Am. J. Hum. Genet. 54:899-908, 1994

Evidence of Linkage Disequilibrium in the Spanish Polycystic Kidney Disease 1 Population

Belén Peral,*,† Christopher J. Ward,* José L. San Millán,† Sandra Thomas,* Raymond L. Stallings,‡,1 Felipe Moreno,† and Peter C. Harris*

*MRC Molecular Haematology Unit, Institute of Molecular Medicine, Headington, Oxford; †Unidad de Genética Molecular, Hospital Ramón y Cajal, Madrid; and ‡Life Sciences Division and Center for Human Genome Studies, Los Alamos Laboratory, Los Alamos, NM

Summary

Forty-one Spanish families with polycystic kidney disease 1 (PKD1) were studied for evidence of linkage disequilibrium between the disease locus and six closely linked markers. Four of these loci—three highly polymorphic microsatellites (SM6, CW3, and CW2) and an RFLP marker (BLu24)—are described for the first time in this report. Overall the results reveal many different haplotypes on the disease-carrying chromosome, suggesting a variety of independent PKD1 mutations. However, linkage disequilibrium was found between BLu24 and PKD1, and this was corroborated by haplotype analysis including the microsatellite polymorphisms. From this analysis a group of closely related haplotypes, consisting of four markers, was found on 40% of PKD1 chromosomes, although markers flanking this homogeneous region showed greater variability. This study has highlighted an interesting subpopulation of Spanish PKD1 chromosomes, many of which have a common origin, that may be useful for localizing the PKD1 locus more precisely.

Introduction

Autosomal dominant polycystic kidney disease (ADPKD) is the most common inherited disorder that leads to renal failure in adults, with a worldwide prevalence of approximately 1 in 1,000 (Gabow 1991). The disease is characterized by the enlargement, often massively, of kidneys because of the formation of innumerable fluid-filled cysts, which cause renal insufficiency in half the affected individuals before the age of 70 years. Although renal cysts do not always cause morbidity, ADPKD accounts for ~10% of all cases of end-stage kidney disease in Caucasians. Other manifestations of the disease are hepatic cysts, cardiac valvular abnormalities, cerebral aneurysms, and hypertension (Gabow 1991).

In the majority of cases, ADPKD is due to mutation of a locus, designated "polycystic kidney disease 1" (PKD1), which has been assigned to 16p13.3 (Reeders et al. 1985). However, in a minority of families, the ADPKD phenotype is not linked to PKD1 markers (Kimberling et al. 1988; Romeo et al. 1988; Bachner et al. 1990), and a second locus, on 4q, probably accounting for most of the remaining families, has recently been described (Kimberling et al. 1993; Peters et al. 1993). In the Spanish population the frequency of PKD1-linked mutations has been estimated at 85% (Peral et al. 1993), a value similar to that obtained from a study of 273 European families (Peters and Sandkuijl 1992). By study of PKD1 families by genetic linkage analysis, the PKD1 locus has been localized proximal to the probe GGG1 (D16S259), but distal to 26.6PROX (D16S125) and SM7 (D16S283), in a region spanning 600 kb (Harris et al. 1991; Germino et al. 1992; Somlo et al. 1992) (fig. 1).

To localize the PKD1 gene more precisely, we looked for evidence of linkage disequilibrium between the PKD1 locus and closely linked markers. This strategy

Received November 23, 1993; accepted for publication January 20, 1994.

Address for correspondence and reprints: Peter Harris, MRC Molecular Haematology Unit, Institute of Molecular Medicine, Headington, Oxford OX3 9DU, United Kingdom.

[1] Present address: University of Pittsburgh, Department of Human Genetics, Pittsburgh.

© 1994 by The American Society of Human Genetics. All rights reserved.

