Close linkage of α and β interferons and infrequent duplication of β interferon in humans

(restriction fragment length polymorphism/segregation analysis/haplotypes/linkage disequilibrium)

MONICA OHLSSON*, JOHN FEDER†, LUIGI L. CAVALLI-SFORZA†, AND ALEXANDER VON GABAIN*

*Unit of Applied Cell and Molecular Biology, Umea, Sweden; and †Department of Genetics, Stanford University, Stanford, CA 94305

Contributed by Luigi L. Cavalli-Sforza, March 18, 1985

ABSTRACT Five restriction fragment length polymorphisms in the human α/β interferon (IFN- α/β) gene region were identified, three with an IFN- α probe and two with an IFN- β probe. Heterozygosities are 74% for IFN- α , 57% for IFN- β , and 87% jointly, making IFN- α/β genes excellent markers for the short arm of chromosome 9. The pedigrees of about 25 families of Caucasian background were studied. Segregation analysis disclosed the occurrence of 12 of 32 possible haplotypes. No recombinant was found between IFN- α and - β genetic markers; linkage disequilibrium within the IFN- α markers is of a similar order of magnitude as that between the IFN- α and - β markers. The IFN- α and - β genes might cluster within several hundred kilobases. In two parents, the IFN- β gene is duplicated; the duplications segregate regularly. They are of independent, probably recent, origin. Therefore, some degree of multiplicity might even be found for IFN- β , at least in some individuals.

Three interferon (IFN) gene clusters have been described: IFN- α , - β , and - γ genes, of which the first two show great homology to each other and are located in the same or adjacent bands of chromosome 9 short arm, while IFN- γ gene is located in chromosome 12 long arm (1). Both IFN- α and - β genes have no introns. There is unequivocal evidence that IFN- α is encoded by a gene family in humans (1). The number of different IFN- α gene sequences isolated so far has been given as 15 (2), but when unpublished observations are included, it may be as large as 30. Some of this may be due to polymorphic variation (3), although this has not yet been directly observed.

We describe in this paper restriction fragment length polymorphisms for IFN- α and - β genes and study their segregation, recombination, and linkage disequilibrium. The existence of a polymorphism at the protein level will still require direct verification, but the variation described here will make it easier to recognize it and provides the basis for a study of disease association with qualitative or quantitative variation of IFNs.

MATERIALS AND METHODS

Blood samples were obtained from Caucasian individuals belonging to pedigrees from the San Francisco Bay Area (4) and French families from CEPH. DNA was extracted from buffy coats (5).

DNA (8–10 μ g) was digested under conditions suggested by the manufacturers of the restriction endonucleases. Digested DNA was electrophoresed on 0.8% (wt/vol) horizontal agarose gels and transferred to nitrocellulose or Zetabind (AMF Cuno, Meriden, CT) by the Southern method (6). DNA probes were nick-translated with [α - 32 P]dCTP and [α - 32 P]-dATP to a specific activity of $1\text{--}3 \times 10^8$ cpm/g. Labeled

probe at 10–20 ng was used per ml of hybridization solution. Hybridization was done in 50% formamide/0.75 M NaCl/75 mM sodium citrate/0.1% NaDodSO₄/1 $\times$ Denhardt's solution (0.02% bovine serum albumin/0.02% Ficoll/0.02% polyvinylpyrrolidone)/100 μ g of single-stranded DNA per ml at 42°C for 24–48 hr, followed by high-stringency washes.

The IFN- α probe used is the cDNA insert of an IFN- α_2 clone derived from Namalwa cell line (7), and the IFN- β probe is the *Eco*RI fragment F of pcosIFN β (8) subcloned in a plasmid; in one instance a cDNA insert of an IFN- β clone containing only the coding region (7) was used as probe.

Haplotypes of pedigree founders were reconstructed by hand and also by a computer program being developed by Ellen Wijsman at the Stanford University Genetics Department.

