Global distribution of the *CCR5* gene 32-basepair deletion

Jeremy J. Martinson[1,2], Nicola H. Chapman[1], David C. Rees[1], Yan-Tat Liu[1] & John B. Clegg[1]

A mutant allele of the β-chemokine receptor gene *CCR5* bearing a 32-basepair (bp) deletion (denoted Δ*ccr5*) which prevents cell invasion by the primary transmitting strain of HIV-1 has recently been characterized[1–3]. Homozygotes for the mutation are resistant to infection, even after repeated high-risk exposures[1,4], but this resistance appears not to be total, as isolated cases of HIV-positive deletion homozygotes are now emerging[5]. The consequence of the heterozygous state is not clear, but it may delay the progression to AIDS in infected individuals[2,3,6,7]. A gene frequency of approximately 10% was found for Δ*ccr5* in populations of European descent, but no mutant alleles were reported in indigenous non-European populations. As the total number of non-European samples surveyed was small in comparison with the Europeans the global distribution of this mutation is far from clear. We have devised a rapid PCR assay for Δ*ccr5* and used it to screen 3,342 individuals from a globally-distributed range of populations. We find that Δ*ccr5* is not confined to people of European descent but is found at frequencies of 2–5% throughout Europe, the Middle East and the Indian subcontinent (Fig. 1). Isolated occurrences are seen elsewhere throughout the world, but these most likely represent recent European gene flow into the indigenous populations. The inter-population differences in Δ*ccr5* frequency may influence the pattern of HIV transmission and so will need to be incorporated into future predictions of HIV levels.

The recently-characterized chemokine receptor gene *CCR5* has rapidly become the object of intense interest since the discovery of its role in the entry of HIV-1 into target cells[8,9]. The *CCR5* gene product is a member of the seven-transmembrane, G-protein-coupled receptor family[10,11] which, in response to their normal β-chemokine ligands, are involved in the chemotaxis of leukocytes towards sites of inflammation[12]. *CCR5* also mediates the entry into cells of the M-tropic strain of HIV-1 that is the primary transmitting form of the virus[13–17]. The interaction between HIV-1, CCR5 and CD4 which results in HIV-1 entry is still being characterized[18–20] but this mechanism already shows promise as a therapeutic strategy for blocking HIV-1 infection[21]. A mutant allele of *CCR5* has recently been described which carries a 32-bp deletion (denoted Δ*ccr5* (ref. 2): the product of this truncated gene is not expressed on the cell surface and so prevents viral invasion[1,4]. Homozygotes for the deletion are highly resistant to HIV infection, but the effect of the heterozygous condition is less clear; Samson *et al.*[2] claim that heterozygotes have partial resistance to infection, although several other studies suggest that susceptibility to infection is not lowered but that progression to AIDS is postponed[3,6,7]. The protective effect of the homozygous Δ*ccr5* is not total; a case of an HIV-positive homozygote has recently been reported[5]. *CCR5* is the entry cofactor for the M-tropic strain of HIV-1, but although this is the primary transmitting strain of the virus, most infection events include both M-tropic and T-tropic strains[9]. As T-tropic strains utilize a different entry cofactor this may explain the infection in this case; alternatively, infection may have been by an M-tropic strain using a different, *CCR5*-related cofactor. The tropism of the virus in this case has not yet been determined[5], but it seems that infections of this type are rare, and that Δ*ccr5* is generally highly protective against HIV-1 infection in the homozygous state.

Population surveys estimate the allele frequency of Δ*ccr5* at approximately 10% in people of European descent but have not revealed any instances in black populations[1,2] with the exception of African-Americans[3], in whom admixture with people of European descent has been considerable. Over 4,000 European or Caucasian-American samples, but only 747 indigenous non-European, have been studied so far, and so the reported absence of Δ*ccr5* outside of Europe may reflect the limited dataset used. We have used a PCR-based system to survey 3,342 individuals from a globally-distributed range of populations, as well as 66 unrelated chimpanzee samples (Table 1).

