GENOMICS **40,** 73–79 (1997)
ARTICLE NO. GE964545

The Genomic Structure of the Gene Defective in Usher Syndrome Type Ib (MYO7A)

PHILIP M. KELLEY,* MIKE D. WESTON, ZHENG-YI CHEN,† DANA J. ORTEN,* TAMA HASSON,‡ LARRY D. OVERBECK,* JEFF PINNT,* CATHERINE B. TALMADGE,§ PAUL ING,* MARK S. MOOSEKER,‡ DAVID COREY,† JANOS SUMEGI,§ AND WILLIAM J. KIMBERLING*,[1]

*Center for Hereditary Communication Disorders, Boys Town National Research Hospital, Omaha, Nebraska, 68131; †Department of Neurobiology, Massachusetts General Hospital and Harvard Medical School and Howard Hughes Medical Institute, Boston, Massachusetts 02114; ‡Departments of Biology and Pathology, Yale University, New Haven, Connecticut 06520; and §Department of Pathology and Microbiology, University of Nebraska Medical Center, Omaha Nebraska 68198

Received September 4, 1996; accepted November 27, 1996

Usher syndrome type Ib is a recessive autosomal disorder manifested by congenital deafness, vestibular dysfunction, and progressive retinal degeneration. Mutations in the human myosin VIIa gene (MYO7A) have been reported to cause Usher type Ib. Here we report the genomic organization of MYO7A. An STS content map was determined to discover the YAC clones that would cover the critical region for Usher syndrome type Ib. Three of the YACs (802A5, 966D6, and 965F10) were subcloned into cosmids and used to assemble a preliminary cosmid contig of the critical region. Part of the gene encoding human myosin VIIa was found in the preliminary cosmid contig. A cosmid, P1, PAC, and long PCR contig that contained the entire MYO7A gene was assembled. Primers were designed from the composite cDNA sequence and used to detect intron–exon junctions by directly sequencing cosmid, P1, PAC, and genomic PCR DNA. Alternatively spliced products were transcribed from the MYO7A gene: the largest transcript (7.4 kb) contains 49 exons. The MYO7A gene is relatively large, spanning approximately 120 kb of genomic DNA on chromosome 11q13.

INTRODUCTION

Usher syndrome type Ib has been mapped to chromosome 11q13.5 and is characterized by profound congenital deafness, vestibular areflexia, and progressive retinitis pigmentosa (Kimberling *et al.*, 1992; Smith *et al.*, 1992). This autosomal recessive disorder is the most common cause of combined deafness and blindness in humans, and type Ib is the most common type of Usher syndrome type I. Two other genes have been associated with similar symptoms; Usher type Ia mapping to 14q31 (Kaplan *et al.*, 1992) and Usher type Ic mapping to 11p13–p15 (Smith *et al.*, 1992). Two other Usher syndromes are distinguished by the age of onset, severity of hearing loss, and vestibular effects. Usher syndrome type II patients are "hard of hearing" with a normal vestibular response. Usher syndrome type III patients experience a progressive hearing loss, while vestibular effects are variable. Usher syndrome type IIa maps to chromosome 1q41 (Kimberling *et al.*, 1995; Sumegi *et al.*, 1996). Usher syndrome type III maps to chromosome 3q (Sankila, 1994). Mutations in the MYO7A gene, which encodes the human myosin VIIa protein, have been shown to segregate with Usher syndrome type Ib in humans (Weil *et al.*, 1995; Weston *et al.*, 1996). A nonsyndromic recessive deafness has been mapped to the same region as Usher Ib but as yet no mutations associating it with MYO7A have been reported (Guilford *et al.*, 1994).

A transcript corresponding to the MYO7A gene was first discovered by reverse transcriptase–polymerase chain reaction (RT-PCR) analysis of RNA from human and porcine cell lines and tissues (Bement *et al.*, 1994). The positional cloning of mouse *shaker-1* identified the mouse gene encoding myosin VIIa as causative (Gibson *et al.*, 1995). Mouse probes allowed the identification of the human homologue of the mouse *shaker-1*/myosin VIIa, which proved to contain mutations segregating with Usher syndrome type Ib (Weil *et al.*, 1995). Antibodies have been generated against part of the tail region of myosin VIIa (Hasson *et al.*, 1995). Immunohistochemical analysis using these antibodies has shown expression in the testis, lung, and kidney as well as in tissue relevant to Usher syndrome type Ib: the inner

[1] To whom correspondence should be addressed at Center for Hereditary Communication Disorders, Boys Town National Research Hospital, 555 North 30th Street, Omaha, NE 68131. Telephone: (402) 498-6713. Fax: (402) 498-6331. E-mail: KIMBER@BOYSTOWN.ORG.

0888-7543/97 $25.00

and outer hair cells of the cochlea and the retinal pigmented epithelial cells (Hasson *et al.,* 1995). Differences in tissue-specific expression may explain why mutations associated with MYO7A cause retinitis pigmentosa in humans but mutations in the homologous murine gene do not cause a similar response in *shaker-1* mice (El-Amraoui *et al.,* 1996). A full-length cDNA contig has been assembled by hybridization screening with PCR probes, by direct screening with genomic DNA, and by RT-PCR (Hasson *et al.,* 1995; Chen *et al.,* 1996). A long and a short transcript have been described, both encoding a common head or motor region and control region (IQ domains) but having distinctly different tail regions (Chen *et al.,* 1996). These two transcripts appear to be the result of alternative splicing. Weil and co-workers (1996) have also published a description of the large transcript and multiple splice variants.