0002-9297/94/5405-0020$02.00

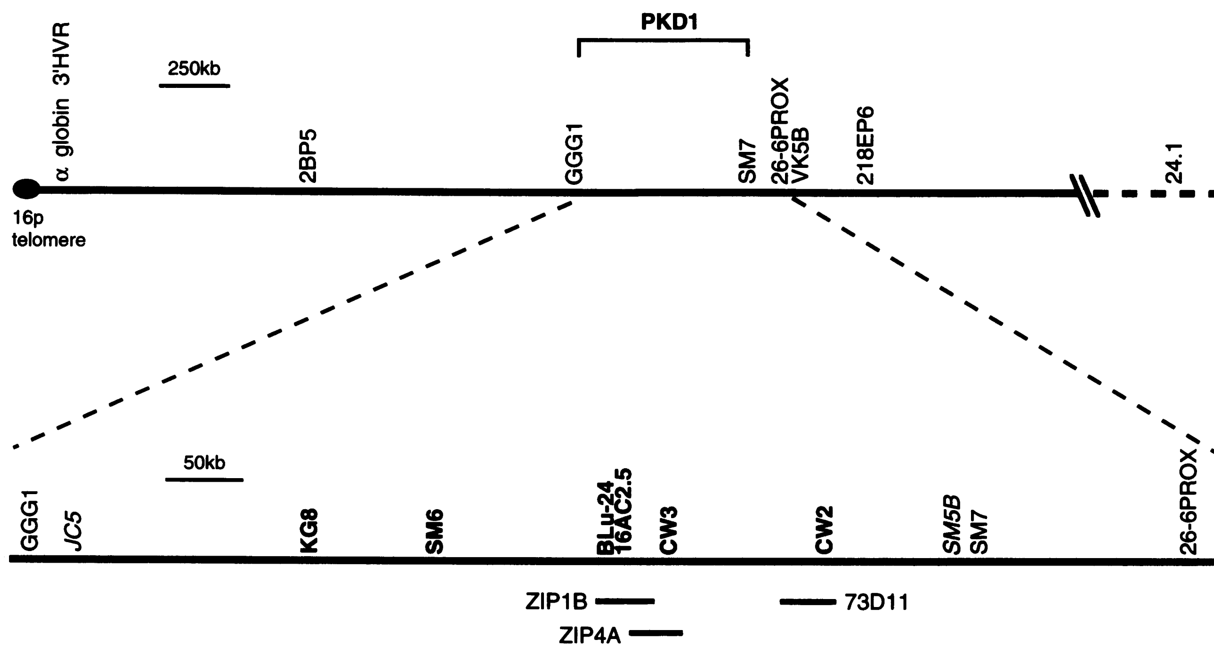

Figure 1 *Top,* Map of the terminal part of chromosome 16p, showing the PKD1 region and polymorphic markers previously used to analyze the Spanish PKD1 families. Data are summarized from Breuning et al. (1990), Harris et al. (1990, 1991, and submitted), and Germino et al. (1992). *Bottom,* Detailed map of the PKD1 area, showing the positions of markers used in this study (**boldface**) and the location of the STSs—*JC5* and *SM5B*—used to screen for YACs (*italics*). The locations of the cosmids ZIP1B and ZIP4A and 73D11 are also illustrated.

was previously used in the isolation of the cystic fibrosis gene (Kerem et al. 1989) and, recently, correctly identified the area containing the Huntington disease gene (Huntington's Disease Collaborative Research Group 1993), despite conflicting results from genetic linkage studies. However, this approach will only be successful if the new mutation rate is low and, hence, if many patients are related through a common ancestor. Previous studies have suggested a relatively high mutation rate for ADPKD (Dalgaard 1957), on the basis of an apparent high frequency of ADPKD subjects with unaffected parents and the perceived reproductive disadvantage of individuals with polycystic kidney disease. In a more recent report (Davies et al. 1991), five possible new mutations were identified in 209 Welsh ADPKD families from a population of 2.1 million, suggesting a lower, but nevertheless detectable, rate of new mutation. Despite these observations, linkage disequilibrium has been previously identified in PKD1 families; significant allelic association was found between the proximal flanking marker—VK5B (D16S94)—and PKD1, in a preliminary analysis of 33 Scottish ADPKD families (Pound et al. 1992).

We have studied 41 previously characterized Spanish PKD1 families in which no evidence of linkage disequilibrium between PKD1 and seven markers that flank the PKD1 locus (3'HVR, 2BP5, GGG1, 26.6PROX, VK5B, 218EP6, and 24.1; fig. 1) had been found (Peral et al. 1993; authors' unpublished results). We have now refined this study by using six markers that map in the PKD1 candidate region. Our results indicate that a wide variety of haplotypes are represented on PKD1 chromosomes, suggesting that many different independent mutations have given rise to PKD1. Nevertheless, within this population there is evidence of a subgroup of PKD1 chromosomes, present in 40% of affected families that may be ancestrally related.

Subjects, Material, and Methods

Families and Patients

The diagnosis of ADPKD was based on the finding of at least one cyst in one kidney and of two or more in the other, by ultrasonic scanning (Bear et al. 1984). Each family included in this study had at least either three affected persons or two affected and two unaffected members, although most families were considerably larger. All the families had been previously characterized as being PKD1; a heterogeneity test was carried out using the HOMOG program (Ott 1985), and families in

Table 1

Sequence of Primers Used for PCR Amplification of Markers Described in Present Study