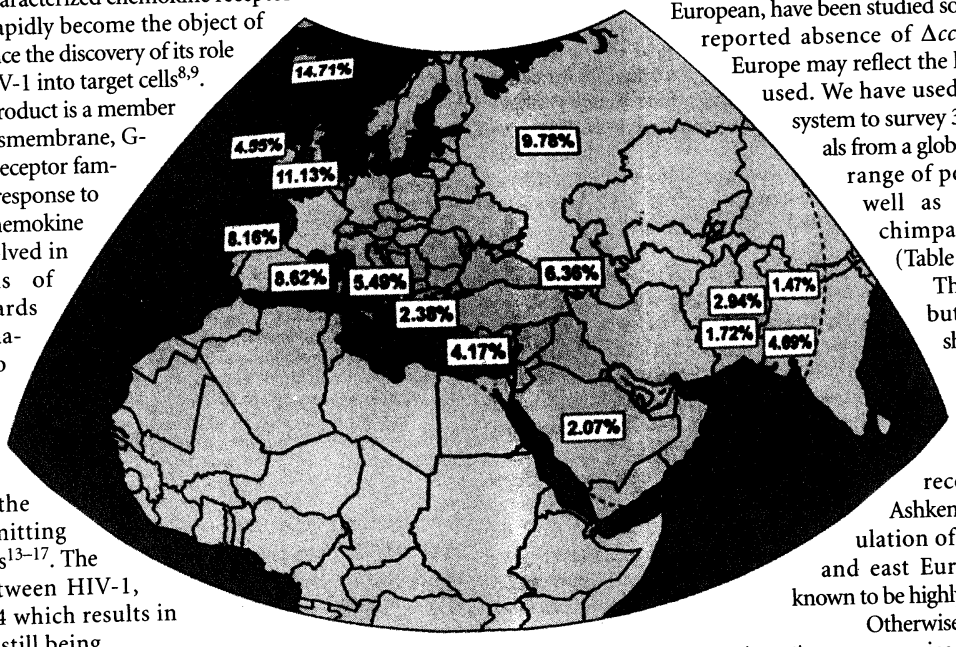


Fig. 1 Δ*ccr5* allele frequencies in Europe, the Middle East and India. This map shows the extent of the distribution of Δ*ccr5* in Europe and Asia. Populations correspond to those shown in Table 1. The dashed line shows the boundary between those populations where Δccr5 frequencies >1.5% were observed and those in which the deletion was absent, or present only as isolated observations. The Ashkenazi samples are omitted from this figure.

The global distribution of Δ*ccr5* is shown in Table 1. The highest allele frequency recorded is 20.93%, recorded in the Ashkenazi Jews, a population of ancient Israeli and east European descent known to be highly endogamous[22]. Otherwise, the highest frequencies observed are in European populations, but considerable variation in frequency is seen even

[1]*MRC Molecular Haematology Unit, Institute of Molecular Medicine, John Radcliffe Hospital, Headington, Oxford, OX3 9DU, UK;* [2]*University of Pittsburgh, Dept. Human Genetics, Graduate School of Public Health, 130 DeSoto St., Pittsburgh, Pennsylvania 15261, USA. Correspondence should be addressed to J.J.M. e-mail: jmartins@helix.hgen.pitt.edu*