We report here an STS content map of YACs covering the critical region of Usher Ib on chromosome 11q13.5. We describe a cosmid, P1, PAC, and genomic PCR DNA contig of the region and the intron–exon structure of the MYO7A gene. We describe the size of introns and the sequences of the intron–exon junctions.More than 189 Usher families have been examined for mutation in the first 14 exons of MYO7A (Weil *et al.,* 1995; Weston *et al.,* 1996). Twenty-three mutations, 13 unique, have been found in 20 families that cosegregate with MYO7A. Mutations causing the syndrome in 169 Usher Ib families have not yet been discovered. The data reported in this paper will enable a rapid analysis of these Usher patients to discover additional causative mutations.

MATERIALS AND METHODS

STS markers and YAC screening. Sequences for STS markers from the region 11q13–q14 used in the YAC content map can be obtained through the Genomic Data Base web site (http://gdbwww.gdb.org/gdb/shortcuts.html) except for D11S1789, which is equivalent to AFM282xcl (Guilford *et al.,* 1994). YACs were obtained by screening the CEPH Mega YAC library for STS content by PCR (Albertsen *et al.,* 1990; Chumakov *et al.,* 1992; Green and Olson, 1990). The CEPH A and B Mega YAC libraries (Albertsen *et al.,* 1990) were initially obtained from Dr. H. Drabkin (University of Colorado Medical Center). An overlapping array of the same libraries was later obtained from Research Genetics (Huntsville, AL). YACs were assayed for chimerism by FISH (Talmadge *et al.,* 1995) and Southern analysis of yeast chromosomal DNA separated by pulsefield gel electrophoresis (Bio-Rad). Only nonchimeric YACs were considered for further analysis.

Subcloning of YAC DNA into cosmids. YAC DNA was purified from yeast grown in selective media as described previously (Smith *et al.,* 1988). A 100-ml culture generally yielded about 300 μg of DNA with an average size of 200 kb. YAC DNA was subcloned into SuperCos1 using a protocol provided by Stratagene. Vector and YAC DNA were ligated overnight at room temperature, then packaged using Gigapack II Gold (Stratagene). The cosmids were titered and then plated onto nylon membrane. After overnight growth, two replicas were made and grown to 1-mm colony size. Filters were fixed and hybridized to random-primed total human DNA. Positive colonies were picked into microtiter plates containing 100 μl of YT medium and grown for 6–8 h at 37°C, then adjusted to 15% with glycerol and frozen at −80°C.

P1 and PAC clones and primers used to construct physical contig. P1 and PAC clones were obtained by PCR screening done by Genome Systems, Inc. (St. Louis, MO) or by screening pooled arrays of their PAC library with primer pairs as described below. P1 and PAC clones in the contig for MYO7A are identified by their clone type and number followed by the designation given by Genome Systems: PAC 7144, PAC-12-C10; P1 8079, DMPC-HFF# 1-795-10A; and P1 10419, DMPC-HFF#1-0262-C3. PCR of genomic DNA was accomplished using the AmpliTaq polymerase (Perkin–Elmer) with a $MgCl_2$ concentration of 1.5 m*M*. Templates were denatured at 94°C for 3 min and then subjected to 35 cycles of 94°C for 30 s, 55°C for 30 s, and 68°C for 30 s. The following primers were used to construct the physical contig containing the MYO7A gene. The primer name is followed by the sequence 5′ to 3′, and amplimer pairs are followed by amplimer size: MYO3PRF (TGAGAATGTGCCGTTAGG)/MYO3PRR (CCAGGGGAAAAGAAGATG) (748 bp); 7144T7F (GATCAGGAACCAAGGGCCC)/7144T7R (CCATCCAGCACAAGACCACC) (250 bp); M7C7200F (TACTCTGAAACCGAAGGGACTG)/M7C7306R (CATGCTTTATGCCCTGAGGT) (125 bp); M7C6702F (ACTGGGCTACAAGATGGATGAC)/M7C6924R (AGAGGCCCAGTTGGTATAGTCAGT) (245 bp).

Sequencing large templates. Primers were designed from cDNA sequence using OLIGO 4.0. Total P1 or PAC DNA was sequenced using the Promega f-mol DNA Sequencing system with [γ-^{32}P]ATP. Templates were denatured at 95°C for 3 min and then subjected to 30 cycles of 95°C for 30 s, 55°C for 30 s, and 70°C for 30 s. Sequence of the intron–exon boundaries has been submitted to GenBank and is available under Accession No. U61413–U61489.

Long PCR. Long PCR of genomic DNA was accomplished using the BRL eLONGase Amplification system with a $MgCl_2$ concentration of 1.5 m*M*. Templates were denatured at 94°C for 30 s and then subjected to 35 cycles of 94°C for 30 s, 55°C for 30 s, and 68°C for 10 min. Primers for long PCR were as follows: M7C6100F (ATTGTGCCCTCACTCACCTACC)/M7C6547R (TGAGGCTGACCCCATACTTGTT) (4.5 kb).

RESULTS

YAC Contig and STS Content Map

Linkage data had defined OMP and D11S911 as flanking markers for the Usher syndrome type Ib critical region (unpublished results). An STS content map of YACs was assembled to discover a minimum number of YAC clones that would span this critical region (Fig. 1). A single YAC, 802A5, was identified with the OMP STS marker by PCR in the Généthon MEGA YAC library. Three additional YACs, 878D8, 966D6, and 984B3, were identified by PCR screening an arrayed version of the same library with the STS markers D11S937, D11S1321, D11S906, D11S911, and OMP. Additional YACs 745E12, 796E11, 725D9, 868B8, and 946F4 were identified from the Généthon physical mapping data available through the internet and verified by PCR assay (Fig. 1). The STS content map order is derived from our best interpretation of marker order published on the internet by Généthon, CHCL, and the Whitehead Institute.

P1, PAC, Cosmid, and Long PCR Contig

Three YACs (802A5, 965F10, 966D6) containing STS markers (OMP and/or D11S911) that were known to flank the critical region were used to generate cosmid subclones. The cosmid clone 7D11 (subcloned from YAC

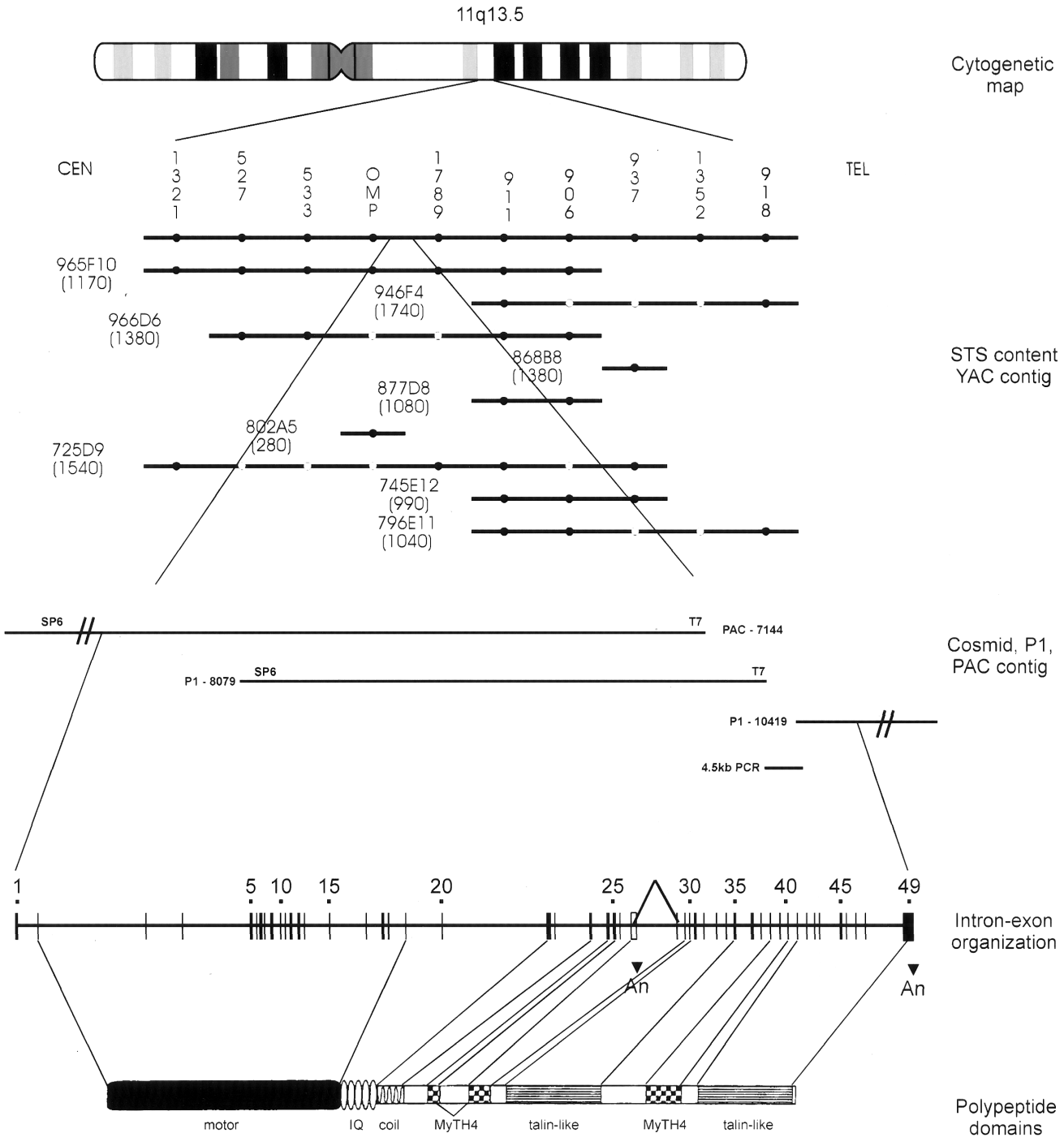

FIG. 1. Schematic representation of the MYO7A gene: its cytogenetic localization, genomic structure, exon–intron organization, and the polypeptide domains of the gene product. (Cytogenetic map) Band 11q13.5 is identified on the ISCN (1995) cytogenetic map (Mitelman, 1995). (STS content contig) The top line is the approximate order of individual STS markers based on linkage data without regard to genetic or physical distance. Individual STSs are designated by their Genome Data Base designation, that is, 1321 is D11S1321. Directly below each YAC name is the molecular weight in kilobases determined by Généthon or by PFGE. STS content of the YAC is represented with a solid circle; open circles indicate that the above marker was not found in the identified YAC while markers presumed to flank it were detected. CEN and TEL designate the centromere and the telomere of chromosome 11 relative to the contig. (Cosmid, P1, and PAC contig) A schematic of the P1, PAC, and long PCR contig from which the intron–exon organization was derived. The orientation of the clones is centromeric to telomeric which is also the direction of transcription. (Intron–exon organization) This part of the figure is a graphical summary of the data in Table 1, indicating where alternative splicing can take place in exon 27 leading to a shortened version of myosin VIIa. A_n shows where a poly(A) tail addition occurs.

802A5) contained two *Eco*RI fragments that were recognized by a 4.6-kb mouse cDNA probe (Gibson *et al.*, 1995). The two cosmid *Eco*RI subclones were sequenced, and exons 5–14 were recognized by their sequence homology to other myosins.

To extend the sequence further into the 3′ direction, a partial MYO7A transcript sequence (3492 bp) was obtained from Hasson and co-workers (1995). Unfortunately, the PCR primer pair, MYO3PR/F, designed from the sequence of the 3′ end of this transcript did not amplify any products from the cosmid libraries generated from the YACs 802A5, 965F10, and 966D6. As an alternative, this PCR primer pair was used to identify a PAC clone, 7144, successfully. Both ends of PAC clone 7144 were sequenced and PCR primer pairs designed from those sequences. The primer pair from the SP6 end of PAC 7144 amplified the appropriate sized DNA fragment from the cosmid 7A5 but not cosmid 4C2, 7D11, or 7A2. This placed the SP6 end of clone 7144 at the T7 end of 7A5, indicating that 4C2, 7D11, and 7A2 are included in clone 7144. The primer pair 7144T7F/R, from the T7 end of 7144, detected no new positive clones from the PAC library but did detect a new P1 clone, 8079 (Fig. 1). A 10-kb *Eco*RI fragment common to 8079 and 7144 was then used to identify new segments of the MYO7A transcript (Chen *et al.*, 1996), allowing assembly of a cDNA contig that corresponds to a 7351-bp transcript.

PCR primers prepared from the sequence of the 3′ end of the MYO7A cDNA contig did not amplify the expected product from P1 clone 8079. Sequence from the T7 end of 8079 indicated that this clone terminated within the MYO7A transcript in the intron following exon 44. A PCR primer pair designed from sequence obtained from the T7 end of 8079 failed to detect any new P1 or PAC clones. Two sets of primer pairs, M7C6702F/M7C6924R and, M7C7200F/M7C7306R, designed from the sequence of the 3′ end of the long MYO7A transcript, detected a single P1 clone, 10419. A primer, M7C6100F, from transcript sequence contained in 8079 and a primer, M7C6547R, from transcript sequence contained in 10419 were used to PCR amplify a gap fragment of 4.5 kb from genomic DNA. The PAC clone 7144, P1 clones 8079 and 10419, and the 4.5-kb PCR product constitute a contig that contains the entire MYO7A gene (Fig. 1).

Intron–Exon Structure of the MYO7A Gene

Once a physical contig was constructed for MYO7A, the intron–exon boundaries could be determined. Primers were designed based on the sequence of the MYO7A transcript and used to sequence directly from the cosmids, P1 clones, PAC clone, or long PCR product. Intron–exon boundaries were then identified by comparing MYO7A genomic and transcript sequence. Forty-nine exons were discovered in the 7351-bp transcript. The intron–exon boundary sequences are shown in Table 1, along with the exon and intron sizes. A graphic summary of these data is also shown in Fig. 1. The size of introns was determined by directly sequencing the intron (1–14) or by PCR amplification of the introns (15–48) from both total genomic DNA and the P1 or PAC clone from that region.

The first exon and part of the second exon of MYO7A consists of 5′-untranslated sequence. The motor domain begins within exon 2 and extends through exon 19 (Table 1 and Fig. 1). Exons 2 through 4 are flanked by large introns and occupy a region of about 25 kb. Exons 2 and 3 were identified during the search of cosmid clones for candidate genes but had no homology to previously reported myosin cDNAs. Exons 5 through 14 are clustered, with an average intron size of about 600 bp. Exons 15 through 22 are interrupted by introns that range from about 3000 to 12,000 bp. A cluster of exons precedes the 27th intron, which is relatively large (5 kb) and contains an alternatively spliced 3′ end with a consensus poly(A) addition site. The remaining exons 28 through 48 are relatively evenly distributed with an average intron size of about 1200 bp. The final 3′ exon is separated from the penultimate exon by about 4 kb.

DISCUSSION

Our initial strategy for obtaining a sequencible contig for MYO7A depended on subcloning YACs into cosmids. Cosmids containing the 3′ end of the MYO7A could not be obtained from critical YACs, so P1 clones, a PAC clone, and long PCR segments were used to construct a sequencible contig. No P1 or PAC clone that contained the entire MYO7A gene was found. The P1 clone 8079 (approximately 85 kb) contained exons 4 through 44. Adding the introns and exons excluded from this clone at the 5′ end (approximately 18.5 kb) and at the 3′ end (approximately 10.5 kb) results in an estimate that 114 kb are required for the 49 exons and 48 introns of MYO7A on chromosome 11q13.5. This estimate of the gene size does not include the promoter region, which has not yet been characterized. Both linkage analysis and FISH analysis of metaphase chromosomes indicated that OMP was centromeric to MYO7A (unpublished data). The Genomic Database assigns OMP to a range 11q14 to 11q21, while MYO7A is assigned to 11q13.5.

Of the 49 exons included in the long major transcript of MYO7A, 48 include coding sequence. The average size of the exons excluding the most 5′ and 3′ exons is 140 bp (Table 1). Only 553 bp of transcribed sequence is found in the first 20 kb of the gene. Motor domain exons are clustered in the next 7–8 kb followed by a series of larger exons separating the motor from the tail domain that are relatively evenly spaced, except for larger exons following exon 27 and preceding the 3′ exon 49 (Fig. 1). Most of the intron–exon junctions followed the canonical GT-AG rule. Two exceptions were noted and are yet to be resolved. A TG was found at the 3′ splice acceptor site of alternative exon 26 (26a;

TABLE 1

Organization of the Human MYO7A

Exon No.	3′ splice acceptor	5′ end of exon	Exon size	3′ end of exon	5′ splice donor	Intron size
1			221	CCCAG	gtaagg	2500
2	tggtctctctccctgcag	AACTG	64	CAG CAG	gtcagt	12000
3	agactctctctcgcccatag	GGG GAC	114	GAC AAT	gtgagt	4000
4	gacgttctggctccccgcag	GAA CAC	153	ATC TAC	gtgagt	8000
5	ccatctcttgctgcccgcag	ACG TAT	185	ATC AG	gtgggc	533
6	gcxactccctccctctgcag	T GGG GAA	122	CTG GAA G	gtagga	80
7	acaccctactcactccgcag	CA TTT GGG	143	CGC CAG	gtgggc	274
8	cactgtgcccacattttcag	GCC CTG	114	GCC ATG	gtgagg	884
9	tctcccaccctgcccaccag	GGT AAC	154	TAT GAG G	gtgagg	1017
10	gtgccccttgctgccaacag	CA CGC	77	CTT GAG	gtcagt	639
11	tggggtgttgcctgtaccag	GTG AAC	120	GTA AAG	gtgggc	690
12	ggctgatcactgtctttcag	GGG ATC	143	AAC AG	gtaccg	1005
13	ccatccctgtgcccctgcag	C TTT GAG	211	CCC AAG	gtgggc	522
14	gctgcccctccactccccag	GGC ACA	136	ACC CAA G	gtacag	3129
15	ctgctgtttgctgcttgcag	GC TTC CTG	107	GCC ATG	gtaagc	4000
16	gctgccgtccgtcccccag	GGC GCC	138	CCC ATG	gtgagt	2072
17	ctggctgtgtgcctggcag	CTG TTC	159	AAG CAG	gtacag	600
18	tgttccccatcctcactccag	GGC GAC	93	CTG AAG	gtgagc	2249
19	ctcaggagctctgcctcctag	GAC CAC	95	GAC AG	gtgcgt	4000
20	cgcgccactactgctgtttcag	G TCT AAC	85	GGG CTG	gtgagc	12000
21	tggggtggctgtccttgcag	ATG CGT	219	GCT GAG	gtgagg	815
22	gatgccccctccctcag	TAT CTG	108	CAT CAG	gtgagc	4000
23	gtgtcaccccaattgcccag	GAG CGC	210	TTT GAG	gtacca	1951
24	catacctcttgtctccttcag	GAC CTG	204	CAG CTG	gtaagg	474
25	acatgcgcgctctgccccag	GCA GCC	177	GGC GAG	gtgagg	677
26a	acttgaccctgatccctgctg	GTC CTG CAG	99	GAG GAG	gtgagg	1606
26	gatccctgctggtcctgcag	GCC CAG	90	GAG GAG	gtgagg	1606
27-3′a	gtcctctccctctggcccagg	GTG ACC	281			
27	gtcctctccctctggcccag	GTG ACC	128	CTC CG	gtcagt	5260
28	ttgccctgctgcctgcccag	G GAC GAG	127	GTC AAG	gtagga	764
29	cagcgggtactctggctgag	TAC CTG	120	CTG CAG	gttcgt	520
30	accctcttggggcactccag	GCC ACC	174	GAC AAG	gtatgg	607
31	ccctcccctctgtgcccacag	GTG T CCT	228	GAG AAG	gtgagt	1438
32	cgaggcctcccccccacctag	GAG GAC	171	AAG AAG	gtagaa	1535
33	tctgtccctctctccttccag	GGG ATT	118	TTC TCA G	gtaccc	1082
34	gtgccttggtctcgtcccag	GC CCC AGT	127	AGC AGC AG	gtgagg	1163
35a	aactggcctgatctccttcag	GGA GTG	284	AAC CCC G	gtgagt	1149
35	ttctccgtgttggtcctgcag	G GGA GCG	170	AAC CCC G	gtgagt	1795
36	ccactcacctctgctctacag	CA GGC GAG	191	ATT GTG	gtatgt	842
37	cccctgtctcttggtccctag	GCC CTG	125	TAC TTC AG	gtgatg	769
38	ggtccctgtgctgcggcag	G CCC CCA	158	TTC ATT G	atatccg	1050
39	ccttccttgacaggccccag	CT GTG CTC	154	CAC ATC AG	gtgagc	1417
40	agcccacgcctcctcctgcag	G TAC AGC	156	GCC CTG AG	gtacag	677
41	tgcccctgctgccttttcag	A AAC GGG	106	GAC GAG	gtgagg	1275
42	gtgctcactgcccctcccag	GCC TTC	114	GAC AAG	gtgggt	1433
43	atgccctgaccccag	GTC CTC	88	AAG GAC G	gtaatg	200
44	ccgactgccctgtgctgcag	GA ATT GTG	107	TAC CAG	gtgggc	2603
45	acctgctctgtctctgacag	GAG TTG	186	AAG CGG	gtgagc	638
46	gaccgccctgtccccatag	TCC ATC	111	TTC GAG	gtggag	1163
46a	gaccgccctgtccccatag	TCC ATC	117	GTG GAG	gtacac	1157
47	cttctcatctttttttctag	CAA ACT	84	ACG AAG	gtgagc	988
48	atgccctttctgctcccccag	GAT ATC	120	TCA CTG	gtgagg	4000
49-3′	ttctcaccctgcttcctag	GGC TAC	634			

Table 1 and see below), and an AT was found at the 5′ splice donor site of exon 38. Repeated sequence of these regions indicates that a mutation may have existed in the original clone or arisen during the process of cloning. Exons 2 through 19 specify the motor region of this myosin (Table 2 and Fig. 1). Exon 6 contains the ATP binding domain, whereas exon 16 encodes the actin binding domain. The splice junction between exons 4 and 5 encoding IY/TY is conserved in all metazoan myosins, including both conventional and unconventional myosin types (Strehler *et al.*, 1986). The splice junctions for exon 5/6 and exon 6/7 are also conserved with other myosins, such as bovine brush border myosin I (Kawakami *et al.*, 1992). Interestingly, the junctions for exons 15/16 and 16/17, although not conserved in sequence, are conserved in location, as exon 16 encodes

TABLE 2

Exon Location of Functional Domains for MYO7A

Nucleotide	Amino acid	Exon	Domain
268–2494	1–742	2–19	Motor
748–761	161–165	6	ATP binding
2164–2178	633–637	16	Actin binding
2495–2837	743–856	19–21	IQ
2838–3087	857–939	21–23	Coiled-coil
3088–3322	940–1017		
3323–3428	1018–1052	24–25	Myosin IV like
3429–3699	1053–1142		
3700–3919	1143–1216	27–29	Myosin IV like
3920–4049	1217–1259		
4050–4920	1260–1549	30–35	Talin-like
4921–5401	1550–1709		
5402–5732	1710–1819	38–40	Myosin IV like
5733–5877	1820–1867		
5878–6763	1868–2162	41–49	Talin-like
6764–6792	2163–2172		

a loop found on the surface of the myosin motor, unique to each myosin, which serves as an actin binding contact site. The final conserved junction is seen for exon 18/19, which defines the end of the motor domain and is conserved for all myosins.

MYO7A has five IQ motifs, repeats that serve as the sites for myosin light chain binding, located in three exons. Each of these exons contains a precise number of complete IQ repeats; exon 19 has one (Table 2), exon 20 has one, and exon 21 has three. These IQ domains bind calmodulin or other members of the EF-hand family of calcium binding proteins and serve to regulate the myosin motor. Alternative splicing is responsible for controlling the number of IQ domains in at least two myosins, chicken brush border myosin I (Halsall and Hammer, 1990) and rat myosin I (myr 1) (Ruppert *et al.*, 1993). Alternative splicing has not been detected in this region for the MYO7A gene, but exons in the region are flanked by large introns that could contain additional alternatively spliced exons that have not yet been detected.

We discovered a possible polymorphism in the genomic sequence by comparison of the cDNA sequence of the smaller transcript (Chen *et al.*, 1996) with the genomic sequence derived from P1 clone 8079 using primers designed from that cDNA sequence. When compared with sequence from the short transcript clone, the genomic sequence from P1 clone 8079 had an unexpected 22-bp insertion 10 bp downstream from the 3′ end of exon 27 in intron 27. This region of intron 27 is also in the alternative 3′ exon 27, which terminates the MYO7A short transcript 2 (Chen *et al.*, 1996). This insertion did not disrupt the 5′ splice donor site of intron 27 but shifted the reading frame, uncovering a stop signal that would result in a different C-terminus for the protein product of the short transcript. Instead of a 32-amino-acid C-terminus, this insertion would specify a unique 10-amino-acid C-terminus SVPGGGDTRA. The frequency of this variant in the general population is currently under study. The function of the protein encoded by the smaller (4.2 kb) transcript is not known. It is not known whether the integrity of the C-terminus of this polypeptide is required for its function.

The work reported here will allow completion of our mutation analysis of the 312 Usher I patients diagnosed as having Usher syndrome type Ib. Distribution of mutations in the first 14 exons shows no unusual distribution but suggests that mutations will be found throughout the MYO7A gene (Weil *et al.*, 1995; Weston *et al.*, 1996). Initially our analysis will focus on searching for point mutations that disrupt critical functional domains, alter sequence required for exon–exon splicing, or provide stop codons for premature termination of translation. Potential functional domains have been identified by analysis of the primary structure and include the IQ domains, where mutations could disrupt the number or juxtaposition of myosin light chain binding sites involved in motor regulation. Mutations in talin domains might alter the membrane or cytoskeletal binding properties of this myosin, interfering with its function in moving such substrates within the cell. Mutations in the "myosin IV like domain" (Myth4) could shed light on the function of this domain, which as yet is unascribed.

Mutations not found in the exonic sequence or their boundary will be difficult to determine because of the size of this gene. Of 114 kb only about 7.5 kb is transcribed. The search for promoter and intronic mutations might be accelerated if mRNA from these individuals could be analyzed but preliminary studies indicate that the gene is not expressed in conveniently sampled tissue.

Correlation of phenotypic idiosyncrasies with specific mutations will allow us to understand better the disease prognosis. Efficient screening of the deaf population for early diagnosis of Usher syndrome is needed for career planning to enhance safety and for the genetic counseling of the families involved. This work lays the foundation for future investigations into the nature, distribution, and frequency of different mutations causing Usher syndrome type Ib.

ACKNOWLEDGMENTS

This work was supported by NIH-NIDCD PO1 DC01813-01 (W.J.K.), NIH R01 EY05627 (W.J.K.), NIH DCO2281 (D.P.C.), the Deafness Research Foundation (Z-Y.C.), and Grant 89-198 from The Foundation Fighting Blindness (Hunt Valley, MD). D.J.O. was supported by NIH NIDCD P60 DC00982-06. P.M.K. thanks Bi Fang Li, Peter Berglund, Arpad Lanyi, and Judy Kenyon for their patience and assistance.

REFERENCES

Albertsen, H. M., Abderrahim, H., Cann, H. M., Dausset, J., Le Paslier, D., and Cohen, D. (1990). Construction and characterization of a yeast artificial chromosome library containing seven haploid human genome equivalents. *Proc. Natl. Acad. Sci. USA* **87:** 4256–4260.

Bement, W. M., Wirth, J. A., and Mooseker, M. A. (1994). Cloning and mRNA expression of human unconventional myosin-IC: A homologue of amoeboid myosins-I with a single IQ motif and an SH3 domain. *J. Mol. Biol.* **243:** 356–363.

Chen, Z-Y., Hasson, T., Kelley, P. M., Schwender, B., Schwartz, M. F., Mooseker, M. S., Kimberling, W. J., Mooseker, M. S., and Corey, D. P. (1996). Molecular cloning and domain structure of human myosin VIIa, the gene product defective in Usher syndrome 1B. *Genomics* **36:** 1–9.

Chumakov, I., Rigault, P., and Guillou, S. (1992). Continuum of overlapping clones spanning the entire human chromosome 21q. *Nature* **359:** 380–387.

El-Amraoui, A., Sahly, I., Picaud, S., Sahel, J., Abitbol, M., and Petit, C. (1996). Human Usher 1B/mouse *shaker-1:* The retinal phenotype discrepancy explained by the presence/absence of myosin VIIA in the photoreceptor cells. *Hum. Mol. Genet.* **5:** 1171–1178.

Gibson, F., Walsh, J., Mburu, P., Varela, A., Brown, K. A., Antonio, M., Beisel, K. W., Steel, K. A., and Brown, S. D. M. (1995). A type VII myosin encoded by the mouse deafness gene *shaker-1. Nature* **274:** 62–64.

Green, E. D., and Olson, M. V. (1990). Systematic screening of yeast artificial chromosome libraries by use of the polymerase chain reaction. *Proc. Natl. Acad. Sci. USA* **87:** 1213–1217.

Guilford, P., Ayadi, H., Blanchard, S., Chaib, H., Le Paslier, D., Weissenbach, J., Drira, M., and Petit, C. (1994). A human gene responsible for neurosensory, non-syndromic recessive deafness is a candidate homologue of the mouse *sh-1* gene. *Hum. Mol. Genet.* **3:** 989–993.

Halsall, D. J., and Hammer, J. A. (1990). A second isoform of chicken brush boarder myosin I contains a 29-residue inserted sequence that binds calmodulin. *FEBS Lett.* **267:** 126–130.