RESULTS

Polymorphisms. A search for restriction fragment length polymorphisms was made by testing the restriction patterns of DNA from 7–10 individuals with the following enzymes: *Apa* I, *Bam*HI, *Ban* II, *Bcl* I, *Bgl* I, *Bgl* II, *Bst*EII, *Dra* I, *Eco*RI, *Eco*RV, *Hinc*II, *Hind*III, *Hinf*I, *Mbo* I, *Msp* I, *Pst* I, *Pvu* I, *Rsa* I, *Sst* I, *Sin* I, and *Taq* I for IFN- α ; and *Ban* II, *Eco*RI, *Hinc*II, *Hind*III, *Hpa* I, *Msp* I, *Pst* I, *Pvu* II, *Stu* I, *Taq* I, and *Xma* I for IFN- β .

With several enzymes, individual differences were found. Of these, *Msp* I showed most polymorphic bands for IFN- α genes. For IFN- β genes, both *Msp* I and *Ban* II showed two different high-frequency polymorphisms.

Altogether five enzymes showed markers segregating at relatively high frequency which, therefore, can be considered as polymorphic—three with the IFN- α probe (all with the same enzyme, *Msp* I) and two with the IFN- β probe with different enzymes (*Msp* I and *Ban* II and/or *Sst* I). The polymorphism found with the IFN- β probe and *Msp* I is identical to one already described (9). Fig. 1 shows the restriction patterns of DNA from individuals of two families tested with *Msp* I and the IFN- α probe. Altogether, 15 *Msp* I bands could be seen, 2 of which [4.8 and 4.6 kilobases (kb), respectively] have a higher multiplicity. Three different IFN- α gene fragment polymorphisms showed high frequency; two of these can be considered as codominant, and their family segregations are shown in Table 1.

(i) The *Msp* I fragments of 7.5 and 8.6 kb have gene frequencies of 0.447 and 0.553 ± 0.059 , respectively (always calculated on the parents of the families). Hardy-Weinberg equilibrium is satisfactory, and so are the Mendelian segregations of Table 1 in families in which the 8.6-kb band is clearly legible. In others this band was too faint to be legible unambiguously, but here the 7.5-kb band segregates as a clear-cut dominant marker.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: IFN- α and - β , α and β interferons; kb, kilobase(s).

FIG. 1. Autoradiogram after hybridization to the IFN- α -specific probe of *Msp* I-digested DNAs from two families. (A) Pattern of one family composed of parents (lanes 1 and 2) and their children (lanes 3 and 4). (B) Pattern of another family composed of grandparents (lanes 1 and 2), their children (3 and 4), second-generation parents (lanes 5/6 and 7/8), and their children (third generation) (lanes 9/10 and 11/12, respectively). The molecular weights (kb) of the analyzed DNA fragments are indicated.

(ii) The *Msp* I fragment of 3.1 and 11.3 kb have gene frequencies of 0.190 and 0.810 ± 0.043 , respectively. As above, Hardy-Weinberg equilibrium on parents and segregation are satisfactory. Again, the high molecular weight band is difficult to read in four families not reported in Table 1, but the 3.1-kb band segregates well as a dominant marker.

(iii) The 7.1-kb band segregates well as a dominant marker, as

Table 1. Segregation of the IFN- α markers in nuclear families with the *Msp* I enzyme