Marker	Locus Symbol	Repeat Sequence[a]	PCR Primers	
SM6	D16S665	$(TG)_{10}C(CG)_{10}TGC\ (GT)_3$	SM6A	5'-AGCTGGGGTCTCAGGGGAGCT-3'
			SM6B	5'-GCGCACACAGCACTAACACG-3'
CW3	D16S664	$(AC)_{20}$	CW3A	5'-AATATCTCATCAGACTTTGC-3'
			CW3B	5'-TAAATTGTTGTATGTATTTTAC-3'
CW2	D16S663	$(AC)_{26}$	CW2A	5'-GTCTTTCTAGGAATGAAATCAT-3'
			CW2B	5'-ATTGCAGCAAGACTCCATCT-3'
BLu24	D16S662	...	BLuA	5'-AATTCCCCCTCGGCTAAGCTG-3'
			BLuB	5'-TGTCGAGTTGTGAGCCCCTGC-3'

[a] Size of most frequent allele.

which the probability of linkage to PKD1 was >.90 were included for further analysis (Peral et al. 1993).

Isolation of Markers

YACs were isolated from the CEPH YAC library (Albertsen et al. 1990) by the PCR screening method (Green and Olson 1990) with primers for SM5B (Harris et al. 1991) and, for the locus JC5 (which was subcloned from the cosmid JCII; Germino et al. 1990), with the primers JC5E(A) 5' GACTCTTAAGGGTCTATCCCT 3' and JC5E(B) 5' AGCCAGGCTCTAGATCACTCA 3', which yielded a product of 110 bp. DNA was isolated, either in solution or in agarose blocks, from the yeast and was characterized by standard methods. Alu-PCR was performed as described by Nelson et al. (1989), with the primer 938 (Hirst et al. 1991). A partial *Mbo*I library in bacteriophage λ EMBL3 was produced from total yeast DNA containing the YAC y77C8 (described below). Cosmids were isolated from the Los Alamos human chromosome 16 library (Stallings et al. 1990), and fragments containing $(CA)_n$ were identified, isolated, and sequenced as described elsewhere (Harris et al. 1991).

Characterization of Polymorphisms

The primer sequences used to amplify the newly described microsatellites, as well as the nature of the repeated areas, are shown in table 1. The previously described CA repeats—16AC2.5 (Thompson et al. 1992) and KG8 (Snarey et al., submitted)—were identified independently (see Results), and thus novel primers were used to amplify these markers: for 16AC2.5, CW4A 5' AAGGCTGGCAGAGGAGGTGA 3' and CW4B 5' CTAATCGGCGGGGAGCTCTA 3'; and, for KG8, BP1A 5' CACAGCCAGCTCCGAGGG 3' and BP1B 5' TCCTCCTGGGGGCTGGCTC 3'. The microsatellites were amplified and scored on native acrylamide gels as described elsewhere (Harris et al. 1991). The $MgCl_2$ concentrations and annealing temperatures used for amplification of the microsatellites were as follows: KG8, 0.5 mM, 58°C; SM6, 1.25 mM, 60°C; 16AC2.5, 1.5 mM, 60°C; CW3, 2 mM, 52°C; and CW2, 1.25 mM, 55°C. An *Rsa*I polymorphism (BLu24) was identified with BLu2, a 1.5-kb *Sau*3AI single-copy subclone of ZIP1B, by the method of Murray et al. (1987), using a somatic cell hybrid (N-BuHIA) as the source of a normal single chromosome 16. The primers used to amplify the BLu24 RFLP are shown in table 1, with amplification as for the microsatellites with 0.75 $MgCl_2$ and annealing at 65°C. The resulting product was digested with *Rsa*I and was visualized on a 2% agarose gel with alleles of 529 bp, for L1, and 413/116 bp, for L2.

Statistical Analysis

By analysis of pedigrees, we unequivocally identified the allele associated with PKD1 for each marker. To test for linkage disequilibrium a χ^2 test with the null hypothesis of no linkage disequilibrium was calculated for BLu24, with 1 df. For multiallelic loci, we used a previously defined χ^2 statistic (Hill 1975; Weir and Cockerham 1978).

Results

Isolation of Polymorphic Markers

To provide new markers for detailed genetic analysis of PKD1 families, microsatellite markers were sought within the PKD1 candidate region between GGG1 and SM7 (see fig. 1). In an attempt to clone this region, the YACs—y77C8 and y376B1—were isolated from the

