```
CFTR 122-136     Y L G I G L C L L F I V R T L
                 : :   :   :   :   : : . :   :
GLNP 198-212     Y L I I T L V L S F I L R R L

CFTR 307-319     S S A F F F S G F F V V F
                 :     : : : . : : : .   :
COX   89-101     S E V F F F A G F F W A F

CFTR 701-713     I L N P I N S I R K F S I
                 : :   : . :   : : : . . :
NaCh 111-123     I L T P F N P I R K L A I

CFTR 1425-1442   D S I Q K L L N E R S L F R Q A I S
                 : : :   : :     : :   : :   :   . :
raf   578-595    D S I K K L R D E R P L F P Q I L S
```

GLNP, glutamine permease of *E. coli* [T. Nohno, T. Saito, J. Hóng, *Mol. Gen. Genet.* **205**, 260 (1986)]; COX, human cytochrome c oxidase polypeptide III [S. Anderson *et al.*, *Nature* **290**, 457 (1981)]; NaCh, rat brain sodium channel III (*32*); raf, the serine-threonine kinase proto-oncogene of *Xenopus laevis* (*31*). The first two sequences are within membrane spanning segments and probably reflect only coincidental arrangements of the hydrophobic residues suited to this function. In contrast, the latter two sequences are both in polar hydrophilic regions of the proteins. The large extent of amino acid conservation (11 of 13 residues) implies some functional relation between these short segments of the primary structure of the Na^+ channel and CFTR. Similarities between sequences at the same relative locations with respect to the COOH-termini of the raf kinase and CFTR suggest that they may also share at least a small facet of their structures and functions.

31. R. LeGuellec, K. LeGuellec, J. Paris, M. Philippe, *Nucleic Acids Res.* **16**, 10357 (1988).
32. M. Noda *et al.*, *Nature* **312**, 121 (1984); L. Salkoff *et al.*, *Science* **237**, 744 (1987).
33. P. R. Schofield *et al.*, *Nature* **328**, 221 (1987).
34. F. Sanger, S. Nicklen, A. R. Coulsen, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463 (1977).
35. D. Eisenberg, E. Schwarz, M. Komaramy, R. Wall, *J. Mol. Biol.* **179**, 125 (1984).
36. A. P. Feinberg and B. Vogelstein, *Anal. Biochem.* **132**, 6 (1983).
37. J. Rommens *et al.*, *Am. J. Hum. Genet.* **43**, 645 (1988).
38. S. J. Foote *et al.*, *Cell* **57**, 921 (1989).
39. J. P. McGrath and A. Varshavsky, *Nature* **340**, 400 (1989).
40. B. P. Surin, H. Rosenberg, G. B. Cox, *J. Bacteriol.* **161**, 189 (1985).
41. C. F. Higgins *et al.*, *Nature* **298**, 723 (1982).
42. E. Gilson, H. Nikaido, M. Hofnung, *Nucleic Acids Res.* **10**, 7449 (1982).
43. I. D. Hiles, M. P. Gallagher, D. H. Jamieson, C. F. Higgins, *J. Mol. Biol.* **195**, 125 (1987).
44. A. W. Bell *et al.*, *J. Biol. Chem.* **261**, 7652 (1986).
45. R. F. Doolittle *et al.*, *Nature* **323**, 451 (1986).
46. I. J. Evans and J. A. Downie, *Gene* **43**, 95 (1986).
47. D. R. Gill, G. H. Hatfull, G. P. C. Salmond, *Mol. Gen. Genet.* **205**, 134 (1986).
48. J. E. Walker, M. Saraste, M. J. Runswick, N. J. Gay, *EMBO J.* **1**, 945 (1982).
49. D. C. Fry, S. A. Kuby, A. S. Mildvan, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 907 (1986).

50. We thank O. Augustinas for the collection of tissues; T. Jensen and R. Baird for the culturing of epithelial cells; L. Naismith for the isolation of RNA; D. Kennedy and D. Markiewicz for technical assistance and M. Buchwald and M. Dean for discussions. Supported by NIH grants DK39690 (F.S.C.) and DK34944 (L.C.T.), the Cystic Fibrosis Foundation (U.S.A.), the Canadian Cystic Fibrosis Foundation, and the Sellers Fund. J.M.R. holds a postdoctoral fellowship from the Medical Research Council (MRC) of Canada and F.S.C. is an Associate Investigator of the Howard Hughes Medical Institute.

7 August 1989; accepted 18 August 1989

Identification of the Cystic Fibrosis Gene: Genetic Analysis

BAT-SHEVA KEREM, JOHANNA M. ROMMENS, JANET A. BUCHANAN, DANUTA MARKIEWICZ, TARA K. COX, ARAVINDA CHAKRAVARTI, MANUEL BUCHWALD, LAP-CHEE TSUI

Approximately 70 percent of the mutations in cystic fibrosis patients correspond to a specific deletion of three base pairs, which results in the loss of a phenylalanine residue at amino acid position 508 of the putative product of the cystic fibrosis gene. Extended haplotype data based on DNA markers closely linked to the putative disease gene locus suggest that the remainder of the cystic fibrosis mutant gene pool consists of multiple, different mutations. A small set of these latter mutant alleles (about 8 percent) may confer residual pancreatic exocrine function in a subgroup of patients who are pancreatic sufficient. The ability to detect mutations in the cystic fibrosis gene at the DNA level has important implications for genetic diagnosis.

ALTHOUGH THE FREQUENCY OF CYSTIC FIBROSIS (CF) IS not uniformly high among all Caucasian populations, a consensus estimate is that it occurs once in 2000 live births (*1*). On the basis of the autosomal recessive mode of inheritance for this disease, a mutant allele frequency of 0.022 may be derived. Several different mechanisms, including high mutation rate (*2*), heterozygote advantage (*3*), genetic drift (*4*), multiple loci (*5*), and reproductive compensation (*6*), have been proposed in attempts to explain the high incidence and, indirectly, the nature of the CF mutations. Although some of these hypotheses could not be further addressed because of the lack of knowledge about the basic defect in CF, several important observations have been made during the past few years through genetic analysis of the families of affected individuals (*7–20*).

Extensive linkage analysis provides evidence for the existence of a single CF locus on human chromosome 7 (region q31) (*7–10, 21*). The detection of allelic and haplotype association between the CF

B. Kerem, J. M. Rommens, J. A. Buchanan, D. Markiewicz, M. Buchwald and L.-C. Tsui are in the Department of Genetics, Research Institute, The Hospital for Sick Children, Toronto, Ontario M5G 1X8, Canada. T. K. Cox and A. Chakravarti are in the Department of Human Genetics, University of Pittsburgh, Pittsburgh, PA 15261. M. Buchwald and L.-C. Tsui are also members of the Departments of Medical Genetics and Medical Biophysics, University of Toronto, Toronto, Ontario M5S 1A8, Canada.

locus (*CF*) and closely linked DNA markers further suggests a small number of mutations in CF (*12–16, 19*). A single mutational event may account for most CF mutations in Northern European populations (*12*); another mutation appears to be prevalent in some Southern European populations (*15*). Additional haplotype data indicate that there could be several other mutations (*15, 16, 19, 20*). The two clinical subgroups of CF patients, those with pancreatic insufficiency (CF-PI) and those with pancreatic sufficiency (CF-PS), can be explained by different mutations on the basis of family studies (*18*) and haplotype data (*19, 20*). Patients with CF-PI, which constitute the larger proportion (about 85 percent) of the CF population, appear to be genetically more homogeneous than those with CF-PS (about 15 percent); most individuals with CF-PS appear to be compound heterozygotes (*19*). We must emphasize, however, that these studies were based on linked DNA markers whose exact relation to *CF* was not entirely certain. Further understanding of the pathophysiology, haplotype association, and population distribution of CF would require detailed molecular knowledge about the mutations.