Three high-frequency IFN- α RFLPs						
Matings*	No. of families	No. of children with <i>Msp</i> I markers			χ^2	df
		8.6/ 8.6	8.6/ 7.5	7.5/ 7.5		
8.6/8.6 $\times$ 8.6/7.5	4	6	6	0	0	1
8.6/8.6 $\times$ 7.5/7.5	2	0	7	0		
8.6/7.5 $\times$ 8.6/7.5	6	6	3	4	4.38	2
8.6/7.5 $\times$ 7.5/7.5	6	0	14	10	0.67	1
7.5/7.5 $\times$ 7.5/7.5	1	0	0	2		
Total	19				5.05	4
					$P > 5\%$	
		11.3/ 11.3	11.3/ 3.1	3.1/ 3.1		
11.3/11.3 $\times$ 11.3/11.3	9	24	0	0		
11.3/11.3 $\times$ 11.3/3.1	8	10	8	0	0.22	1
11.3/11.3 $\times$ 3.1/3.1	1	0	2	0		
11.3 $\times$ 3.1 $\times$ 11.3/3.1	3	5	3	2	3.40	2
Total	21				3.62	3
					$P > 5\%$	
		7.1 ⁺	7.1 ⁻			
7.1 ⁺ $\times$ 7.1 ⁺	1	2	0		0.57	1
7.1 ⁺ $\times$ 7.1 ⁻	8	10	11		0.05	1
7.1 ⁻ $\times$ 7.1 ⁻	15	0	49			
Total	24				0.62	2
					$P > 5\%$	

df, Degrees of freedom; RFLPs, restriction fragment length polymorphisms.

*Gene markers are indicated by restriction fragment size in kb.

shown in the last part of Table 1; no allelic band is detectable. The gene frequency of the 7.1-kb allele (0.110 ± 0.032) is obtained as $1 - (\text{frequency of negative phenotype})^{1/2}$. The segregation proportions shown in Table 1 were calculated by using Snyder's ratios (10) and agreed well with the observed ones.

In Table 2 are the segregations of the two markers of the IFN- β gene, which were treated as being codominant: (i) The *Msp* I fragments of 3.9, and 2.5 kb are allelic with frequencies $0.83/0.177 \pm 0.039$. (ii) The *Ban* II fragments 3.0 and 2.6 kb (with *Sst* I, fragment sizes are 9.4 and 2.6 kb) are also allelic with frequencies of 0.268 and 0.732 ± 0.059 but are different from the *Msp* I fragments. Segregations were in agreement with Mendelian expectations, as shown by the χ^2 values (Table 2). Also, the frequencies of genotypes among the parents were in agreement with Hardy-Weinberg expectations (χ^2 not given).

Table 2. Segregation of the IFN- β markers in nuclear families with two restriction enzymes

Matings*	No. of families	No. of children with <i>Msp</i> I markers			χ^2	df
		3.9/8.6	3.9/2.5	2.5/2.5		
3.9/3.9 $\times$ 3.9/3.9	10	35	0	0	—	—
3.9/3.9 $\times$ 3.9/2.5	11	14	16	0	0.13	1
3.9/3.9 $\times$ 3.9/2.5	3	1	6	1	2.0	2
Total	24				2.13	3
Matings*	No. of families	No. of children with <i>Sst</i> I/ <i>Ban</i> II markers†			χ^2	df
		9.4/9.4	9.4/2.6	2.6/2.6		
9.4/9.4 $\times$ 9.4/2.6	1	1	3	0	1.0	1
9.4/2.6 $\times$ 9.4/2.6	4	4	9	3	0.37	2
9.4/2.6 $\times$ 2.6/2.6	4	0	9	7	0.25	1
2.6/2.6 $\times$ 2.6/2.6	5	0	0	13	—	—
Total	14				1.62	4

df, Degrees of freedom.

*Gene markers are indicated by restriction fragment size in kb.

†Fragment size of 3.0 kb with *Ban* II is equivalent to a fragment size of 9.4 kb with *Sst* I.

Haplotypes and Linkage Disequilibrium. The reconstruction of haplotypes was made for all genotypes contributing independently to pedigree information. In addition to manual reconstruction of haplotypes, a computer program now being developed by E. Wijsman was also used and gave results in complete agreement with the manual reconstruction. Altogether 65 haplotypes were reconstructed (Table 3) of which 41 were fully defined for all five markers, while for the rest (24 haplotypes), information was missing for one of the five markers. With five diallelic markers, $2^5 = 32$ different haplotypes could exist, but only 12 were observed in this sample. Heterozygosities are 73.8% for IFN- α gene, 56.9% for IFN- β gene, and 86.9% for both together, which make IFN genes excellent genetic markers for the short arm of chromosome 9.