Table 1 • Global distribution of *Δccr5*

Population	No. studied	Genotype *CCR5/ CCR5*	*CCR5/ Δccr5*	*Δccr5/ Δccr5*	%*Δccr5*	95% C.I.	HWE χ^2 (*P*-value)
Europe							
Ashkenazi	43	26	16	1	20.93	16.54–25.32	0.6631 (n.s.)
Iceland	102	75	24	3	14.71	12.23–17.19	0.3930 (n.s.)
Britain	283	223	57	3	11.13	9.81–12.45	0.0925 (n.s.)
Spain: Basque	29	24	5	0	8.62	4.94–12.31	0.2581 (n.s.)
Spain: Catalonia	49	41	8	0	8.16	5.40–10.93	0.3872 (n.s.)
Italy	91	81	10	0	5.49	3.81–7.18	0.3076 (n.s.)
Ireland	44	40	4	0	4.55	2.32–6.77	0.0998 (n.s.)
Cyprus	84	77	7	0	4.17	2.62–5.71	0.1588 (n.s.)
Greece	63	60	3	0	2.38	1.02–3.74	
Middle East							
Caucasus: Daghestan	110	96	14	0	6.36	4.72–8.01	0.5081 (n.s.)
Saudi Arabia	241	231	10	0	2.07	1.43–2.72	0.1082 (n.s.)
Yemen	34	34	0	0	-		
Asia							
Russia: Udmurtia	46	38	7	1	9.78	6.69–12.88	0.8746 (n.s.)
Gujarat	32	30	1	1	4.69	2.05–7.33	13.5314 ($P<0.0005$)
Pakistan	34	32	2	0	2.94	0.89–4.99	0.0312 (n.s.)
Sindh	29	28	1	0	1.72	0.01–3.43	
Punjab	34	33	1	0	1.47	0.01–2.93	
Bengal	25	25	0	0	-		
Hong Kong	50	50	0	0	-		
Taiwan	83	83	0	0	-		
Philippines: Filipino	26	26	0	0	-		
Philippines: Negrito	30	30	0	0	-		
Mongolia	59	59	0	0	-		
Sri Lanka	37	37	0	0	-		
Burma	67	67	0	0	-		
Thailand	101	100	1	0	0.50	0.00–0.99	
Borneo	151	151	0	0	-		
Sumatra	72	72	0	0	-		
Andaman Islands	24	24	0	0	-		
Africa							
Nigeria	111	110	1	0	0.45	0.00–0.90	
Gambia	56	56	0	0	-		
Central African Republic	52	52	0	0	-		
Kenya	80	80	0	0	-		
Ivory Coast	87	87	0	0	-		
Malawi	80	80	0	0	-		
Zambia	96	96	0	0	-		
Kalahari San	36	36	0	0	-		
Oceania							
New Guinea Coast	96	96	0	0	-		
French Polynesia	94	94	0	0	-		
Aboriginal Australian	98	96	2	0	1.02	0.30–1.74	
Guam	59	58	1	0	0.85	0.00–1.69	
Fiji	17	17	0	0	-		
Americas							
Nuu-Chah-Nulth	38	37	1	0	1.32	0.01–2.62	
Mexico (Huicholes)	52	52	0	0	-		
Brazilian Amerindian	98	98	0	0	-		
Jamaica	119	119	0	0	-		
Primates							
Chimpanzee	66	66	0	0	-		
TOTAL:	3,408						

Each population was initially collected as part of anthropological surveys or investigations of non-HIV-related medical conditions. In each case, samples with known non-indigenous ancestry were eliminated from the study. For populations where the *Δccr5* frequency was >2.0%, the genotype distribution was tested for agreement with Hardy-Weinberg equilibrium (HWE); the value of χ^2, with the associated *P*-value for significant departures, is shown. Good agreement was seen for each population, with the exception of the Gujerati.

within Europe, with southern European and Mediterranean populations generally showing lower frequencies than northern Europeans. Outside of Europe, *Δccr5* is seen at frequencies above 5% in the Caucasus and Russian western Asia and at frequencies of 2–5% in populations from Saudi Arabia as far east as India and Pakistan. Outside of this region, the mutation is either absent or only present as isolated individual occurrences (Table 1). The distribution of *Δccr5* in Europe, the Middle East and Asia is shown in Fig. 1.