Table 2

Details of Polymorphic Markers Used in Present Study

Allele	Allele Length (bp)	No. (%) of Chromosomes		χ^2	df	*P*
		Normal	Affected[a]			
KG8 (*n* = 148[b]):				6.06	7	.53
A1	130	7 (4.7)	1 (2.4)			
A2	128	15 (10.1)	9 (22.0)			
A3	126	1 (0.6)	1 (2.4)			
A4	124	5 (3.4)	1 (2.4)			
A5	122	95 (64.2)	23 (56.1)			
A6	120	2 (1.3)				
A7	118	10 (6.8)	2 (4.9)			
A8	116	13 (8.8)	4 (9.8)			
SM6 (*n* = 218[b]):				18.95	15	.22
B1	140	3 (1.4)				
B2	138	2 (0.9)				
B3	136	5 (2.3)	1 (2.4)			
B4	134	6 (2.8)				
B7	128	1 (0.5)				
B8	126	2 (0.9)	4 (9.8)			
B9	124	53 (24.3)	12 (29.3)			
B10	122	108 (49.5)	18 (43.9)			
B11	120	4 (1.8)				
B12	118	5 (2.3)				
B13	116	16 (7.3)	5 (12.2)			
B14	114	1 (0.5)				
B15	112	4 (1.8)	1 (2.4)			
B17	108	1 (0.5)				
B24	94	4 (1.8)				
B28	86	3 (1.4)				
16AC2.5 (*n* = 218[b]):				7.29	8	.50
C2	121	2 (0.9)				
C3	119	18 (8.2)	4 (9.8)			
C4	117	49 (22.5)	9 (22.0)			
C5	115	49 (22.5)	9 (22.0)			
C6	113	35 (16.0)	2 (4.9)			
C7	111	8 (3.6)	3 (7.3)			
C8	109	51 (23.4)	14 (34.1)			
C9	107	5 (2.3)				
C10	105	1 (0.5)				
CW3 (*n* = 214[b]):				9.38	6	.15
D3	87	4 (1.9)	3 (7.3)			
D4	85	42 (19.6)	4 (9.8)			
D5	83	26 (12.1)	4 (9.8)			
D6	81	123 (57.5)	28 (68.3)			
D7	79	12 (5.6)				
D8	77	6 (2.8)	2 (4.9)			
D9	75	1 (0.5)				
CW2 (*n* = 220[b]):				12.74	12	.39
F2	129	1 (0.5)				
F3	127	9 (4.1)	1 (2.4)			
F4	125	20 (9.1)	2 (4.9)			
F5	123	64 (29.1)	12 (29.3)			
F6	121	38 (17.3)	3 (7.3)			
F7	119	34 (15.5)	11 (26.8)			
F8	117	18 (8.2)	4 (9.8)			
F9	115	1 (0.5)				
F10	113	1 (0.5)	1 (2.4)			

(*continued*)

Table 2 (continued)

Allele	Allele Length (bp)	No. (%) of Chromosomes		χ^2	df	*P*
		Normal	Affected[a]			
CW2 (*n* = 220[b]) (*continued*):						
F11	111	3 (1.4)				
F12	109	3 (1.4)	2 (4.9)			
F13	107	27 (12.3)	4 (9.8)			
F14	105	1 (0.5)	1 (2.4)			
BLu24 (*n* = 208[b]):				6.81	1	.009
L1	529	158 (76.0)	23 (56.1)			
L2	413/116	50 (24.0)	18 (43.9)			

[a] Forty-one affected chromosomes were analyzed.
[b] No. of normal nonrelated chromosomes analyzed in the Spanish population.

CEPH human YAC library by using the sequence-tagged sites (STSs) JC5 and SM5B, respectively (fig. 1). y376B1 was unstable, and, although the isolated ends of the YAC were mapped in genomic DNA, 350 kb apart (right end between 26-6 DIS and 26-6 PROX and left end between SM6 and GGG1), analysis by pulsed-field gel electrophoresis revealed deleted YACs of 290 and 150 kb. Alu-PCR was used to generate human-specific probes from this YAC, which were hybridized to a cosmid library. One positive, novel cosmid, 73D11, was obtained, which mapped ~100 kb distal to SM5B. Two $(CA)_n$ repeats were isolated from this cosmid: CW1 (the same as W5.2; Snarey et al., submitted) and CW2 (fig. 1).

y77C8 is a 300-kb YAC but has a complex rearrangement, in that it both is chimeric with chromosome 3 and has a large interstitial deletion within the segment of chromosome 16 DNA. The junction between the chromosome 3 material and the chromosome 16 material was subcloned in the bacteriophage λ77A. An internal 4.2-kb *Bam*HI fragment from this phage, which mapped to the proximal part of the PKD1 region, was used to isolate the cosmids, ZIP1B and ZIP4A; these two overlapping cosmids contain the microsatellites CW4 (same as 16AC2.5; Thompson et al. 1992) and CW3, respectively (fig. 1). An additional probe, BLu24, which identifies an *Rsa*I polymorphism, was isolated from ZIP1B. The orientation of these markers on the chromosome was determined from a crossover between CW3 and 16AC2.5, which placed CW3 proximal to 16AC2.5, and, with additional mapping, BLu24 was located 5–10 kb distal of 16AC2.5 (fig. 1). The precise locations of CW2 and CW3 were determined by mapping them to the previously described PKD1 cosmid contig (Germino et al. 1992; Snarey et al., submitted). The two most distal microsatellites were isolated from clones described elsewhere—SM6 from SMII and BP1 (same as KG8; Snarey et al., submitted) from AH4 (European Chromosome 16 Tuberous Sclerosis Consortium 1993; fig. 1).