The reconstruction of haplotypes showed that no recombinant between IFN- α and - β markers or within each of the two clusters of markers was observed. The number of gametes tested for IFN- α/β gene recombination was 55, and the upper confidence limit of recombination (at $P = 95\%$) is as high as 5.3%.

From the haplotype frequencies, the linkage equilibria (D) between the five markers taken in all possible ways were calculated using only haplotypes that were completely known and are given in Table 4. They are all of moderate size; their average, ignoring the sign of individual values, varies slightly but is not significantly higher between IFN- α and - β markers ($D = 0.28$) than the average of D values within the α or the β cluster of markers ($D = 0.20$). Two of the 10 D values observed were statistically significant, 1 of them at the $P = 1\%$ level, even though the numbers of haplotypes observed are small.

A Duplication of the IFN- β Gene. In almost all families, the IFN- β gene segregates as a single-copy gene. In two families, however, there was a marked discrepancy in the density of the bands of some individuals (see Fig. 2); in particular, heterozygotes showed a departure from the expected 1:1

Table 4. Association between loci in the population

Loci*	Linkage disequilibrium correlations				
	IFN- α <i>Msp</i> I markers			IFN- β markers	
	7.5/8.6	7.1	3.1	<i>Msp</i> I 3.9/2.5	<i>Ban</i> II 3.0/2.6
7.5/8.6		0.244 (3.5)	0.35† (7.26)	-0.12 (0.85)	-0.31‡ (3.93)
7.1			-0.236 (3.61)	0.169 (1.85)	0.18 (1.52)
3.1				0.155 (1.56)	-0.26 (3.19)
3.9/2.5					0.250 (2.89)
9.4/2.6					

Correlations (r) measuring linkage disequilibrium in and between IFN- α and IFN- β loci. In parentheses below r values are the calculated χ^2 values with 1 degree of freedom each.

*Indicated by restriction fragment sizes in kb.

† $P < 1\%$.

‡ $P < 5\%$.

density ratio. Densitometric measurements were carried out and confirmed the visual impression. The aberrant density ratio in the two abnormal heterozygotes was observed also with another enzyme (*Sst* I) that can recognize the same polymorphic site. Finally, the same results were obtained when a cDNA insert of an IFN- β clone (7) comprising only the coding region was used in the respective experiments.

The two families can be explained on the assumptions that in some individuals the IFN- β gene region is duplicated, and that in the two families (which are totally unrelated) a different duplication is found, indicating that duplications occurred repeatedly in the IFN- β gene region.

In family 10 the light variant seems to be entirely duplicated in the mother and is inherited by 4 of the 10 children. The duplication is observed with two restriction enzymes (*Sst* I and *Ban* I) that distinguish a light (L) and a heavy fragment (H); in both cases, the light fragment has duplicated density. The genotypes of family 10 can be explained by assuming that the father is a normal heterozygote L/H and the mother has a normal L chromosome and one containing the duplicated LL—i.e., is L/LL. Four types of children are observed: L/LL (2 children), LL/H (2), L/L (1), and L/H (5), and these four are expected in equal numbers (likelihood χ^2 for equality is 1.4 with 3 degrees of freedom in good agreement with expectation). In family 5 the father is a regular heterozygote H/L and the mother is presumably HH/L; of the two children, one is H/L and the other probably HH/L. Other duplications could explain this family. Also here the aberrant density ratio in the two abnormal heterozygotes was observed with both enzymes (*Sst* I and *Ban* II).