The extensive genetic and physical mapping data made it possible for molecular cloning studies to be focused on a small segment of DNA on chromosome 7 (*12, 13, 17, 21, 22*). A putative CF gene has now been identified through chromosome walking and jumping (*23*) and cDNA clones have been isolated and characterized (*24*). Because of the lack of chromosome deletions and rearrangements in CF and the lack of a well-developed functional assay for the CF gene product, the identification of the CF locus required a detailed characterization of the locus itself and comparison between the CF and normal (N) alleles. Random, phenotypically normal, individuals could not be included as controls in the comparison because of the high frequency of symptomless carriers in the population. As a result, only parents of CF patients, each of whom by definition carries an N and a CF chromosome, were suitable for the analysis. Moreover, because of the strong allelic association observed between *CF* and some of the closely linked DNA markers (*12–15, 21*), it was necessary to exclude the possibility that sequence differences detected between N and CF were polymorphisms associated with the disease locus.

Identification of RFLP's and family studies. To determine the relation of each of the DNA segments isolated from the chromosome walking and jumping experiments to *CF*, restriction fragment length polymorphisms (RFLP's) were identified and used to study families where crossover events between *CF* and other flanking DNA markers had previously been detected (*25, 26*). In all, 18 RFLP's were detected in the 500-kb region of chromosome 7; 17 of them (from E6 to CE1.0) are listed in Table 1; some of them corresponded to markers previously reported (*12, 13*). Five of the RFLP's were identified with cDNA and genomic DNA probes derived from the putative CF locus, namely 10-1X.6, T6/20, H1.3, and CE1.0 (*23, 24*). The RFLP data (Table 1) are presented with markers in the regions of *MET* and *D7S8* included for comparison. The physical distances between these markers as well as their relation to *MET* and *D7S8* are shown in Fig. 1.

Because families with informative crossovers only gradually became available [through the courtesy of investigators (*27*)] over the course of the study, not all the DNA markers were examined in these families (*28*). The recombination breakpoints were localized for the two families [out of four that showed a crossover between *D7S23* and *CF* (*26*)] that were informative for the DNA markers tested. One of them showed a crossover between the sites defined by JG2E1 (also known as KM19) and E2.6 (*29*), and the other between JG2E1 and E4.1 (Mp6d.9) (*30*) (Fig. 1). These results indicated that *CF* mapped to the telomeric side of JG2E1. The two other families were not informative in the analysis. Only a single family (SLC1380) was documented to show a recombination event between *CF* and *D7S8* (*25*), and more recently the crossover point was narrowed to between *CF* and *D7S424*, a jump clone derived from *D7S8* (*31*); no crossover was detected with the above-described DNA markers in this family. *CF* was therefore delimited to the region between JG2E1 and *D7S424*, spanning less than 900 kb (Fig. 1). Because of the small number of recombinant families available for the analysis, which was expected from the close distance between the markers and *CF*, and because of the possibility of misdiagnosis, alternative approaches were necessary in further fine mapping of the CF locus.

Allelic association. Allelic association (linkage disequilibrium)

Table 1. RFLP's associated with the CF locus.

Probe name	Enzyme	Fragment (kb)	N*	CF-PI*	|A|†	|Δ|‡	Source
metD	Ban I	7.6	28	48	0.60	0.10	(*11*); TS§
		6.8	59	25			
metD	Taq I	6.2	74	75	0.66	0.06	(*9*); TS
		4.8	19	4			
metH	Taq I	7.5	45	49	0.35	0.05	(*9*); TS
		4.0	38	20			
E6	Taq I	4.4	58	62	0.45	0.06	(*19*); TS
		3.6	42	17			
E7	Taq I	3.9	40	16	0.47	0.07	TS
		3+0.9	51	57			
pH131	Hinf I	0.4	81	33	0.73	0.15	(*13*); TS
		0.3	18	47			
W3D1.4	Hind III	20	82	33	0.68	0.13	(*19*); TS
		10	22	47			
H2.3A (XV2C)	Taq I	2.1	39	53	0.64	0.09	(*12*); TS
		1.4	37	11			
EG1.4	Hinc II	3.8	31	69	0.89	0.17	TS
		2.8	56	7			
EG1.4	Bgl II	20	27	69	0.89	0.18	TS
		15	62	9			
JG2E1 (KM19)	Pst I	7.8	69	10	0.88	0.18	(*12, 19*); TS
		6.6	30	70			
E2.6 (E.9)	Msp I	13	34	6	0.85	0.14	TS
		8.5	26	55			
H2.8A	Nco I	25	22	55	0.87	0.18	TS
		8	52	9			
E4.1 (Mp6d.9)	Msp I	12	37	8	0.77	0.11	(*29*); TS
		8.5+3.5	38	64			
J44	Xba I	15.3	40	70	0.86	0.13	TS
		15+0.3	44	6			
10-1X.6	Acc I	6.5	67	15	0.90	0.24	TS
		3.5+3	14	60			
10-1X.6	Hae III	1.2	14	61	0.91	0.25	TS
		0.6	72	15			
T6/20	Msp I	8	56	66	0.51	0.54	TS
		4.3	21	8			
H1.3	Nco I	2.4	53	7	0.87	0.15	TS
		1+1.4	35	69			
CE1.0	Nde I	5.5	81	73	0.41	0.03	TS
		4.7+0.8	8	3			
J32	Sac I	15	21	24	0.17	0.02	(*31*); TS
		6	47	38			
J3.11	Msp I	4.2	36	38	0.29	0.04	(*10*); TS
		1.8	62	36			
J29	Pvu II	9	26	36	0.36	0.06	(*31*); TS
		6	55	36			

*The number of N and CF-PI (CF with pancreatic insufficiency) chromosomes was derived from the parents in the families used in our linkage analysis (*36*). †Standardized association (*A*), which is less influenced by the fluctuation of DNA marker allele distribution among the N chromosomes, is used here for the comparison. Yule's association coefficient (A) = $(ad - bc)/(ad + bc)$, where a, b, c, and d are the numbers of N chromosomes with DNA marker allele 1, CF with 1, N with 2, and CF with 2, respectively. Relative risk can be calculated from the relation $RR = (1 + A)/(1 - A)$ or its reverse. ‡Allelic association (Δ), calculated according to (*33*) on the basis of a frequency of 0.02 for CF chromosomes in the population, is included for comparison. §TS, this study.

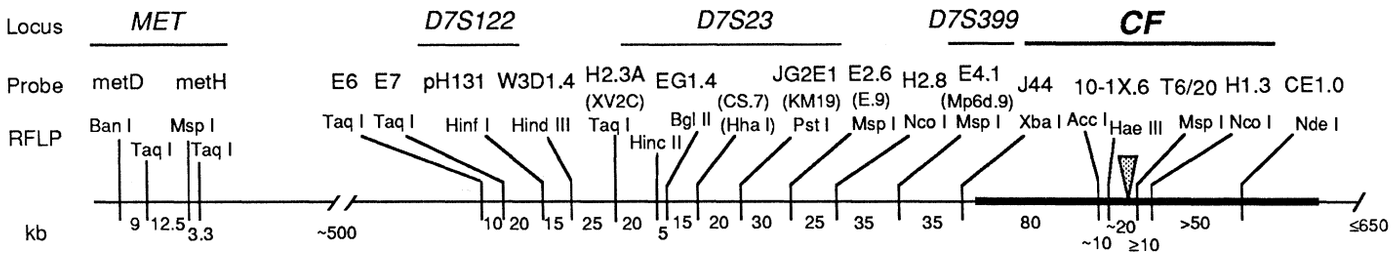
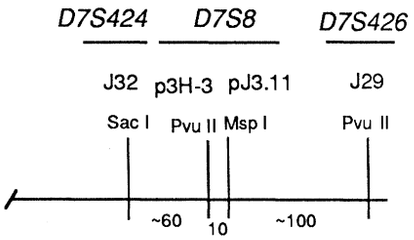


Fig. 1. Map of the RFLP's closely linked to the CF locus. Details of the RFLP's are shown in Table 1. The inverted triangle indicates the location of the ΔF_{508} mutation.



has been detected for many closely linked DNA markers (*32, 33*). Although the utility of using allelic association for measuring genetic distance is uncertain, an overall correlation has been observed between *CF* and the flanking DNA markers. A strong association with *CF* was noted for the closer DNA markers *D7S23* (*12, 14*) and *D7S122* (*13*), whereas little or no association was detected for the more distant markers *MET*, *D7S8*, or *D7S424* (*21, 31*; Fig. 1).