The sporadic appearance of *Δccr5* in the other populations (Table 1) can most easily be explained by gene flow from Caucasian populations into these indigenous peoples in the recent colonial past. It is of course possible that the same mutation has

occurred independently in these populations, but as the sizes of the deletion alleles in each case are identical this seems unlikely. Also, the genomic structure of *CCR5* does not indicate any intragenic repeat structures likely to facilitate the removal of this 32 bp region by recombination mechanisms[11,23]. Recent gene flow from Caucasian populations cannot be excluded as the cause of Δ*ccr5* frequencies in the Asian Indian samples, but substantial amounts of gene flow would be necessary to produce the observed frequencies. The allele frequency data alone are not sufficient to resolve these issues, but as the genomic location and sequence of *CCR5* is now known[11,23] the determination and analysis of flanking sequence polymorphisms may allow the construction of linked haplotypes which can then be used to determine the probability of the mutation being produced more than once on different chromosomal backgrounds, as has been done by ourselves for the factor V Leiden mutation[24] and the haemoglobinopathies[25].

The fluctuation in gene frequency seen in Europe is consistent with the effects of genetic drift acting on a neutral polymorphism, that is, one that confers no selective advantage or disadvantage on its host[26]. The introduction of HIV-1 into European populations has been too recent to have affected the distribution of Δ*ccr5*. No deleterious phenotypes have been reported for either heterozygous or homozygous individuals, and, with the exception of the Gujerati sample, the genotypes themselves are in Hardy-Weinberg equilibrium (HWE) (Table 1) suggesting that no genotype confers an advantage or disadvantage in the absence of HIV-1. The deviation from HWE seen in the Gujerati sample could reflect the presence of individuals with unknown European ancestry, but we feel that the deviation is due to the small size of this sample.

The experimental approach adopted in this study reports only on the distribution of the 32-bp Δ*ccr5* deletion, and not on any other mutations which may impair the ability of *CCR5* to act as an entry cofactor for HIV-1. Dean *et al.*[3] reported several other mutations detected by SSCP analysis, but did not characterize them all in detail. A more exhaustive population-based survey of variation across the entire *CCR5* gene would reveal whether any of these mutations achieve high frequencies in particular populations, and would not suffer from a bias towards the Δ*ccr5* deletion. None of these mutations was seen with a frequency >1% in the populations studied by Dean *et al.*, however, it still seems that the 32-bp Δ*ccr5* deletion is the most prevalent mutation in the *CCR5* gene—and the one whose distribution and function needs the closest examination.

The general absence of Δ*ccr5* in sub-Saharan Africa, and in the 66 chimpanzee samples we studied (Table 1) suggests a recent origin of the mutation, but also demonstrates that it is not a significant factor in the prevalence or transmission of HIV-1 within Africa. Rowland-Jones *et al.*[27] studied six HIV-resistant, but multiply-exposed, female prostitutes in The Gambia and found that three of these could be explained by the presence of HIV-specific cytotoxic T-lymphocytes. Δ*ccr5* was absent from the Gambian samples we studied, suggesting that at least one other mutation (in *CCR5* or elsewhere) is responsible for the remaining individuals who are known to be resistant to HIV infection. The difference in Δ*ccr5* allele frequencies between Europe and the other areas of the world where HIV incidence is rising may be a factor in the different rates of transmission seen in these regions. A gene frequency of 10% will, under Hardy-Weinberg equilibrium, result in a heterozygote frequency of 18% and a homozygote frequency of 1%; a population in which almost 20% of individuals are wholly or partially resistant to infection may show much lower rates of disease transmission than one in which all individuals are equally susceptible. The inter-population variation in Δ*ccr5* frequency may be a significant factor in the prediction of AIDS endemicity in future studies[28].

A variety of other genetic polymorphisms are now known to have a similar Europe-centred distribution, such as the Δ*F508* cystic fibrosis mutation[29], α-1 antitrypsin deficiency[30], factor V Leiden[31] and haemochromatosis[32]. These differ from Δ*ccr5* in that they are deleterious conditions whose persistence needs to be explained by other factors such as resistance to infectious disease in the heterozygous state. The high frequency of Δ*ccr5* in Europe may indicate that it too has had a long-term selective advantage against infectious diseases using *CCR5* as entry cofactors, as suggested by Dean *et al.*[3], but another explanation is possible—that it is a previously-neutral polymorphism whose selective advantage is just becoming apparent.