DISCUSSION

It is well known that the IFN- α molecules are produced by a gene family. An exact estimate of the number of gene copies in the family is difficult, but an examination of the number and density of bands in Southern blots indicates that 15 might be a reasonable "guesstimate" for the number of copies of the IFN- α genes. Since the number of different sequences known today, as mentioned before, is close to 30, it seems inevitable to conclude that there must be polymorphism at the protein level. The demonstration of considerable polymorphism at the DNA level contributes to the credibility of this statement, but the DNA polymorphisms observed so far are not necessarily in the coding segments.

The IFN- β gene traditionally has been considered as a single-copy gene by most people and has so behaved in most

Table 3. Haplotypes and their frequency

Haplotype	IFN- α/β markers*					Number observed
	IFN- α			IFN- β		
	7.5	7.1	3.1	3.9/2.5	9.4/2.6	
1	—	—	—	3.9	9.4	7
2	—	—	—	3.9	2.6	4
1 or 2	—	—	—	3.9	?	8
3	—	—	—	2.5	2.6	2
4	—	—	—	2.5	9.4	0
5	—	—	+	3.9	9.4	1
6	—	—	+	3.9	2.6	2
7	—	+	—	3.9	9.4	2
8	—	+	—	3.9	2.6	1
9	+	—	—	3.9	9.4	3
1 or 9	?	—	—	3.9	9.4	1
10	?	—	—	3.9	2.6	5
9 or 10	+	—	—	3.9	?	4
11	+	—	—	2.5	2.6	2
12	+	—	—	2.5	9.4	0
11 or 12	+	—	—	2.5	?	4
13	+	—	+	3.9	9.4	2
14	+	—	+	3.9	2.6	10
13 or 14	+	—	+	3.9	?	1
15	?	—	+	2.5	2.6	1
5 or 14	?	—	+	3.9	2.6	1
Total						61

*Gene markers are indicated by restriction fragment size in kb. Presence (+) or absence (—) of a specific marker or the selection of a marker in a choice of two are indicated.

FIG. 2. Autoradiogram after hybridization to the IFN- β -specific probe of *Ban* II-digested DNAs from two families. (A) Patterns of one family: parents (lanes 1 and 2) and their children (lanes 3 and 4). (B) Pattern of another family: parents (lanes 1 and 2) and their children (lanes 3–12). The molecular sizes are indicated in kb. Densitometric evaluation of the bands revealed within the tracks normal, 2- and 3-fold intensities; among the tracks the same multiple intensities were found with some exceptions, but their departure from the expected intensities can be explained on the basis of DNA input.

of our families. There are indications, however, that in at least 2 unrelated individuals out of ≈ 28 , the IFN- β gene is duplicated. The duplication must be recent, and the duplicates are probably adjacent or very close since they always segregate together and can be recognized clearly because of inequality of density of the bands in heterozygotes. This phenomenon has been confirmed with different enzymes and with densitometric measurements using two probes: a chromosomal segment including flanking regions of the gene (10) and a cDNA containing only the coding region (7). On the other hand, the data do not answer the question of whether the duplicated IFN- β genes are functional in the particular individuals.

Sagar *et al.* (11) have maintained that the IFN- β gene exists also in multiple copies (in different chromosomes, however), a statement with which most other laboratories disagree. It is possible that some degree of multiplicity exists also for the IFN- β gene depending on which individual is chosen for cloning.

The markers we have studied (three for IFN- α gene and two for IFN- β gene) have always segregated as a block, and no recombinants have been observed either within or between the IFN- α/β gene markers. With the number of gametes we have tested, this would still be compatible with a frequency of recombination as high as a few unit %, but observations on linkage disequilibrium also indicate that IFN- α and IFN- β genes must be close, if not directly adjacent. In fact the disequilibrium between IFN- α and IFN- β genes is similar in magnitude or only slightly greater than that between IFN- α markers or between IFN- β markers. Two of 10 disequilibrium values are significantly different from zero, indicating that linkage disequilibrium is real; its average value would correspond theoretically (unpublished data) to a distance of several hundreds of kilobases, which is more than the length of the whole region if it were represented exclusively IFN- α and IFN- β genes. However, linkage disequilibrium depends on the amount of recombination, and this is correlated only loosely with physical length. In fact, recombination shows considerable local variation (unpublished data; ref. 12).

The close linkage of IFN- α and - β genes which we have observed is in good agreement with cytological observations (1), which place them close to each other. The position of the grains in the chromosome maps shown in the figures given by these authors actually suggest that the IFN- α and - β genes are in different but adjacent bands. It is doubtful, however, if the current state of the art of "in situ" hybridization allows such a degree of resolution. In any case, more data will be necessary also to improve the accuracy of recombination and linkage disequilibrium estimates.

The variability observed with IFN- α and - β genes is very

high, and the heterozygosity obtained for this region is among the highest observed, making the complex of IFN- α and - β genes one of the best genetic markers available today. Because one of the two IFN- β restriction fragment length polymorphisms is elicited by the same enzyme, *Msp* I, that reveals the IFN- α gene polymorphisms and because the bands of IFN- α markers do not overlap with those of IFN- β , one can use simultaneously the two probes together and, thus, increase the power of analysis beyond that possible with the IFN- α gene alone, testing one of the two IFN- β polymorphisms on the same filter. This has been done successfully, although for the results given above we have preferred to test IFN- α/β separately in order not to miss possible new variants.

The detection of a considerable number of variants of IFN- α and - β genes allows one to easily distinguish 12 or more human haplotypes and, hence, many more different genotypes, which may produce different quantities or qualities of the various IFNs. This makes it possible and easy to study the genetic basis of (i) individual differences in IFN production and (ii) resistance to diseases that might be controlled by IFNs.

We thank Eric Lundgren for stimulating discussions and Ellen Wijsman for making available her computer program. This work was supported by grants from KabiGen AB, Stockholm, Sweden, and by National Institutes of Health Grants GM 28428 and GM 20467.

1. Trent, J. M., Ohlsson, S. & Lawn, R. M. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 7809–7813.
2. Collins, J. (1983) in *Interferons*, eds. Burke, D. C. & Morris, A. G. (Cambridge Univ. Press, New York), pp. 35–66.
3. Lund, B., Edlund, T., Lindenmaier, W., Ny, T., Collins, J., Lundgren, F. & von Gabain, A. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 2435–2439.
4. Migone, N., Feder, J., Cann, H., van West, B., Hwang, J., Takahashi, N., Honjo, T., Piazza, A. & Gavalli-Sforza, L. L. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 467–471.
5. Steffen, D. & Weinburg, R. A. (1978) *Cell* **15**, 1003–1010.
6. Southern, E. M. (1975) *J. Mol. Biol.* **98**, 503–517.
7. Gross, G., Mayr, V., Bruns, W., Grosfeldt, F., Dahl, H.-H. M. & Collins, J. (1981) *Nucleic Acids Res.* **9**, 2495–2507.
8. Sehgal, P. G., Lester, T. M., Sagar, A. D., La Forge, K. S. & Inouye, M. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 3632–3635.
9. Riggan, C. H. & Pitha, P. M. (1982) *DNA* **1**, 267–271.
10. Cavalli-Sforza, L. L. & Bodmer, W. F. (1971) *The Genetics of Human Populations* (Freeman, San Francisco), p. 63.
11. Sagar, A. D., Sehgal, P. B., Lester, T. M., Inouye, M., Slate, D. L., Shulman, L. & Ruddle, F. H. (1984) *Science* **223**, 1312–1315.
12. Chakravarti, A., Buetow, K. H., Antonarakis, S. E., Waber, P. G., Boehm, C. D. & Kazazian, H. H. (1984) *Am. J. Hum. Genet.* **36**, 1239–1258.