The degree of association between DNA markers and *CF* (as measured by the Yule's association coefficient) increased from 0.35 for metH and 0.17 for J32 to 0.91 for 10-1X.6 [only CF-PI families were used in the analysis as they appeared to be genetically more homogeneous than CF-PS (*19*)]. The association coefficients appeared to be rather constant over the 300-kb interval between EG1.4 and H1.3; the fluctuations detected at several locations, most notably at H2.3A, E4.1, and T6/20, were probably due to the variation in the allelic distribution among the N chromosomes (Table 1). These data are therefore consistent with the result from the study of recombinant families (Fig. 1). A similar conclusion could also be made by inspection of the extended DNA marker haplotypes associated with the CF chromosomes (as discussed below). However, the strong allelic association detected over the large physical distance between EG1.4 and H1.3 did not allow further refined mapping of the CF locus. Since J44 was the last genomic DNA clone isolated by chromosome walking and jumping before a cDNA clone was identified (*24*), the strong allelic association detected for the JG2E1-J44 interval prompted us to search for candidate gene sequences over this entire interval. The highest degree of allelic association was in fact detected between *CF* and the two RFLP's detected by 10-1X.6, a region near the major CF mutation.

Strong allelic association was also detected among subgroups of RFLP's on both the CF and N chromosomes, namely, between adjacent markers E6 and E7, between pH131 and W3D1.4, between the Acc I and Hae III polymorphic sites detected by 10-1X.6 and among EG1.4, JG2E1, E2.6 (E.9), H2.8, and E4.1 (*34*). The two groups of distal markers in the MET and D7S8 regions also showed some degree of linkage disequilibrium among themselves, but they showed little association with markers from E6 to CE1.0, consistent with the distant locations for *MET* and *D7S8*. The lack of association between DNA markers that are physically close may indicate the presence of recombination hot spots (*33, 34*). Examples of these potential hot spots are the region between E7 and pH131, around H2.3A, and between J44 and the regions covered by the probes 10-1X.6 and T6/20 (Fig. 1). These regions, containing frequent recombination breakpoints, were useful in the subsequent analysis of extended haplotype data for the CF region.

The major CF mutation. Molecular cloning experiments have allowed us to identify a gene in the J44-*D7S424* interval (*23, 24*). Sequence analysis of overlapping cDNA clones of this gene predicted a protein with properties consistent with membrane association and nucleotide binding (*24*). Comparison of the nucleotide sequences of the cDNA clones derived from N and CF individuals revealed a 3-bp deletion in the CF gene, resulting in the loss of a single phenylalanine residue (position 508) in the predicted amino acid sequence (*24*).

Table 2. Distribution of CF and non-CF (N) chromosomes with and without the 3-bp deletion (ΔF_{508}). The data for the CF-PI and CF-PS chromosomes were derived from the CF families included in our linkage analysis (*36*). These families were originally selected without our knowledge about PI or PS; the 14 CF-PS families subsequently identified were not included as part of this calculation. The unclassified CF chromosomes were derived from CF families for which pancreatic function data were not available.

	Chromosomes	
	CF	N
Without the deletion		
CF-PI	24	
CF-PS	9	
Unclassified	36	
Subtotal	69 (32%)	198
With the deletion		
CF-PI	62	
CF-PS	5	
Unclassified	78	
Subtotal	145 (68%)	0
Total	214	198

To investigate the proportion of CF patients carrying this deletion (ΔF_{508}), we used genomic DNA samples from patients and their parents. Each sample was amplified with oligonucleotide primers flanking the mutation in a polymerase chain reaction (PCR) (*35*) and hybridized to ^{32}P-labeled oligonucleotides specific for the normal and the putative mutant sequences (Fig. 2). The result of this analysis showed that 68 percent (145/214) of CF chromosomes in the general patient population had the ΔF_{508} mutation (Table 2). In contrast, none (0/198) of the N chromosomes had the deletion (Table 2) ($\chi^2 = 207$, $P < 10^{-57.5}$!), suggesting that this sequence alteration is specific to CF and that it is the major mutation causing the disease. No recombination has been detected between the ΔF_{508} mutation site and *CF*. However, the possibility that this sequence difference corresponded to a tightly linked DNA polymorphism had yet to be excluded.

Haplotype analysis. Extended haplotypes based on 23 DNA markers were generated for the CF and N chromosomes in the collection of families previously used for linkage analysis (*36*). Assuming recombination between chromosomes of different haplotypes, it was possible to construct several lineages of the observed

Table 3. DNA marker haplotypes spanning the CF locus.

Haplotypes*									CF†				N
Region									ΔF_{508}		Others		
1	2	3	4	5	6	7	8	9	PI	PS	PI	PS	
						Group Ia							
A	A	A	A	A	A	A	A	A	10	1			
A	A	A	A	A	A	–	A	A	3				
A	A	A	A	–	A	A	–	A	1				
A	A	A	A	–	–	A	–	A					1
A	A	A	A	A	A	A	A	B	10				1
A	A	–	A	A	A	A	A	B	4				
A	A	A	A	–	A	A	A	B	1				
A	A	–	A	A	A	A	A	C	1				
B	A	A	A	A	A	A	A	A	4				
B	A	–	A	A	A	A	A	A	1				
B	A	A	A	–	A	A	A	A		1			
B	A	A	A	A	A	A	A	–	1				
B	A	A	A	–	A	A	–	A	1				
A	B	A	A	A	A	A	A	A	1				
A	D	A	A	A	A	A	A	A	1				
A	G	A	A	A	A	A	A	A	1				
B	B	A	A	A	A	A	A	A	1				
B	C	A	A	A	A	A	A	A	2				
E	B	A	A	–	–	A	–	A	1				
D	B	A	A	–	A	–	A	A	1				
D	B	B	A	A	A	A	A	A	1				
B	A	–	A	A	A	A	A	B	1				
C	A	–	A	A	A	A	A	B	1				
A	D	A	A	A	A	A	A	B	1				
D	C	A	A	A	A	A	A	B		1			
A	D	A	A	–	A	A	A	B	1				
D	D	–	A	A	A	A	A	B					1
B	B	–	A	A	A	–	A	B	1				
A	B	A	A	A	A	A	A	E	2				
A	B	–	A	A	A	A	A	E	1	1			
A	E	B	A	A	A	A	A	E	1				
A	C	A	A	A	A	A	A	B	1				
A	C	–	C	–	A	A	A	B		1			
A	B	A	B	A	A	A	–	A					1
B	C	B	A	–	A	A	A/D	B	1				
						Group Ib							
A	C	–	A	A	A	A	A	A				1	
A	C	A	A	A	A	A	A	–			1		
D	C	–	A	A	A	A	A	B			1		
D	C	A	A	A	A	A	A	D					1
F	C	–	A	A	A	A	A	B			1		
B	C	A	A	A	A	A	A	B			3		
						Group Ic							
B	C	A	B	C	A	A	D	A					1
B	C	A	B	C	A	A	D	B			1		
F	C	A	B	C	A	A	D	B					1
F	A	A	B	C	A	A	D	B					1
A	B	A	B	C	A	A	D	B					1
B	B	A	B	C	A	A	D	B					1
B	D	A	B	C	A	–	D	C					1
A	B	A	B	A	A	–	D	A					1
						Group Id							
D	B	A	A	A	A	A	C	A					1
B	C	B	C	A	A	A	C	B					1
						Subtotals			57	5	7	1	14
						Group IIa							
B	A	–	B	B	B	A	C	B			1		
–	B/C	B	B	B	B	A	C	B			1		
B	A	–	B	–	B	A	A/C	B				1	
A	B	B	B	B	B	A	C	B			1		
B	B	B	B	B	B	A	C	A					3
A	C	B	B	B	B	A	C	A					1
A	C	B	B	B	B	–	C	A					1
F	C	B	B	B	B	A	C	A					1
A	C	B	B	B	B	A	C	B					1
A	C	–	B	B	B	–	C	C					1
B	C	B	B	–	B	A	C	C				1	
B	C	B	B	B	B	A	C	B					1
B	C	B	B	B	B	A	C	A					1
B	C	B	B	B	B	A	C	D					1
B	C	–	B	B	B	A	C	B					1
B	C	B	B	B	B	–	C	B					1
D	C	B	B	B	B	A	C	B					2
D	–	B	B	–	B	A	C	B					1
F	C	B	B	B	B	A	C	B					1
C	C	–	B	B	B	A	C	B					1