Analysis of PKD1 Families with the Polymorphic Markers

A total of 369 individuals from 41 Spanish PKD1 families, 159 of whom were affected, were analyzed with the five microsatellites—KG8, SM6, 16AC2.5, CW3, and CW2—and with the RFLP marker BLu24. No recombination between any of the markers studied and the PKD1 locus was observed in these families. The allele sizes and frequencies for each of the microsatellites and for BLu24, analyzed on normal and PKD1 chromosomes in the Spanish population, are summarized in table 2. Each of the microsatellite markers is highly informative, with an observed heterozygosity of .56, .69, .81, .60, and .83, for KG8, SM6, 16AC2.5, CW3, and CW2, respectively. BLu24 has an observed heterozygosity of .36.

These results clearly show that, with all of the markers studied, many different alleles are associated with PKD1. If, by application of the χ^2 test, each of the microsatellites is tested formally for evidence of linkage disequilibrium with PKD1, the results in table 2 show that none of these gives a significant *P* value. This finding is emphasized if extended PKD1-associated haplotypes, including all six markers, are analyzed; different

haplotypes are represented on each of the 41 PKD1 chromosomes. These results are consistent with the view that many different mutations have given rise to PKD1 in this population. Alternatively, this population may contain a few ancient PKD1 mutations, but the original haplotypes associated with these could have diverged by recombination and marker mutation over a period of time.

Although no significant allelic association was evident with the microsatellites, we did detect linkage disequilibrium with the RFLP marker BLu24. In this case a significant P value was found ($P < .01$; table 2), with overrepresentation of the L2 allele on PKD1 chromosomes (44%, compared with 24% on normal chromosomes).

Extended PKD1 Haplotypes

This evidence of significant allelic association between the BLu24 allele L2 and PKD1 highlighted a subpopulation that could be studied for the presence of a common PKD1 haplotype. As a first step, we analyzed a haplotype consisting of the physically close markers: BLu24, 16AC2.5, and CW3 (see fig. 1), to determine which alleles were associated with PKD1, and we compared the results found on the BLu24, L1 and L2 chromosomes. The observed haplotype distribution is shown in table 3, with 16 different haplotypes discerned from 41 PKD1 chromosomes.

Twelve different PKD1 haplotypes were found associated with the L1 allele of BLu24. None of these were found significantly more frequently on PKD1 than on normal chromosomes. The most frequent haplotypes on PKD1 chromosomes were also overrepresented on the normal ones, with only one haplotype just present in the PKD1 population; and this haplotype was only on one chromosome. Although it is possible that these various haplotypes have evolved by recombination and marker mutations from one or a small number of common haplotypes, it seems most likely that many different PKD1 mutations have occurred on L1 chromosomes.

However, in the case of haplotypes associated with the L2 allele, we observed two haplotypes that were overrepresented in the PKD1 population. The haplotype most often associated with PKD1—L2,C8,D6—accounted for 30% of the disease chromosomes but for only 9% of normal ones; and the second most frequent L2,PKD1 haplotype—L2,C7,D6—was found on 7.3% of affected chromosomes, whereas only 1 of 208 normal chromosomes carried this haplotype (table 3).

The observation that the haplotypes associated with the L2 allele of BLu24 form a relatively homogeneous group strengthens the case—suggested by the allelic association with BLu24 (L2)—that many of these PKD1 chromosomes may be ancestrally related. If we examine the association between L2 and the D6,CW3 allele, we see that it is highly enriched in the PKD1 population, with the association not being explained by normal linkage disequilibrium between these markers. The L2,D6 haplotype is present on 89% (16/18) of PKD1 chromosomes associated with L2, compared with 58% (29/50) of normal L2 chromosomes. The other PKD1 haplotype with the L2 allele (L2,C8,D5) differs from the D6 allele by just one CA repeat. It is possible that CW3 has undergone a slippage mutation to gain a repeat unit and to alter the D6 allele (81 bp) to a D5 (83 bp) (see Discussion).

