., Kassam, N., Ruffing, N., Rottman, J.B., Sullivan, N., Choe, H., Sodroski, J., Newman, W., Koup, R.A. and Mackay, C.R. (1997) CCR5 levels and expression pattern correlate with infectability by macrophage-tropic HIV-1, *in vitro*. *J. Exp. Med.*, **185**, 1681–1691.
31. Granelli-Piperno, A., Moser, B., Pope, M., Chen, D., Wei, Y., Isdell, F., O'Doherty, U., Paxton, W., Koup, R., Mojsov, S., Bhardwaj, N., Clark-Lewis, I., Baggiolini, M. and Steinman, R.M. (1996) Efficient interaction of HIV-1 with purified dendritic cells via multiple chemokine coreceptors. *J. Exp. Med.*, **184**, 2433–2438.
32. Garred P., Eugen-Olsen, J., Iversen, A.K., Benfield, T.L., Svejgaard, A. and Hofmann, B. (1997) Dual effect of CCR5 delta 32 gene deletion in HIV-1-infected patients. Copenhagen AIDS Study Group. *Lancet*, **349**, 1884.
33. McCombie, W. R., Wilson, R., Chen, E., Gibbs, R., Zuo, L., Johnson, D., Nhan, M., Parnell, L., Dedhia, N., Ansari, A., Mardis, E., Schutz, K., Gnoj, L., de la Bastide, M., Kaplan, N., Greco, T., Touchman, J., Muzny, D., Chen, C.-N., Evans, C., FitzGerald, M., See, L. H., Tang, M., Porcel, B. M., Dragan, Y., Giacalone, J., Pae, A., Powell, E., Solinsky, K. A., DeSilva, U., Diaz-Perez, S., Zhou, X., Yu, Y., Watanabe, M., Doggett, N., Garcia, D. and Sagripanti, J.-L. (1997) Human BAC clone 110P12. EMBL/Genbank U95626.
34. Weber, J.L. and Wong, C. (1993) Mutations of human short tandem repeats. *Hum. Mol. Genet.* **2**, 1123–1128.
35. Rubinsztein, D.C., Amos, W., Leggo, J., Goodburn, S., Jain, S., Li, S.-H., Margolis, R.L., Ross, C.A. and Ferguson-Smith, M.A. (1995) Microsatellite evolution—evidence for directionality and variation in rate between species. *Nature Genet.*, **10**, 337–343.
36. Hästbacka, J., de la Chapelle, A., Sistonen, P., Weaver, A. and Lander, E. (1992) Linkage disequilibrium mapping in isolated founder populations: diastrophic dysplasia in Finland. *Nature Genet.*, **2**, 204–211.
37. Wennberg, C. and Kucinkas, V. (1994) Low frequency of the ΔF508 mutation in Finno-Ugrian and Baltic populations. *Hum. Hered.*, **44**, 169–171.
38. Beckman, L.E., Saha, N., Spitsyn, V., Van Landeghem, G. and Beckman, L. (1997) Ethnic differences in the HFE codon 282 (Cys/Tyr) polymorphism. *Hum. Hered.*, in press.
39. Sajantila, A., Lahermo, P., Anttinen, T., Lukka, M., Sistonen, P., Savontaus, M.L., Aula, P., Beckman, L., Tranebjaerg, L., Gedde Dahl, T., Issel Tarver, L., DiRienzo, A. and Paabo, S. (1995) Genes and languages in Europe: an analysis of mitochondrial lineages. *Genome Res.*, **5**, 42–52.
40. Motulsky, A.G. (1995) Jewish diseases and origins. *Nature Genet.*, **9**, 99–101.
41. Cavalli-Sforza, L.L., Menozzi, P. and Piazza, A. (1993) *The History and Geography of Human Genes*. Princeton University Press, Princeton, NJ.
42. Deng, H., Unutmaz, D., KawalRamani, V.N. and Littman, D. (1997) Expression cloning of new receptors used by simian and human immunodeficiency viruses. *Nature*, **388**, 296–300.
43. Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.