Methods

Samples. Samples studied were obtained from unrelated individuals in each of the regions studied. Genomic DNA (50–100ng) from each individual was amplified in a total volume of 21 μl, in a buffer containing 100 mM Tris-Cl pH8.8, 50 mM KCl, 1.5 mM $MgCl_2$, 0.2 mM each dNTP, 1 U *Taq* polymerase and 12.5 pmoles each primer. Primer sequences were: Δ*ccr5*F 5'–ATCACTTGGGTGGTGGCTGTGTTTGCGTCTC–3' and Δ*ccr5*R 5'–AGTAGCAGATGACCATGACAAGCAGCGGCAG–3', corresponding to bases 505–535 and 667–697 respectively of the published sequence[23]. Cycling conditions were: 94 °C 5min, 1 cycle; 94 °C 30s, 70 °C 30 s with 1s per-cycle increment for 30 cycles. The polymorphism was detected by electrophoresis through 4% MetaPhor agarose gels (FMC corporation): allele sizes were 193 bp for the normal allele, 161 bp for the deletion allele. Amplifications were performed in 96 well microtitre plates, affording a high throughput of samples.

Acknowledgements

We are grateful to B. Sykes, M. Richards, J. Old, D. Tumen, K. Bulayeva, N. van den Broek, N. Raben, S. Thein, D. Higgs, R. Ward and J. Bertranpetit for access to DNA samples. This work was supported by the MRC and the Wellcome Trust.

Received 10 September 1996; accepted 31 March 1997

1. Liu, R. *et al.* Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. *Cell* **86**, 367–77 (1996).
2. Samson, M. *et al.* Resistance to HIV-1 infection in caucasian individuals bearing mutant alleles of the CCR5 chemokine receptor gene. *Nature* **382**, 722–725 (1996).
3. Dean, M. *et al.* Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the *CKR5* structural gene. *Science* **273**, 1856–1862 (1996).
4. Paxton, W.A. *et al.* Relative resistance to HIV-1 infection of CD4 lymphocytes from persons who remain uninfected despite multiple high-risk sexual exposures. *Nature Med.* **2**, 412–417 (1996).
5. Biti, R., French, R., Young, J., Bennetts, B. & Stewart, G. HIV-1 infection in an individual homozygous for the *CCR5* deletion allele. *Nature Med.* **3**, 252–253 (1997).
6. Huang, Y. *et al.* The role of a mutant CCR5 allele in HIV-1 transmission and disease progression. *Nature Med.* **2**, 1240–1243 (1996).
7. Michael, N.L. *et al.* The role of viral phenotype and *CCR5* gene defects in HIV-1 transmission and disease progression. *Nature Med.* **3**, 338–40 (1997).
8. Hill, C.M. & Littman, D.R. Natural resistance to HIV? *Nature* 382, 668–669 (1996).
9. Fauci, A.S. Resistance to HIV-1 infection: It's all in the genes. *Nature Med.* **2**, 966–967 (1996).
10. Neote, K., DiGregorio, D., Mak, J.Y., Horuk, R. & Schall, T.J. Molecular cloning, functional expression, and signalling characteristics of a C-C chemokine receptor. *Cell* **72**, 415–425 (1993).
11. Samson, M., Labbe, O., Mollereau, C., Vassart , G. & Parmentier, M. Molecular cloning and functional expression of a new human CC-chemokine receptor gene. *Biochemistry* **35**, 3362–367 (1996).
12. Murphy, P.M. The molecular biology of leukocyte chemoattractant receptors. *Annu. Rev. Immunol.* **12**, 593–633 (1994).
13. Alkhatib, G. *et al.* CC CKR5: A RANTES, MIP-1α, MIP-1β receptor as a fusion cofactor for macrophage-tropic HIV-1. *Science* **272**, 1955–1958 (1996).

14. Choe, H. *et al.* The β-chemokine receptors CCR3 and CCR5 facilitate infection by primary HIV-1 isolates. *Cell* **85**, 1135–1148 (1996).
15. Deng, H. *et al.* Identification of a major co-receptor for primary isolates of HIV-1. *Nature* **381**, 661–666 (1996).
16. Doranz, B.J. *et al.* A dual-tropic primary HIV-1 isolate that uses fusin and the β-chemokine receptors CKR-5, CKR-3 and CKR-2b as fusion cofactors. *Cell* **85**, 1149–1158 (1996).
17. Dragic, T. *et al.* HIV-1 entry into CD4+ cells is mediated by the chemokine receptor CC-CKR-5. *Nature* **381**, 667–673 (1996).
18. Atchison, R.E. *et al.* Multiple extracellular elements of CCR5 and HIV-1 entry: Dissociation from response to chemokines. *Science* **274**, 1924–1926 (1996).
19. Trkola, A. *et al.* CD4-dependent, antibody-sensitive interactions between HIV-1 and its co-receptor CCR5. *Nature* **384**, 184–187 (1996).
20. Wu, L. *et al.* CD4-induced interaction of primary HIV-1 gp120 glycoproteins with the chemokine receptor CCR5. *Nature* **384**, 179–183 (1996).
21. Arenza-Seisdedos, F. *et al.* HIV blocked by chemokine antagonist. *Nature* **383**, 400 (1996).
22. Motulsky, A. Jewish diseases and origins. *Nature Genet.* **9**, 99–101 (1995).
23. Raport, C.J., Gosling, J., Schweickart, V.L., Gray, P.W. & Charo, I.F. Molecular cloning and functional characterization of a novel human CC chemokine receptor (CCR5) for RANTES, MIP-1β, and MIP-1α. *J. Biol. Chem.* **271**, 17161–17166 (1996).
24. Cox, M.J., Rees, D.C., Martinson, J.J. & Clegg, J.B. Evidence for a single origin of factor V Leiden. *Br. J. Haematol.* **92**, 1022–1025 (1996).
25. Flint, J., Harding, R.M., Clegg, J.B. & Boyce, A.J. Why are some genetic diseases common? Distinguishing selection from other processes by molecular analysis of globin gene variants. *Hum. Genet.* **91**, 91–117 (1993).
26. Kimura, M. *The Neutral Theory of Molecular Evolution* (Cambridge Univ. Press, Cambridge, 1983).
27. Rowland-Jones, S. *et al.* HIV-specific cytotoxic T-cells in HIV-exposed but uninfected Gambian women. *Nature Med.* **1**, 59–64 (1995).
28. DeGruttola, V., Seage III, G.R., Mayer, K.H. & Horsburgh, C.R. Jr Infectiousness of HIV between male homosexual partners. *J. Clin. Epidemiol.* **42**, 849–856 (1989).
29. Morral, N. *et al.* The origin of the major cystic fibrosis mutation (ΔF508) in European populations. *Nature Genet.* **7**, 169–175 (1994).
30. Cox, D.W. α_1-Antitrypsin deficiency. in *The Metabolic and Molecular Bases of Inherited Disease* 7th edn. (eds Scriver, C.R., Beaudet, A.L., Sly, W.S. & Valle, D.) 4125–4158 (McGraw-Hill, New York, 1995).
31. Rees, D.C., Cox, M. & Clegg, J.B. World distribution of factor V Leiden. *Lancet* **346**, 1133–1134 (1995).
32. Edwards, C.Q. *et al.* Prevalence of hemochromatosis among 11,065 presumably healthy blood donors. *N. Engl. J. Med.* **318**, 1355–1362 (1988).