The overrepresentation of the C8,16AC2.5 allele with PKD1 (table 2) is entirely due to association in the L2 population (table 3). However, detailed analysis in this case shows that this association can largely be explained by very strong linkage disequilibrium between the C8 and L2 alleles in the normal population—38/50 (76%), compared with 14/18 (78%) in the PKD1 population. If, however, the argument of slippage mutation,

Table 3

Frequencies of BLu24, 16AC2.5, and CW3 Haplotypes on PKD1 and Normal Chromosomes

Allele			No. (%) of Chromosomes	
BLu24	16AC2.5	CW3	PKD1	Normal
L2	C8	D6	12 (29.26)	19 (9.13)
L2	C7	D6	3 (7.32)	1 (.48)
L2	C8	D5	2 (4.88)	8 (3.85)
L2	C4	D6	1 (2.44)	6 (2.88)
Others			0	16 (7.69)
L1	C5	D6	4 (9.76)	24 (11.54)
L1	C4	D6	4 (9.76)	22 (10.58)
L1	C3	D6	3 (7.32)	9 (4.33)
L1	C5	D4	2 (4.88)	18 (8.65)
L1	C5	D3	2 (4.88)	1 (.48)
L1	C5	D5	1 (2.44)	3 (1.44)
L1	C4	D4	2 (4.88)	13 (6.25)
L1	C4	D3	1 (2.44)	1 (.48)
L1	C4	D8	1 (2.44)	0
L1	C6	D6	1 (2.44)	26 (12.50)
L1	C6	D8	1 (2.44)	2 (.96)
L1	C3	D5	1 (2.44)	5 (2.40)
Others			0	34 (16.35)
Total			41	208

Table 4

Extended Haplotypes on PKD1 Chromosomes with the BLu24,L2 Allele

KG8	SM6	BLu24	16AC2.5	CW3	CW2
A5	B10	L2	C8	D6	F7
A2	B10	L2	C8	D6	F7
A1	B10	L2	C8	D6	F7
A7	B10	L2	C8	D6	F6
A2	B10	L2	C8	D6	F5
A5	B10	L2	C8	D6	F12
A5	B10	L2	C8	D5	F4
A5	B10	L2	C8	D5	F10
A2	B10	L2	C7	D6	F7
A3	B10	L2	C7	D6	F7
A2	B9	L2	C8	D6	F7
A5	B9	L2	C8	D6	F7
A2	B9	L2	C8	D6	F14
A8	B9	L2	C8	D6	F5
A8	B9	L2	C7	D6	F5
A5	B8	L2	C8	D6	F8
A2	B8	L2	C8	D6	F7
A5[a]	B13	L2	C4	D6	F8

NOTE.—Each haplotype is present on one PKD1 chromosome. (The extended haplotypes of PKD1 chromosomes associated with the L1 allele can be obtained from the author on request.)

[a] This PKD1 haplotype is very different from the others and is probably an independent mutation.

as described above for CW3, is used to suggest that the C7 allele of 16AC2.5, seen on three L2-associated PKD1 chromosomes, is also derived from the more frequent C8 allele, then 17 of the 18 L2-associated PKD1 haplotypes may have a common origin. The other L2,PKD1 chromosome has a much larger 16AC2.5 allele (C4) and so is probably due to an independent mutation.

If the L2,C8,D6 haplotype is a PKD1 founder chromosome, further analysis of chromosomes with this or closely related haplotypes, with the proximal marker CW2 and distal flanking markers SM6 and KG8 may be expected to also show similarities. Table 4 lists the extended haplotypes of the 18 L2,PKD1 chromosomes. We can see from this analysis that 18 different haplotypes are associated with PKD1; each patient has a different haplotype. Clearly, at first sight, this result does not lend support to the idea of a common origin. However, further analysis does show some interesting associations between each marker and the L2 allele; either a much more limited range of alleles is associated with the L2,PKD1 chromosomes than in the normal L2 population, or a specific allele is highly enriched in the L2,PKD1 population, compared with normal L2 chromosomes.

For SM6 we find just three alleles, which differ by only 2 bp from each other, associated with the L2,PKD1 population, compared with 11 different alleles on normal L2 chromosomes. The most frequent allele, B10, is found on 55.6% of L2,PKD1 chromosomes and on 44% of normal L2 chromosomes. Seventeen of the 18 L2,PKD1 chromosomes have similar haplotypes with the four markers—SM6, BLu24, 16AC2.5, and CW3—with each of these haplotypes found more frequently on PKD1 chromosomes. The most distal marker, KG8, shows many different alleles (six of a total of eight observed) associated with the disease on L2-bearing chromosomes, and these differ in size by as many as 14 bp. Nevertheless, A2, the allele most enriched in the PKD1 population (table 2), is represented on 33.3% of affected L2 chromosomes but on only 8% of normal L2 chromosomes. On the proximal side, again one allele—F7 of CW2—is present on a much higher proportion of L2,PKD1 chromosomes (44%), compared with normal chromosomes (26%). However, overall, eight CW2 alleles are associated with PKD1 on L2 chromosomes, and, although some of these differ from the most frequent (F7) allele by just one dinucleotide, others differ by as many as seven CA repeats. Beyond the immediate PKD1 region, we could not find evidence of association between the L2,PKD1 chromosomes and particular alleles of the flanking distal (GGG1) or proximal (26.6 PROX, VK5B, or 218EP6) markers.

Discussion

We have described three new microsatellite markers from the PKD1 region, which, in association with four previously described CA repeats from this area (Harris et al. 1991; Thompson et al. 1992; Snarey et al., submitted), provide a formidable set of markers for both linkage analysis and presymptomatic diagnosis of PKD1. The recent identification of the tuberous sclerosis 2 locus (TSC2) in this area (European Chromosome 16 Tuberous Sclerosis Consortium 1993) means that these polymorphisms will also be useful for genetic analysis of TSC2 families. Studies are underway to determine if the PKD1 candidate region can be further refined by analysis of known crossovers between PKD1 and existing flanking markers, using these microsatellites. However, the number of known crossovers is small, and the size of the candidate region, at ~600 kb (Harris et al. 1991; Germino et al. 1992; Somlo et al. 1992), will be

difficult to further refine by this method. For this reason we have searched for evidence of linkage disequilibrium between the microsatellites and PKD1, to determine if this method would help to localize this locus. However, the data from this study reveal a complex structure to the Spanish PKD1 population and indicate that the results need to be interpreted with caution.

The clearest result from this study of linkage disequilibrium is that many different haplotypes, constructed with markers from within the PKD1 candidate region, are associated with PKD1. The simplest explanation for this finding is that many independent mutations have given rise to PKD1. This result is consistent both with other studies' description of a significant new mutation rate for PKD1 (Dalgaard 1957; Davies et al. 1991) and with our own observations of occasional cases that appear to be new mutations. Despite the evidence of multiple different PKD1 mutations, there is, nevertheless, evidence that part of this population may arise from a common ancestor.

The only significant linkage disequilibrium seen with a single marker in this study is with the RFLP probe, BLu24. It is likely that linkage disequilibrium is seen with this marker because the association is with the rare allele, L2 (seen in just 24% of normals and in 44% of PKD1 individuals), and therefore is more readily identified. Support for a common origin of at least part of the BLu24,L2,PKD1 population is found by analyzing the microsatellites flanking this marker. For markers KG8, 16AC2.5, CW3, and CW2 the alleles most enriched in the whole PKD1 population, compared with normal controls—i.e., A2 (22% in PKD1 and 10% in normal controls), C8 (34% in PKD1 and 23% in normal controls), D6 (68% in PKD1 and 57% in normal controls), and F7 (22% in PKD1 and 15% in normal controls) (see table 2)—are all overrepresented on L2,PKD1, but not on L1,PKD1, chromosomes. For KG8, 6 of the 9 A2,PKD1 alleles are associated with L2, while, for CW2, 8 of 11 F7,PKD1 alleles are on L2 chromosomes. The CW3 allele found most often on PKD1 chromosomes, D6, is particularly enriched in the L2,PKD1 population—89%, compared with 52% in L1,PKD1 chromosomes. All C8,CW4 alleles associated with PKD1 are on L2 chromosomes; however, this can be explained by strong linkage disequilibrium between CW4 and BLu24 (see Results).

In light of both the significant linkage disequilibrium seen between BLu24,L2 and PKD1 and the enrichment of specific alleles for the other markers in the L2,PKD1 population, why is no single haplotype seen that accounts at least for a proportion of this population? This might be partly explained by a significant rate of new mutation of the microsatellites themselves. If the central part of the haplotype including SM6, BLu24, CW4, and CW3 (table 4) is examined, a founder haplotype consisting of B10,L2,C8,D6 can be proposed that accounts for a third (6/18) of L2,PKD1 chromosomes. Furthermore, 11 of the remaining 12 L2,PKD1 haplotypes are similar to the founder and may have evolved from it because of microsatellite mutations.

It is thought that slippage during reannealing that occurs after DNA replication is the mechanism by which the characteristic 2-bp differences of microsatellite polymorphisms arise (Schlötterer and Tautz 1992). The mutation rate, manifesting as the heterozygosity at a microsatellite locus, is greater with larger runs of uninterrupted repeats (Weber 1990). However, observation of microsatellites in families, their allele distributions, and comparison with the larger minisatellites (Jeffreys et al. 1988) suggest that the generation of new alleles may be complex and that different microsatellites will mutate at different rates. We have seen the appearance of new microsatellite alleles during pedigree analysis, especially for SM6, indicating that this locus may be particularly prone to new mutation. It is also of interest that all of the markers studied show significant linkage disequilibrium with BLu24 (CW2, $P < .05$; CW3, $P < .001$; and 16AC2.5, $P \ll .001$), except for SM6 and the most distal, KG8. It is therefore possible that microsatellite slippage explains the variability of the markers SM6, 16AC2.5, and CW3 from the founder haplotype—in all but the last haplotype in table 4, which probably arose independently. It is worth noting that the marker showing greatest variability from the founder of these four is SM6, which probably has the highest new mutation rate.

Alternatively, it could be argued that different, independent mutations have occurred on the same or related founder haplotype. This is not that unlikely, because the alleles associated with L2 in this haplotype either show strong marker-to-marker disequilibrium (C8,16AC2.5) or are the most frequent alleles in the population (SM6, B10, and CW3,D6).

If we extend the haplotype to include the flanking markers KG8 and CW2, we can see that the L2,PKD1 population is much more varied and that the alleles seen with these two markers are not readily explained by slippage, as many of the markers differ greatly, in size, from the most enriched alleles, A2 and F7, respectively. It is attractive to suggest that the variability seen with these alleles is due to recombination between the PKD1 locus and these two markers. Such an explanation

would indicate that the PKD1 locus is situated in the ~300 kb between these markers. On the proximal side, this interpretation is supported by the observation of crossovers between PKD1 and proximal markers in this interval (Somlo et al. 1992). There is, however, no evidence of corresponding crossovers on the distal side, i.e., between KG8 and SM6.

It is possible, however, that the crossover hypothesis is too simple and that the L2,PKD1 population consists of several independent mutations as discussed above—e.g., one associated with KG8 allele A2, but others on A5 or A8 backgrounds, which nevertheless have similar SM6, CW4, and CW3 haplotypes. It may be possible to determine if this is the case by analyzing the L2,PKD1 families with a number of more stable, RFLP polymorphisms (such as BLu24), from across the candidate region, to see if a single haplotype can account for the majority of L2,PKD1 chromosomes.

Our studies have shed some light on the complexity of a PKD1 population. Although there is evidence of many different PKD1 mutations, there is also an indication that, of the population, ~40% may be ancestrally related. It is difficult, with the present evidence, to say precisely where the PKD1 locus lies, but this study has resulted in the identification of an interesting subpopulation, further study of which may help answer this question.

Acknowledgments

We thank Denis Le Paslier for YAC clones, Janette Lamb for the sequence of JC5, Reginald Boone for computing help, Douglas Higgs for helpful discussion, and Prof. Sir D. J. Weatherall for continued support. This work was supported by the Medical Research Council and the Fondo de Investigaciones Sanitarias of the Ministerio de Sanidad y Consumo of Spain (support to B.P. and J.L.S.M.) and by the EC Concerted Action "Toward Prevention of Kidney Failure Caused by Inherited Polycystic Kidney Disease."

References

Albertsen HM, Abderrahim H, Cann HM, Dausset J, Le Paslier D, 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

Bachner L, Vinet MC, Lacave R, Babron MC, Rondeau E, Sraer JD, Chevet D, et al (1990) Linkage study of a large family with autosomal dominant polycystic kidney disease with reduced expression. Hum Genet 85:221–227

Bear JC, McManamon P, Morgan J, Payne RH, Lewis H, Gault MH, Churchill DN (1984) Age at clinical onset and at ultrasonographic detection of adult polycystic kidney disease: data for genetic counselling. Am J Med Genet 18:45–53

Breuning MH, Snijdewint FGM, Brunner H, Verwest A, Ijdo JW, Saris JJ, Dauwerse JG, et al (1990) Map of 16 polymorphic loci on the short arm of chromosome 16 close to the polycystic kidney disease gene (PKD1). J Med Genet 27:603–613

Dalgaard OZ (1957) Bilateral polycystic disease of the kidneys: a follow-up of two hundred and eighty-four patients and their families. Acta Med Scand suppl 328:1–251

Davies F, Coles GA, Harper PS, Williams AJ, Evans C, Cochlin D (1991) Polycystic kidney disease re-evaluated: a population-based study. Q J Med 79:477–485

European Chromosome 16 Tuberous Sclerosis Consortium (1993) Identification and characterization of the tuberous sclerosis gene on chromosome 16. Cell 75:1305–1315

Gabow PA (1991) Polycystic kidney disease: clues to pathogenesis. Kidney Int 40:989–996

Germino GG, Barton NJ, Lamb J, Higgs DR, Harris P, Xiao GH, Scherer G, et al (1990) Identification of a locus which shows no genetic recombination with the autosomal dominant polycystic kidney disease gene on chromosome 16. Am J Hum Genet 46:925–933

Germino GG, Weinstat-Saslow D, