Am J Hum Genet 39:699–706, 1986

Genetic Analysis of Cystic Fibrosis: Linkage of DNA and Classical Markers in Multiplex Families

A. M. BOWCOCK,[1] J. CRANDALL,[2] L. DANESHVAR,[1] G. M. LEE,[2] B. YOUNG,[1] V. ZUNZUNEGUI,[2] C. CRAIK,[3] L. L. CAVALLI-SFORZA,[1] AND M.-C. KING[2]

SUMMARY

Linkage of cystic fibrosis (CF) to DNA and classical markers was studied in 36 families of two or three generations with at least two living affected children. Among the 79 affected children, no recombinants were detected between the disease and the markers MET and pJ3.11, previously shown to be linked to CF. No linkage between the human trypsin gene family (which appears to include at least 10 members) and CF was found, although not all genes of the trypsin family have been screened yet. In one of the CF families, recombination between MET and pJ3.11 was detected in an unaffected sib. Data from our families suggest that the gene order of markers among chromosome 7q is: (7cen;p8.33)collagen(COL1A2);DOCR1-917;paraoxonase(PON);(MET-cf-J3.11);T-cell receptor beta chain (TCRB);qter. There was no evidence for (or against) either postzygotic selection or meiotic drive to explain the high frequency of CF in Caucasian populations.

INTRODUCTION

The rationale for linkage analysis of CF is explained in the joint analysis [1]. In this paper, we present results from our linkage analysis of DNA and classical

Received July 24, 1986.

This research was supported by grants AM 34942, GM 28428, and CA 27632 from the National Institutes of Health.

[1] Department of Genetics, Stanford University, Stanford, CA 94305.

[2] School of Public Health, University of California, Berkeley, CA 94720.

[3] Departments of Biochemistry, Pharmaceutical Chemistry, and Biophysics, University of California, San Francisco, CA 94143.

polymorphisms in multiplex CF families [2, 3], including markers from chromosome 7q and some members of the trypsin gene family. We also discuss a crossover between MET and pJ3.11 in one of our families, which has made it possible to order these markers with respect to others on 7q. Finally, we study the ratio of carriers to noncarriers among the unaffected siblings of CF children, estimating the sample size necessary to establish a statistically significant deviation, under specific evolutionary hypotheses, from the expected 2:1 ratio.

MATERIALS AND METHODS

The families in our sample are described in table 1. Clinical collaborators who assisted in identifying these families and encouraging their participation were: Drs. B. Nickerson, J. McQuitty, and H. Lipow and Ms. M. Minor (Children's Hospital, Medical Center, Oakland); Drs. C. Wong and A. Osher (Children's Hospital of Los Angeles); Dr. W. F. Howatt (C. S. Mott Children's Hospital, Ann Arbor), Dr. C. J. L. Newth (University of California Hospital, San Francisco), and Dr. L. Taussig (College of Medicine, University of Arizona). We collected 20–40 ml whole blood into ACD from all parents, CF children, and unaffected siblings and (for two families) grandparents and cousins.

All family members were typed for 26 classical markers. The paraoxonase/arylesterase polymorphism was typed as a codominant using a method developed by B. LaDu and S. Adkins (University of Michigan) (personal communication). It involves determining by spectrophotometry the initial rates of hydrolysis of paraoxon by human sera in the presence of 1 M NaCl and then of 10 mm phenyl acetate. The ratio of the paraoxonase activity stimulated by salt over paraoxonase activity without salt distinguishes the AA type from the AB and BB types. The ratio of the paraoxonase activity with salt over the arylesterase activity distinguishes the AA, AB, and BB types. The two ratios were used to validate each other.

DNA probes were obtained from the following sources: human procollagen COL1A2 probe HF32 from J. Myers, p*met*H and p*met*D from R. White, pJ3.11 from J. Schmidtke, the T-cell receptor beta-chain probe PT10 from J. G. Seidman, and 7C22 from R. Williamson. *Hin*dIII and *Hin*cII blots were probed with DORC1-917 by L.-C. Tsui.

TABLE 1

FAMILIES WITH TWO OR MORE AFFECTED CHILDREN USED IN GENETIC ANALYSIS OF CF

CHILDREN LIVING AND SAMPLED		
Affected	Unaffected	No. FAMILIES
2	0	13
2	1	7
2	2	6
2	3	3
2	7	1
3	0	2
3	1	2
3	8*	1
4	5*	1
Total 79	50	36

* 3-generation families, each with two sets of sibs.

DNA obtained from the ACD blood samples was digested, electrophoresed, transferred, and hybridized as described [4]. Hybridization conditions for the rat trypsin cDNA probes were performed under less stringent conditions, using 40% formamide at 42°C in 4× SSPE. The final wash was in 0.1× SSPE, 0.1% SDS at 42°C for 1 hr.

RESULTS

Linkage Data

Table 2 indicates the linkage relationships between CF and all DNA and classical markers in our "CAL" families. The chromosome 7 markers, of course, are of greatest interest. At a recombination fraction of .001, the lod score for MET is 9.294, and for pJ3.11, 9.994. Estimated recombination fractions for the other chromosome 7q markers and CF are approximately .22 for COL1A2, .14 for PON, .15 for DOCR1-917, and .35 for TCRB. These are consistent with the joint data, although none of the lod scores for these markers were above three in our sample alone.

Trypsin Analysis

Using the 222- and 428-base pair (bp) *Pst*I fragments of a rat trypsin gene [5], we detected hybridization to 7–12 fragments of human DNA. These may represent a number of human trypsin genes analogous to those of the rat that may have at least 10 closely linked trypsin gene family members [5]. We were able to detect RFLPs with *Msp*I and *Taq*I using the 222-bp fragment, the 428-bp fragment detected on RFLP with *Hin*dIII. *Pvu*II, *Rsa*I, *Eco*RI, and *Sac*I detected a deletion of certain trypsin specific bands in some individuals. These may be deletions of trypsin pseudogenes. We could exclude linkage of CF with the trypsin *Hin*dIII RFLP closer than 0.02 and linkage of CF to one *Taq*I variant closer than 0.02.

In one family, recombination occurred between the *Taq*I and *Hin*dIII sites, and in another family, between the two *Pvu* sites. Thus, although there may be a number of trypsin genes, they are far enough apart to recombine and may not necessarily be clustered. A number of trypsin markers could be excluded from linkage to the CF gene, but it is unlikely that the entire gene family is represented by the polymorphisms detected thus far.

MET-pJ3.11 Recombination

In one of our families, we detected crossing over between pJ3.11 and MET in a nonaffected CF sib (fig. 1). This enabled us to propose an order for these markers with respect to the other chromosome 7 markers. The recombinant child—195—has inherited a normal chromosome from his mother, but his paternal chromosome reflects a paternal recombination between pJ3.11 and MET. Child 195 is unaffected, but it is not possible to determine his carrier status. Carrier status for the other children in this family was determined from pJ3.11 and MET data, and the inferred phases of maternal or paternal chromosomes are given in figure 1. This and our other families suggest the gene order p8.33-COL1A2-PON-MET-J3.11-TCRB. Given data from all laboratories, this becomes 7cen-COL1A2-DOCR1.917-PON-MET-J3.11-TCRB. Since the

TABLE 2

LOD SCORES FOR LINKAGE OF CYSTIC FIBROSIS TO GENETIC MARKERS

MARKER	CHROMOSOME	INFORMATIVE FAMILIES	RECOMBINATION FRACTION				
			.001	.10	.20	.30	.40
RH	1p32-ter	33	−66.823	−8.294	−2.570	−.663	−.097
PGM1	1p22	17	−15.626	.109	.683	.441	.130
AMY	1p21	6	−3.397	.158	.300	.197	.061
FY	1q13	32	−27.653	−1.210	.547	.584	.204
BNGF	1p21-qter	8	−5.496	−.100	.199	.158	.051
D1S2	1	1	−1.798	.022	.124	.095	.031
POMC	2p23	6	−5.745	−.265	.108	.122	.043
D2S1	2pter-p22	10	−6.848	.028	.213	.114	.029
ACP	2p23-25	27	−48.312	−6.890	−2.463	−.787	−.155
D1S1 *Hin*dIII	3	7	−8.693	−1.130	−.375	−.110	−.020
GC	4q11-13	33	−58.466	−7.576	−2.600	−.813	−.159
MNS	4q28-31	28	−30.140	−1.320	.540	.552	.182
D5S2	5	3	−4.497	−.673	−.254	−.087	−.018
HLA	6p21-23	7	−22.834	−3.598	−1.410	−.501	−.109
BF	6p21-23	10	−13.745	−2.247	−.938	−.358	−.083
GLO1	6p21-223	12	−17.158	−1.854	−.426	−.036	.015
D6S2	6p21-qter	10	−24.236	−3.377	−1.211	−.385	−.075
D7S1	7	10	−22.92	−3.67	−1.45	−.52	−.15
D7S11	7p?	12	−13.560	−.503	.363	.343	.116
P8.33	7cen	13	−15.035	−1.804	−.505	−.095	−.004
COL1A2	7q22	18	−11.353	.143	.529	.333	.007
DOCR1-917	7q	10	−7.190	1.185	1.117	.608	.166
PON	7q	22	−3.466	2.223	1.482	.721	.212
METHD	7q21-31	14	9.294	6.708	4.185	2.000	.517
J3.11	7cen-q22	17	9.994	7.073	4.372	2.143	.411
TCRB	7q3	13	−15.659	−1.147	−.154	.023	.017
13.15	9p	6	−6.397	−.763	−.228	−.056	−.008
AK1	9q34	4	−4.299	−.534	−.168	−.045	−.007
ABO	9q34	26	−23.515	−.925	.566	.547	.184
PHAGE42	9	4	−9.271	−1.540	−.628	−.231	−.052

D11S16	11	11	−10.416	.073	.442	.272	.077
A2M	12	3	−4.196	−.458	−.120	−.023	−.002
D13 S1	13q12-q14	9	−6.744	.344	.591	.387	.119
ESD	13q14	13	−16.874	−3.146	−1.336	−.501	−.113
D13 S5	13q12-q22	9	−7.023	−.304	.224	.140	.039
D13 S6	13q12-q22	5	−5.395	−.088	−.146	.104	.032
D13 S7	13q12-q22	2	−4.797	−.887	−.388	−.151	−.035
D13 S2	13q22	3	−1.973	−.145	−.011	.011	.005
D13 S3	13q22-qter	13	−17.663	−2.458	−.866	−.267	−.049
D13 S4	13q22-qter	5	−11.091	−1.574	−.568	−.185	−.038
PMP1	13qter	2	−6.296	−.671	−.156	−.011	−.007
D14 S4	14	4	−10.791	−1.323	−.385	−.079	−.005
PI	14q24-32	13	−13.515	−.738	.130	.194	.068
D15 S4	15q12-q24	5	−3.647	−.024	.180	.140	.046
GPT1	16pter-11	26	−35.189	−4.953	−1.862	−.649	−.140
HP	16q21-22	27	−36.267	−4.431	−1.195	−.198	.003
LE	19	11	−3.381	.153	.287	.193	.061
C3	19	16	−18.706	−1.928	−.435	−.030	.018
ADA	20p11-qter	3	−4.548	−.705	−.272	−.095	−.020
K	?	8	−.947	.717	.592	.331	.096
TC2	?	24	−28.687	−3.527	−1.060	−.257	−.033
P	?	26	−25.835	−3.258	−1.070	−.309	−.054
TF	?	6	−6.124	−.556	−.093	.012	.011
F13B	?	18	−20.371	−1.908	−.395	−.019	.018
CO	?	4	−1.747	.027	.105	.072	.002
PL39.1	?	8	−18.689	−3.191	−1.316	−.488	−.110
PL3.1	?	11	−17.437	−2.281	−.756	−.220	−.040
PL18.1	?	19	−32.900	−4.436	−1.512	−.453	−.083
PHAGE28	?	5	−4.248	−.502	−.150	−.037	−.006
PHAGE4	?	6	−3.726	−.103	.108	.090	.029
PL19.2	?	12	−6.585	−.323	.559	.352	.106

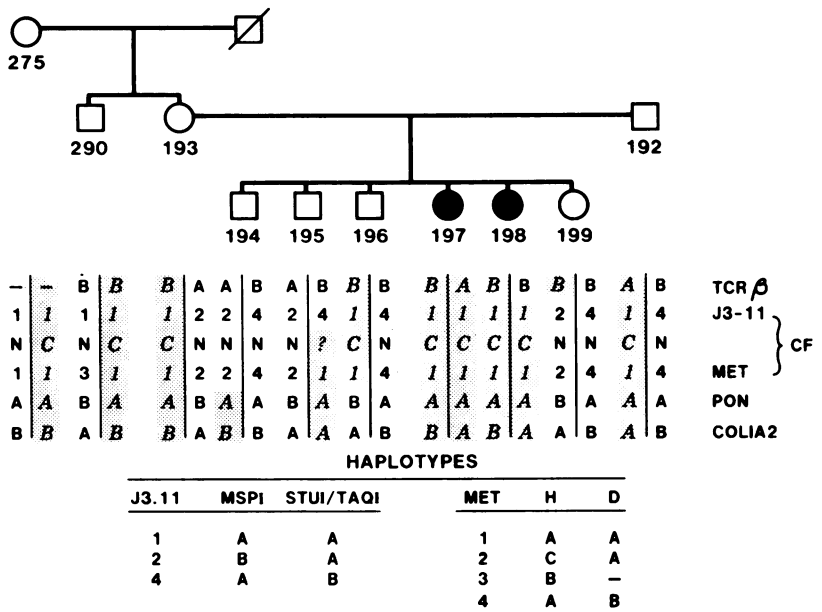

FIG. 1.—Portion of family CAL41 with recombinants illustrated. The data suggest that pJ3.11 is between TCRB and MET and supports PON, DOCR1-917, and COL1A2 on the opposite side of the pJ3.11, CF, MET cluster from TCRB. The most likely order of genes is indicated, but bracketed pairs of loci could have either position; that is, pJ3.11 and CF, or CF and MET, could be reversed. The carrier status of child 195 cannot be determined. This family was also typed for 7C22, but no additional linkage information was obtained.

carrier status of 195 cannot be determined, it is not possible to position CF with respect to MET and pJ3.11 using this family.

Carrier Status of Unaffected Siblings

For the 228 unaffected siblings of CF children, we assessed carrier status based on MET and pJ3.11. Frequencies of carrier, noncarrier, and partially informative siblings are shown in table 9 of the joint report [1]. The partially informative siblings have one chromosome known to be normal. Therefore, under standard Mendelian proportions, we expect equal frequencies of carriers and noncarriers among these siblings.

As demonstrated in the joint report, there is no difference between the observed and expected ratios of carriers to noncarriers. We estimated the sample size necessary to detect a significant excess of carrier siblings as follows. Assume the frequency of the CF allele among Caucasians is in equilibrium at the observed frequency (1/40) and that fitnesses are 0 for CF children, 1 for carriers, and 1 − s for noncarriers. Then $1/40 = s/(1 + s)$, so $s = 1/39$. If there is no selection against noncarrier sibs of CF children, then 1/3 of CF sibs will be noncarriers and 2/3 will be carriers. For selection s against noncarriers, the expected proportion of siblings will be $(1 - s)/(3 - s)$ normal and $2/(3 - s)$ carriers, or .328 noncarriers and .672 carriers for $s = 1/39$. The chi-square

for detecting a difference between these two noncarrier:carrier ratios with N unaffected siblings:

$$\chi^2 = \frac{N\left[\frac{1-s}{3-s} - \frac{1}{3}\right]^2}{\frac{1}{3}} + \frac{N\left[\frac{2}{3-s} - \frac{2}{3}\right]^2}{\frac{2}{3}},$$

so that

$$N = \frac{\chi^2(3-s)^2}{2s^2}.$$

For $P = .05$, chi-square = 3.84, s = 1/39, N = 25,836. This number is about 100-fold higher than that available for testing. Similar calculations hold if meiotic drive is responsible for the deviation. As a further complication, heterozygous advantage might be due to other selective mechanisms that cannot be analyzed by the present data, such as differential fertility.

DISCUSSION

Our data confirm the very close linkage of MET, pJ3.11, and cystic fibrosis; in fact, we did not find a single recombinant between CF and either marker in our 36 families. One recombinant between MET and pJ3.11 was observed and, with our other families, suggests the order 7cen-COL1A2-PON-MET-J3.11-TCRB-7qter. Linkage between CF and trypsin was not found; however, trypsin is a very large gene family and not all members can be excluded.

The close linkage of CF, MET, and pJ3.11 excludes important genetic heterogeneity. The existence of fairly close linkage is, in itself, proof that there cannot be much heterogeneity due to unlinked disease loci, as shown by the following simple model: Assume that there are two disease loci, one unlinked to the marker being examined and the other linked with unknown recombination theta. If the two loci are found in proportions (1 − alpha) and alpha of the families, respectively, the frequency of families with the unlinked marker (a measure of heterogeneity = 1 − alpha) cannot be greater than twice the observed recombination. This is easily deduced from formula 513, p. 110 of Ott [6]; see also Cavalli-Sforza and King [7]. Thus, a first firm result is that there is only one locus contributing to at least 98% of all CF cases (based on twice the 95% confidence limit for MET or pJ3.11-CF recombination given in the joint paper), although so far one cannot entirely exclude other rare loci. This conclusion allows the detection of carriers in families with at least one affected child and an excellent probability of fetal diagnosis for pregnancies after the first affected child. It also paves the way for location of the CF gene.

REFERENCES

1. Beaudet A, Bowcock A, Buchwald M, et al.: Linkage of cystic fibrosis to two tightly linked DNA markers: joint report from a collaborative study. *Am J Hum Genet* 39:691–693, 1986

2. Bowcock AM, Lee GM, Sanders MF, Cavalli-Sforza LL: Linkage analysis of cystic fibrosis with DNA polymorphisms in multiply affected families. *Cytogenet Cell Genet* 40:A587a, 1985
3. Lee GM, Bowcock AM, Lattanzio D, et al.: Linkage analysis of cystic fibrosis in multiply affected families. *Am J Hum Genet* 37:A593, 1985
4. Feder J, Yen L, Wijsman E, et al.: A systematic approach for detecting high-frequency restriction fragment length polymorphisms using large genomic probes. *Am J Hum Genet* 37:635–649, 1985
5. Craik CS, Choo Q-L, Swift GH, Quinto C, MacDonald RJ, Rutter WJ: Structure of two related rat pancreatic trypsin genes. *J Biol Chem* 259:14255–14264, 1984
6. Ott J: *Analysis of Human Genetic Linkage*. Baltimore, Johns Hopkins Univ. Press, 1985
7. Cavalli-Sforza LL, King M-C: Detecting linkage for genetically heterogeneous diseases and detecting heterogeneity with linkage data. *Am J Hum Genet* 38:599–616, 1986