Table 3 continued.

Haplotypes*									CF†				N
Region									ΔF_{508}		Others		
1	2	3	4	5	6	7	8	9	PI	PS	PI	PS	
A	A	A	B	B	B	A	C	B					1
B	G	A	B	B	B	A	C	B					1
F	A	–	B	B	B	A	C	B					1
B	H	–	B	B	B	A	C	B					1
B	B	–	B	B	B	A	C	B					1
A	B	A	B	B	B	A	C	B					1
A	B	A	B	B	B	A	C	A					1
F	D	A	B	B	B	A	C	B					1
C	D	A	B	B	B	A	C	A					1
B	D	A	B	B	B	A	C	A					1
B	C	A	B	B	B	A	C	A					2
A	C	A	B	B	B	A	C	B					1
A	C	A	B	B	B	–	C	B					1
A	C	A	B	B	B	A	C	C					1
B	C	A	B	B	B	A	C	B					1
D	B/C	–	B	B	B	A	C	A					2
C	C	A	B	B	B	A	C	A				1	
D	B	–	B	B	B	A	A/C	B					1
D	B	A	B	–	B	A	C	B					1
A	G	A	B	B	B	A	C	A					1
B	C	–	B	B	B	A	A/C	A					1
A	C	B	D	B	B	A	C	B			1		
A	C	–	D	–	B	A	C	B					1
B	B	B	E	B	B	A	C	C					1
F	D	A	B	B	B	A	C	C				1	
A	A	A	A	A	B	A	C	D					1
–	B/C	A	B	C	B	A	C	B					1
						Group IIb							
A	C	A	B	B	B	A	B	E			1		
A	C	–	B	B	B	A	B	B			1		
						Group IIc							
B	D	–	B	–	B	A	A	A					1
						Subtotals			0	0	6	4	45
						Group IIIa							
B	C	B	A	A	C	B	A	B	1				
						Group IIIb							
B	A	B	A	A	C	B	A	B			1		
B	C	B	A	A	C	B	A	A			1		
B	C	B	A	A	C	B	A	B					1
B	C	–	A	A	C	B	A	B					2
A	B	–	A	A	C	B	A	B					1
A	B	–	A	A	C	B	A	C					1
B	B	B	A	A	C	B	A	B					2
D	C	B	A	A	C	B	A	A				1	
A	B	B	C	A	C	B	A	B					1
B	B	A	A	A	C	B	A	B			2		1
B	B	–	A	A	C	B	A	B			1		1
B	B	A	A	A	C	B	A	A				1	
D	A	A	A	A	C	B	A	B					1
D	C	A	A	A	C	B	A	B					2
A	C	–	A	A	C	B	A	B			1		1
D	B	A	A	A	C	–	A	C					1
						Group IIIc							
A	A	A	B	B	C	B	A	–			1		
F	B	B	B	B	C	B	A	B					1
D	B	B	B	B	C	B	A	A					1
						Subtotals			1	0	7	2	17
						Group IV							
F	C	B	A	A	C	B	C	A				1	
B	C	A	A	A	C	B	C	–					1
A	B	A	A	A	C	–	C	B					1
A	H	B	A	–	C	–	C	B					1
D	B	B	B	B	C	B	C	B					1
						Subtotals			0	0	0	1	4
						Group V							
B	C	B	B	B	C	A	C	A			1		
A	C	B	B	–	–	A	–	A			1		
B	B	B	B	B	C	A	C	B					1
B	C	B	B	B	C	A	C	B					1
B	C	–	B	B	C	A	C	B					1
D	–	A	B	B	C	–	C	B					1
B	C	A	B	C	C	A	C	A				1	
B	C	–	B	C	C	–	C	D					1
						Subtotals			0	0	2	1	5

Table 3 continued.

Haplotypes*									CF†				N
Region									ΔF_{508}		Others		
1	2	3	4	5	6	7	8	9	PI	PS	PI	PS	
Others													
B	C	B	A	A	B	B	A	B					1
B	C	B	A	A	D	B	A	B					1
B	C	B	E	B	A	B	D	A					1
B	C	A	B	B	E	–	C	–					1
B	D	B	B	B	F	A	C	B					1
A	C	–	A	A	C	B	D	A					1
G	B	B	A	A	B/C	A	A/D	B					1
Subtotals									0	0	0	0	7
Unclassified													
–	–	–	–	–	–	–	–	–	4	10	2	18	6
								Total	62	15	24	27	98

†Number of chromosomes scored in each class: CF-PI(ΔF_{508}), CF chromosomes from CF-PI patients with the ΔF_{508} mutation; CF-PS(ΔF_{508}), CF chromosomes from CF-PS patients with the ΔF_{508} mutation; CF-PI (others), other CF chromosomes from CF-PI patients; CF-PS (others), other CF chromosomes from CF-PS patients; N, normal chromosomes derived from carrier parents. *The extended haplotype data are derived from the CF families used in previous linkage studies (*36*) with additional CF-PS families collected subsequently (*19*). The data are shown in groups (regions) to reduce space. The regions are assigned primarily according to pairwise association data (*34*) with regions 6 to 8 spanning the putative CF locus. The ΔF_{508} mutation is located between region 6 and 7. A dash (–) is shown at the region where the haplotype has not been determined because of incomplete data or inability to establish phase. Alternative haplotype assignments are also given where data are incomplete. Unclassified includes those chromosomes with more than three unknown assignments. The haplotype definitions for each of the nine regions are shown below; 1 refers to the larger allele of the RFLP; 2 refers to the smaller allele.

Region 1

	metD Ban I	metD Taq I	metH Taq I
A	1	1	1
B	2	1	2
C	1	1	2
D	2	2	1
E	1	2	–
F	2	1	1
G	2	2	2

Region 2

	E6 Taq I	E7 Taq I	pH131 Hinf I	W3D1.4 Hind III
A	1	2	2	2
B	2	1	1	1
C	1	2	1	1
D	2	1	2	2
E	2	2	2	1
F	2	2	1	1
G	1	2	1	2
H	1	1	2	2

Region 3

	H2.3A Taq I
A	1
B	2

Region 4

	EG1.4 Hinc II	EG1.4 Bgl II	JG2E1 Pst I
A	1	1	2
B	2	2	1
C	2	2	2
D	1	1	1
E	1	2	1

Region 5

	E2.6 Msp I	E2.8 Nco I	E4.1 Msp I
A	2	1	2
B	1	2	1
C	2	2	2

Region 6

	J44 Xba I	10-1X.6 Acc I	10-1X.6 Hae III
A	1	2	1
B	2	1	2
C	1	1	2
D	1	2	2
E	2	2	2
F	2	2	1

Region 7

	T6/20 Msp I
A	1
B	2

Region 8

	H1.3 Nco I	CE1.0 Nde I
A	2	1
B	1	2
C	1	1
D	2	2

Region 9

	J32 Sac I	J3.11 Msp I	J29 Pvu II
A	1	1	1
B	2	2	2
C	2	1	2
D	2	2	1
E	2	1	1

CF chromosomes and, also, to predict the location of the disease locus (*34, 37*).

To obtain further information useful for understanding the nature of different CF mutations, ΔF_{508} was correlated with the extended DNA marker haplotypes (*37*). Five major groups of N and CF haplotypes could be defined by the RFLP's within or immediately adjacent to the putative CF locus (regions 6 to 8) (Table 3). Most recombinations between haplotypes occurred between regions 1 and 2 and between 8 and 9, again in good agreement with the relatively long physical distance between these regions. Other less frequent breakpoints occurred between short distance intervals, and they generally corresponded to the hot spots identified by pairwise allelic association studies (*34*). A striking result of our study is that ΔF_{508} associated almost exclusively with group I, the most frequent CF haplotype, consistent with our hypothesis that this deletion constitutes the major mutation in CF. More important, while ΔF_{508} was detected in 89 percent (62/70) of the CF chromosomes with the AA haplotype (corresponding to regions 6 and 7 flanking the deletion), no deletions were found in the 14 N chromosomes within the same group ($\chi^2 = 47.3$, $P < 10^{-4}$). The ΔF_{508} mutation is therefore not a common sequence polymorphism associated with the core of the group I haplotype (Table 3).

One of the CF chromosomes, detected by the specific oligonucleotide probe for ΔF_{508}, belonged to a different haplotype group (group III). None of the nine other CF chromosomes or 17 N chromosomes within group III hybridized to the probe. This specific hybridization result suggests that the mutation harbored on this chromosome is similiar to ΔF_{508}. Although recombination or gene conversion are possible mechanisms to explain the presence of this deletion on a haplotype that did not belong with group I, it is more likely that this group III chromosome represents a recurrent mutation event, a situation similar to the β^S and β^E mutations at the β-globin locus (*38*). Together, the results of the oligonucleotide hybridization study and the haplotype analysis provide a strong indication that the gene locus described above is the CF locus and that the 3-bp deletion (ΔF_{508}) is the most common mutation in CF.

Other CF mutations. Haplotype association with specific mutations has been demonstrated in studies of other genetic disorders (*39*). As indicated by these examples, knowledge of the haplotypes associated with the disease chromosomes facilitates studies on the molecular defects of each of the respective mutations.

The association of ΔF_{508} with one common and one rare CF haplotype provided further insight into the number of mutational events that could contribute to the present patient population. On the basis of the extensive haplotype data, the two original chromosomes in which the ΔF_{508} mutation occurred are likely to carry the haplotype –AAAAAAA– (group Ia) and –CBAACBA– (group IIIa), as defined (Table 3). The other CF chromosomes in group I carrying the deletion are probably recombination products derived from the original chromosome. If the CF chromosomes in each haplotype group are considered to be derived from the same origin, only three to four additional mutational events would be predicted (Table 3); however, since many of the CF chromosomes in the same group are different from each other, further subdivision within each group is possible. As a result, a greater number of independent mutational events can be considered, and the data suggest that at least seven additional, putative mutations also contribute to the CF-PI phenotype. The mutations leading to the CF-PS subgroup are probably more heterogeneous and are considered below.

The seven additional CF-PI mutations are represented by the haplotypes: –CAAAAAA– (group Ib), –CABCAAD– (group Ic), –––BBBAC– (group IIa), –CABBBAB– (group IIb), –––AACBA– (group IIIb), –AABBCBA– (group IIIc), and –CBBBCAC– (group V). Although the molecular defect in each of these mutations has yet

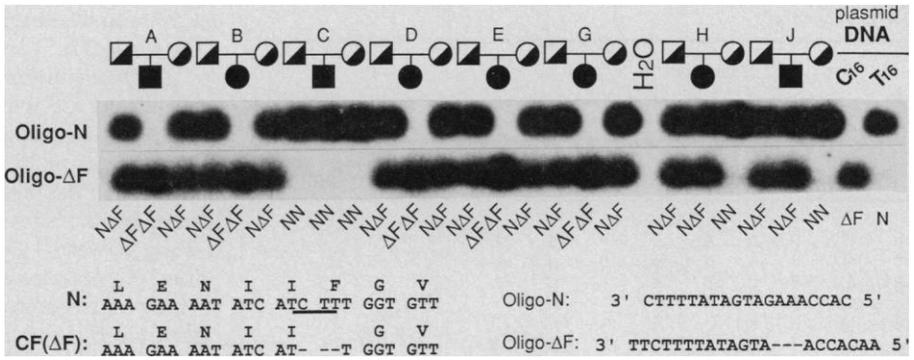


Fig. 2. Detection of the ΔF_{508} mutation by oligonucleotide hybridization. Autoradiographs show the hybridization results of genomic DNA from representative CF families with the two specific oligonucleotide probes as indicated. Oligo-N detects the normal DNA sequence and oligo-ΔF detects the mutant sequence. Genomic DNA sample from each family member was amplified by PCR (*35*), and the products were separated by electrophoresis on a 1.4 percent agarose gel, and transferred to Zetaprobe (Bio-Rad) membrane according to standard procedures. The membrane was hybridized with ^{32}P-labeled oligonucleotide probes, washed, and exposed to Kodak XAR film as described (*35*). Samples without DNA (H_2O) and plasmid DNA, T16 (N cDNA) and C16 (cDNA with the ΔF_{508} deletion), were included as controls. ΔF is the abbreviation for ΔF_{508}. Families B, D, E, and H are CF-PI (diagnosis for the other families not available). The illustration is based on the assumption that the triplet CTT was deleted; the sequencing data do not allow distinction between deletion of these nucleotides or other combinations.

Table 4. Population analysis of CF-PI and CF-PS.

	Assumed genotype*	Predicted frequency†	Observed‡	Expected§
Pancreatic Insufficient (PI)	$\Delta F_{508}\Delta F_{508}$	0.459	21	21.1
	ΔF_{508}S	0.331	14	15.2
	SS	0.060	4	2.7
	Total	0.850	39	
Pancreatic Sufficient (PS)	ΔF_{508}M	0.106	15‖	14.8
	SM	0.038	6	6.2
	MM	0.006		
	Total	0.150	21	

*Allele designations: ΔF_{508}, the 3-bp deletion; S, uncharacterized severe mutant alleles; M, uncharacterized mild mutant alleles. †Assuming that the CF-PI mutant phenotype is recessive to the CF-PS mutant phenotype, the frequency of CF-PI mutant alleles, including the 3-bp deletion, could be estimated from the observed proportion of CF-PI patients in our CF clinic (*18*), that is, $(0.85)^{1/2} = 0.92$. The observed allele frequency for ΔF_{508} in the total CF population is 0.68 (Table 2); the frequency for S is $0.92 - 0.68 = 0.24$; the frequency for M is $1 - 0.92 = 0.08$. The frequency for each genotype was then calculated from the Hardy-Weinberg law. ‡The number of CF-PI and CF-PS patients in each category was obtained by oligonucleotide hybridization analysis as illustrated in Fig. 2. The patients were from the CF families used in our linkage analysis (*36*) with 14 additional CF-PS patients or families from a subsequent study (*19*). §The expected numbers were calculated for CF-PI and CF-PS after normalization within each group. The χ^2 of fit is 0.86, df = 3, $0.74 < P < 0.90$. ‖This number is higher than would be expected (15 observed compared to 9.6) if ΔF_{508} is in Hardy-Weinberg equilibrium among all CF chromosomes ($\chi^2 = 6.48$, df = 1; $P < 0.011$).

to be defined, none of these mutations severely affect the region corresponding to the oligonucleotide binding sites used in the PCR-hybridization experiment (Fig. 2 and legend).