Hasson, T., Heintzelman, M. B., Santos-Sacchi, J., Corey, D. P., and Mooseker, M. S. (1995). Expression in cochlea and retina of myosin VIIa, the gene product defective in Usher syndrome type Ib. *Proc. Natl. Acad. Sci. USA* **92:** 9815–9819.

Kaplan, J., Gerber, S., Bonneau, D., Melki, J., Rozet, J., Briard, M. L., Munnich, A., Dufier, J. L., and Frezal, J. (1992). A gene for Usher syndrome type I (USH1A) maps to chromosome 14q. *Genomics* **14:** 979–987.

Kawakami, H., Moriyoshi, K., Utsumi, T., and Nakanishi, S. (1992). Structural organization and expression of the gene for bovine myosin I heavy chain. *J. Biochem.* **111:** 302–309.

Kimberling, W. J., Moller, C. G., Davenport, S., Priluck, I. A., Beighton, P. H., Greenberg, J., Reardon, W., Weston, M. D., Kenyon, J. B., Grunkemeyer, J., Pieke Dahl, S., Overbeck, L. D., Blackwood, D. J., Brower, A. M., Hoover, D. M., Rowland, P., and Smith, R. J. H. (1992). Linkage of Usher syndrome type I gene (USH1B) to the long arm of chromosome 11. *Genomics* **14:** 988–994.

Kimberling, W. J., Weston, M. D., Moller, C., van Aarem, A., Cremers, C. W., Sumegi, J., *et al.* (1995). Gene mapping of Usher syndrome type IIa: Localization of the gene to a 2.1-cm segment on chromosme 1q41. *Am. J. Hum. Genet.* **56:** 216–23.

Mitelman, F. (Ed.) (1995). "ISCN (1995): An International System for Human Cytogenetic Nomenclature," Karger, Basel.

Ruppert, C., Kroschewski, R., and Bahler, M. (1993). Identification, characterization, and cloning of myr 1, a mammalian myosin I. *J. Cell Biol.* **120:** 1393–1403.

Sankila, E. M. (1994). The existence of Usher syndrome type III proven by assignment of its locus to chromosome 3q by linkage. *Am. J. Hum. Genet.* **55:** A15.

Smith, C. L., Klco, S. R., and Cantor, C. R. (1988). Pulsed-field gel electrophoresis and the technology of large DNA molecules. *In* "Genome Analysis: A Practical Approach" (K. E. Davies, Ed.), pp. 41–72, IRL Press, Oxford.

Smith, R. J. H., Lee, E. C., Kimberling, W., Daiger, S. P., Pelias, M. Z., Keats, B. J. B., Jay, M., Bird, A., Reardon, W., Guest, M., Ayagari, R., and Hejtmancik, J. F. (1992). Localization of two genes for Usher syndrome type 1 to chromosome 11. *Genomics* **14:** 995–1002.

Strehler, E. E., Strehler-Oage, M-A., Perriard, J-C., Periasamy, M., and Nadal-Ginard, B. (1986). Complete nucleotide and encoded amino acid sequence of a mammalian myosin heavy chain gene. *J. Mol. Biol.* **190:** 291–317.

Sumegi, J., Wang, J-Y., Zhen, D-K., Eudy, J. D., Talmadge, C. B., *et al.* (1996). The construction of a yeast artificial chromosome (YAC) contig in the vicinity of the Usher syndrome type IIa (USH2A) gene in 1q41. *Genomics* **35:** 79–86.

Talmadge, C. B., Zhen, D-K., Wang, J-Y., Berglund, P., Li, B. F., Weston, M. D., Kimberling, W. J., Zabarovsky, E. R., Stanbridge, E. J., and Klein, G. (1995). Construction and characterization of *Not*I linking library from human chromosome region 1q25–qter. *Genomics* **29:** 104–114.

Weil, D., Levi, G., Sahly, I., Levi-Acobas, F., Blanchard, S., El-Amraoui, A., Crozet, F., Philippe, H., Abitbol, M., and Petit, C. (1996). Human myosin VIIa responsible for the Usher Ib syndrome: A predicted membrane-associated motor protein expressed in developing sensory epithelia. *Proc. Natl. Acad. Sci. USA* **93:** 3232–3237.

Weil, D., Blanchard, S., Kaplan, J., Guilford, P., Gibson, F., Walsh, J., Mburu, P., Varela, A., Levilliers, J., Weston, M. D., Kelley, P. M., Kimberling, W. J., Wagenaar, M., Levi-Acobas, F., Larget-Piet, D., Munnich, A., Steel, K. P., Brown, S. D. M., and Petit, C. (1995). Defective myosin VIIA gene responsible for Usher syndrome type 1B. *Nature* **374:** 60–61.

Weston, M. D., Kelley, P. M., Overbeck, L. D., Wagenaar, M., Orten, D. J., Hasson, T., Chen, Z.-Y., Corey, D., Mooseker, M., Sumegi, J., Cremers, C., Moller, C., Jacobson, S. G., Gorin, M. B., and Kimberling, W. J. (1996). Myosin VIIA mutation screening in 189 Usher Syndrome Type 1 patients. *Am. J. Hum. Genet.* **59:** 1074–1083.