Pancreatic sufficiency. CF-PS is defined clinically as sufficient pancreatic exocrine function for digestion of food; however, the level of residual pancreatic enzyme activity varies among patients (*1, 40*). Our previous haplotype data suggested that the CF-PI and CF-PS patients have different mutant alleles (*19*). Although the basic biochemical defect in CF has yet to be defined, it is possible that the residual pancreatic enzyme activity in CF-PS patients is a direct reflection of the activity of the mutant CF gene product. Thus, the residual exocrine function conferred by a mild (CF-PS) allele, although much lower than that of the normal gene product, would constitute a dominant phenotype over that of more severe (CF-PI) mutations with little or no function. Therefore, only patients carrying two copies of severe alleles would be CF-PI and patients carrying one or two mild alleles would be CF-PS.

To test this hypothesis, we used the information on the proportion of CF patients carrying the ΔF_{508} deletion. If we assume that a severe mutation is recessive to a mild mutation and a distribution of CF alleles among the patient population according to the Hardy-Weinberg law, the frequency could be estimated to be ~0.92 for the severe alleles and ~0.08 for the mild alleles (M) (Table 4). Since most CF-PI patients were homozygous for ΔF_{508}, it was reasonable to assume that this mutation corresponded to one of the severe alleles. Given the observed frequency of ΔF_{508} (0.68) in the studied CF population (Table 2), the frequency of the remaining severe alleles could be derived. The proportion of $\Delta F_{508}\Delta F_{508}$, SS, MM, ΔF_{508}S, ΔF_{508}M, and SM patients was then calculated. Since individuals with SM and MM could not be distinguished phenotypically or genotypically, they were combined in the analysis. The observed frequencies for all five groups of patients were as expected from this hypothesis (Table 4).

The above analysis thus provides strong support for our hypothesis that CF-PI is due to the presence of two severe alleles and that a CF-PS patient carries either a single severe allele or two mild alleles. This model also explains the lower frequency of ΔF_{508} in the CF-PS than in the CF-PI population and the excess number of CF-PS patients with one copy of the deletion (Table 4).

Given the predicted dominant phenotype conferred by the M alleles, it was necessary to examine the CF chromosomes in CF-PS patients individually in order to identify those carrying the M alleles. Five of the seven representative CF-PS patients carry one copy of the ΔF_{508} deletion; at least five different haplotypes could be assigned to the other CF chromosomes (Table 5). These latter observations further support our previous suggestion that the majority of CF-PS patients are compound heterozygotes (*19*). Further delineation of these and other CF haplotypes observed in our study would require a larger patient population or a more detailed characterization of CF mutations (or both).

Meconium ileus (MI), which occurs in 5 to 10 percent of newborns with CF, is generally ascribed to failure of pancreatic enzyme secretion and digestion of intraluminal contents in utero (*1*). Although only six patients were identified to have MI in our study (*41*), all of them belong to the CF-PI subgroup, five of them homozygous and one heterozygous for ΔF_{508} (Table 5). It is therefore tempting to speculate that homozygous ΔF_{508} (or equivalent severe mutations) may be a prerequisite for development of MI. Moreover, since MI only occurs in a small proportion of CF patients and with only 30 percent concordance within families (*42*), it is probable that this condition is also determined by other genetic or nongenetic factors.

Implications for genetic diagnosis. Previous DNA-based genetic testing for CF has only been available to families with affected children and to their close relatives (*14, 43*). Knowledge of the CF mutations at the DNA sequence level should permit testing any random individual. On the basis of our estimate (Table 4), 46

Table 5. Haplotypes of CF chromosomes in individuals with CF-PS or meconium ileus (MI).

Family no.*	Haplotypes† Region 1	2	3	4	5	6	7	8	9	CF alleles‡
						CF-PS				
3	A	A	A	A	A	A	–	–	A	ΔF_{508} (group Ia)
	D	C	B	A	A	C	B	A	A	M (predicted, group IIIb)
14	B	A	A	A	–	A	A	A	A	ΔF_{508} (group Ia)
	B	C	B	B	–	B	A	C	C	M (predicted, group IIa)
27	A	B	–	A	A	A	A	A	E	ΔF_{508} (group Ia)
	A	C	–	A	A	A	A	A	A	M (predicted, group Ib)
29	A	C	–	C	–	A	A	A	B	ΔF_{508} (group Ia)
	B	A	–	B	–	B	A	A/C	B	M (predicted, group IIa)
40	D	A	A	A	A	A	A	A	B	ΔF_{508} (group Ia)
	F	C	B	A	A	C	B	C	A	M (predicted, group IV)
51	C	C	A	B	B	B/C	A	C	A	M (predicted, group IIa)
	F	D	A	B	B	B/C	A	C	C	M (predicted, group IIa)
54	B	C	A	B	C	C	A	C	A	M or S (predicted, group V)
	B	B	A	A	A	C	B	A	A	M (predicted, group IIIb)
						MI				
4	B	A	A	A	A	A	A	A	A	ΔF_{508} (group Ia)
	B	A	A	A	A	A	A	A	A	ΔF_{508} (group Ia)
10	D	B	A	A	–	A	–	A	A	ΔF_{508} (group Ia)
	A	D	A	A	–	A	A	A	B	ΔF_{508} (group Ia)
23	A	E	B	A	A	A	A	A	E	ΔF_{508} (group Ia)
	B	C	A	A	A	A	A	A	B	S (predicted, group Ib)
28	A	A	–	A	A	A	A	A	C	ΔF_{508} (group Ia)
	A	A	–	A	A	A	A	A	B	ΔF_{508} (group Ia)
33	B	B	–	A	A	A	–	A	B	ΔF_{508} (group Ia)
	B	A	–	A	A	A	A	A	B	ΔF_{508} (group Ia)
49	A	A	A	A	A	A	A	A	B	ΔF_{508} (group Ia)
	A	A	A	A	A	A	A	A	B	ΔF_{508} (group Ia)

*Family numbers from (*36*). †The haplotype definitions are the same as in Table 3. ‡Allele designations are the same as Table 4.

percent of CF patients without a previous family history can be accurately diagnosed by DNA analysis, and 68 percent of the CF carriers in the population can be identified with a probe specific for the ΔF_{508} mutation. It will be necessary to define most of the existing CF mutations before an effective DNA-based genetic testing program can be implemented at the population level.

Since the ΔF_{508} mutation constitutes about 68 percent of all CF mutations, RFLP analysis may be used to supplement genetic diagnosis by direct deletion detection. About 55 percent of the CF parents not carrying the ΔF_{508} mutation are expected to be informative for the DNA marker JG2E1 based on retrospective analysis of our CF linkage families (*36*); an additional 39 percent would be informative if E6 (Taq I) and J3.11(Msp I) were also tested; virtually all parents would be informative if H2.3A (XV2C-Taq I), E2.6(E.9-Msp I), E4.1 (Mp6d.9-Msp I), J44 (Xba I), and metD (Ban I) were included. Clearly, more accurate and simple testing procedures will be developed when additional CF mutations are identified.

As indicated above, variation of DNA marker haplotype association in CF has been detected among different European populations (*9, 15*). Whether the haplotype classifications proposed here correlate with geographic distributions remains to be determined. Our preliminary study showed that all seven CF chromosomes in group Ib (Table 3) were from individuals of French-Canadian ancestry (*44*). It is unexpected that at least two independent mutational events had occurred on a haplotype (groups Ia and Ib) that is rare among the N chromosomes. A detailed description of such haplotypes would be useful for understanding the high incidence of CF in the Caucasian population (*44*) and for further mutational analysis. Further, a complete molecular description of all the CF mutations should provide important insight into the pathophysiology of the disease and form the basis for the development of improved treatment.

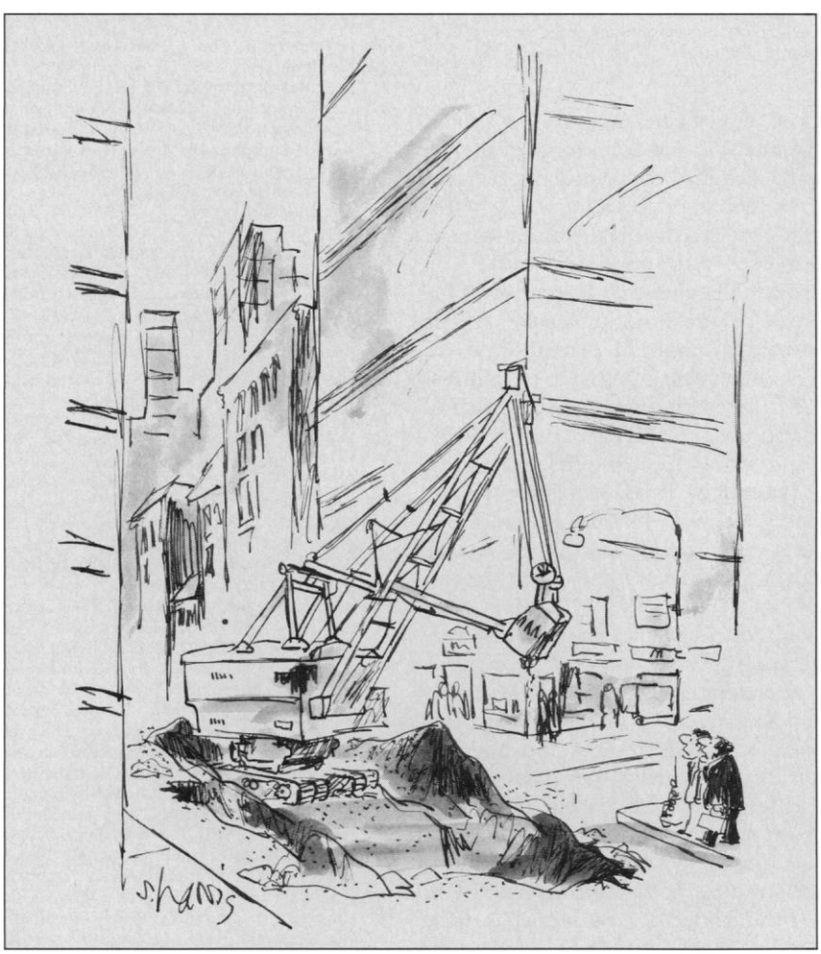
REFERENCES AND NOTES

1. T. F. Boat, M. J. Welsh, A. L. Beaudet, in *The Metabolic Basis of Inherited Disease*, C. L. Scriver, A. L. Beaudet, W. S. Sly, D. Valle, Eds. (McGraw-Hill, New York, ed. 6, 1989), pp. 2649–2680.
2. H. O. Goodman and S. C. Reed, *Am. J. Hum. Genet.* **4**, 59 (1952).
3. D. M. Danks, J. L. Allan, C. M. Anderson, *Ann. Hum. Genet.* **28**, 323 (1965); A. G. Knudson, L. Wayne, W. Y. Mallett, *Am. J. Hum. Genet.* **19**, 388 (1967); P. M. Conneally, A. D. Herritt, P. L. Yu, *Tex. Rep. Biol. Med.* **31**, 639 (1973); D. J. Pritchard, G. R. Hickman, R. Nelson, *Arch. Dis. Child.* **58**, 290 (1983); D. J. Pritchard, *Nature* **330**, 319 (1987); R. S. Meindl, *Am. J. Phys. Anthropol.* **74**, 39 (1987); L. B. Jorde and G. M. Lathrop, *Am. J. Hum. Genet.* **42**, 808 (1988).
4. S. W. Wright and N. E. Morton, *Am. J. Hum. Genet.* **20**, 157 (1968).
5. T. Schaap and M. M. Cohen, in *Cystic Fibrosis: Projection into Future*, J. A. Mangos and R. C. Talamo, Eds. (Stratton International, New York, 1976), pp. 291–303.
6. A. B. Burdick, *Hum. Hered.* **27**, 366 (1977); J. H. Edwards, *Arch. Dis. Child.* **52**, 343 (1977).
7. H. Eiberg *et al.*, *Cytogenet. Cell Genet.* **40**, 623 (1985); K. Schmiegelow *et al.*, *Clin. Genet.* **29**, 374 (1986).
8. L.-C. Tsui *et al.*, *Science* **230**, 1054 (1985); R. G. Knowlton *et al.*, *Nature* **318**, 380 (1985); K. Klinger *et al.*, *Nucleic Acids Res.* **14**, 8681 (1986); J. Schmidtke *et al.*, *Hum. Genet.* **76**, 337 (1987).
9. R. White *et al.*, *Nature* **318**, 382 (1985).
10. B. J. Wainwright *et al.*, *ibid.*, p. 384.
11. J. E. Spence *et al.*, *Am. J. Hum. Genet.* **39**, 729 (1986).
12. X. Estivill *et al.*, *Nature* **326**, 840 (1987); X. Estivill *et al.*, *Genomics* **1**, 257 (1987).
13. J. M. Rommens *et al.*, *Am. J. Hum. Genet.* **43**, 645 (1988).
14. A. L. Beaudet *et al.*, *N. Engl. J. Med.* **318**, 50 (1988); A. L. Beaudet, G. L. Feldman, S. D. Fernbach, G. J. Buffone, W. E. O'Brien, *Am. J. Hum. Genet.* **44**, 319 (1989).
15. X. Estivill *et al.*, *Am. J. Hum. Genet.* **43**, 23 (1988).
16. C. Ober *et al.*, *ibid.* **41**, 1145 (1987); G. Cutting *et al.*, *ibid.* **44**, 307 (1989); T. M. Fujiwara *et al.*, *ibid.*, p. 327, L.-C. Tsui, *ibid.*, p. 303.
17. M. A. Spence and L.-C. Tsui, *Cytogenet. Cell Genet.* **46**, 170 (1987); M. Dean, *Genomics* **3**, 93 (1988); L.-C. Tsui, M. Farrall, H. Donis-Keller, *Cytogenet. Cell Genet.* **51**, 166 (1989).
18. M. Corey, P. R. Durie, D. Moore, G. Forstner, H. Levison *J. Pediatr.* **115**, 274 (1989).
19. B. Kerem *et al.*, *Am. J. Hum. Genet.* **44**, 827 (1989).
20. M. Devoto *et al.*, *ibid.* **43**, A82 (1988).
21. A. L. Beaudet *et al.*, *ibid.* **38**, 681 (1986); G. M. Lathrop *et al.*, *ibid.* **42**, 38 (1988).
22. A.-M. Poustka, H. Lehrach, R. Williamson, G. Bates, *Genomics* **2**, 337 (1988); M. L. Drumm *et al.*, *ibid.*, p. 346; T. R. Fulton *et al.*, *Nucleic Acids Res.* **17**, 271 (1989); J. M. Rommens *et al.*, *Am. J. Hum. Genet.*, in press.
23. J. M. Rommens *et al.*, *Science* **245**, 1059 (1989).
24. J. R. Riordan *et al.*, *ibid.*, p. 1066.
25. These families include GM1076 [L.-C. Tsui, K. Buetow, M. Buchwald, *Am. J. Hum. Genet.* **39**, 720 (1986)]; family SLC1380, reported in R. White *et al.*, *ibid.*, p. 694; family 1 from the German Democratic Republic, reported in Berger *et al.*, *Hum. Genet.* 77, 197 (1987) and in (*26*); family 2 from France, family 3 from Italy, family 4 from Yugoslavia and family 5 from the United States, reported in (*26*).
26. M. Farrall *et al.*, *Am. J. Hum. Genet.* **43**, 471 (1988).
27. Family SLC1380 (*22*) was a generous gift from P. O'Connell; family 3 (*22*) from G. Romeo; family 4 (*22*) from D. Halley; family 2 (*22*) from A. Kitzis; family 1 (*22*) from C. Coutelle and R. Williamson.
28. The method for family DNA analysis has been described previously (*13, 19*).
29. G. Romeo, M. Devoto, L.-C. Tsui, unpublished data.
30. X. Estivill *et al.*, *Am. J. Hum. Genet.* **44**, 704 (1989).
31. M. C. Iannuzzi *et al.*, *ibid.*, p. 695.
32. W. F. Bodmer and J. G. Bodmer, *Br. Med. Bull.* **34**, 309 (1978); S. E. Antonarakis, C. D. Boehm, P. J. Giardina, H. H. Kazazian, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 137 (1982); J. C. Murrary, K. A. Mills, C. M. Demopulos, S. Hornung, A. G. Motulsky, *ibid.* **81**, 3486 (1984).
33. A. Chakravarti *et al.*, *Am. J. Hum. Genet.* **36**, 1239 (1984).
34. B. Kerem *et al.*, unpublished data.
35. A specific region of genomic DNA, as defined by the oligonucleotide primers C16B (5′-GTTTTCCTGGATTATGCCTGGGCAC-3′) and C16D (5′-GTTGGCATGCTTTGATGACGCTTC-3′), was amplified by PCR [R. K. Saiki *et al.*, *Science* **230**, 1350 (1985); R. K. Saiki *et al.*, *ibid.* **239**, 487 (1988)]. Briefly, 200 to 400 ng of genomic DNA from either cultured lymphoblasts or peripheral blood samples of CF individuals and their parents were used in each PCR with the oligonucleotides primers indicated above [see (*24*) for their corresponding regions]. The oligonucleotides were purchased from the HSC DNA Biotechnology Service Center and purified with Oligonucleotide Purification Cartridges (Applied Biosystems) or NENSORB PREP columns (Dupont), with procedures recommended by the suppliers. The primers were annealed at 62° for 45 seconds, extended at 72° for 120 seconds (with 2 units of Taq DNA polymerase) and denatured at 94°C for 60 seconds, for 28 cycles with a final cycle of 7 minutes for extension in a Perkin-Elmer/Cetus automatic thermocycler. Portions of the PCR products were separated by electrophoresis on 1.4 percent agarose gels, and transferred to Zetabind (Bio-Rad) membrane according to standard procedures. Oligonucleotide probes (10 ng each) were labeled separately with 10 units of T4 polynucleotide kinase (Pharmacia) in a 10-μl reaction containing 50 m*M* tris-HCl (*p*H 7.6), 10 m*M* $MgCl_2$, 500 μ*M* dithiothreitol, 10 m*M* spermidine, 1 m*M*

EDTA and 30 to 40 μCi of [γ-^{32}P]ATP for 20 to 30 minutes at 37°C. The unincorporated radionucleotides were removed with a Sephadex G-25 column before use. The hybridization conditions have been described (*13*), except that the temperature was 37°C. The membranes were washed twice at room temperature with 5× SSC and twice at 39°C with 2× SSC (1× SCC is 150 m*M* NaCl and 15 m*M* sodium citrate). Autoradiography was performed at room temperature overnight.

36. The families used for linkage analysis were described by L.-C. Tsui, S. Zengerling, H. F. Willard, M. Buchwald, *Cold Spring Harbor Symp. Quant. Biol.* **51**, 325 (1986). Haplotypes for N and CF parental chromosomes were constructed by inspection of the patient chromosomes in these families.
37. T. Cox *et al.*, manuscript in preparation.
38. Y. W. Kan and A. M. Dozy, *Science* **209**, 388 (1980); S. E. Antonarakis *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 853 (1984); S. E. Antonarakis *et al.*, *ibid.* **79**, 6608 (1982); H. H. Kazazian *et al.*, *Am. J. Hum. Genet.* **36**, 212 (1984); for review, see S. H. Orkin, H. H. Kazazian, *Annu. Rev. Genet.* **18**, 131 (1984).
39. Y. W. Kan and A. M. Dozy, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 5631 (1978); Y. W. Kan, K. Y. Lee, M. Furbetta, A. Angius, A. Cao, *N. Engl. J. Med.* **302**, 185 (1980); H. H. Kazazian *et al.*, *Nature* **310**, 152 (1984); R. Chakraborty *et al.*, *Hum. Genet.* **76**, 40 (1987); S. L. C. Woo, *Am. J. Hum. Genet.* **43**, 781 (1988); S. P. Daiger *et al.*, *ibid.* **45**, 310 (1989); S. P. Daiger *et al.*, *ibid.*, p. 319; D. W. Cox, S. L. C. Woo, T. Mansfield, *Nature* **316**, 79 (1985).
40. K. J. Gaskin *et al.*, *J. Pediatr.* **100**, 857 (1982); K. J. Gaskin, P. R. Durie, L. Lee, R. Hill, G. G. Forstner, *Gastroenterology* **86**, 1 (1984).
41. The haplotypes detected in the MI patients are shown in Table 5. The criteria for diagnosis have been described by E. Kerem *et al.*, *J. Pediatr.* **114**, 767 (1989).
42. J. L. Allan, M. Robbie, P. D. Phelan, D. M. Danks, *Eur. J. Pediatr.* **135**, 291 (1981).
43. M. Farrall *et al.*, *Lancet* **i**, 1402 (1986); J. E. Spence *et al.*, *Hum. Genet.* **76**, 5 (1987); M. Super *et al.*, *Lancet* **ii**, 782 (1987); L. Mathy *et al.*, *Hum. Genet.* **75**, 359 (1987); C. E. Nugent *et al.*, *Obstet. Gynecol.* **71**, 213 (1988).
44. Most of the French-Canadian CF families were from the Chicoutimi Clinic, Quebec; likewise, terms such as Northern European refer only to geographic distribution.
45. We thank F. S. Collins for discussions; P. Ray for providing CF family DNA; R. Worton for continued encouragement and critical reading of this manuscript; N. Plavsic, D. Kennedy, M. Zsiga for technical assistance, and those investigators listed in (*27*) for sharing recombinant family materials. This research is supported by grants from the Canadian Cystic Fibrosis Foundation (CCFF), the National Institutes of Health (DK34944, GM33771), the Cystic Fibrosis Foundation (CFF) (USA), and the Sellers Fund. J.M.R. and J.A.B. are Postdoctoral Fellows of the Medical Research Council (MRC) of Canada and CFF (USA), respectively; A.C. is a recipient of a Research Career Development Award from the NIH (HD00774); L.-C.T. is a Research Scholar of CCFF and recipient of an MRC Scientist Award.

7 August 1989; 18 August 1989



"It turned out there just wasn't any ordinance against *strip mining on this street."*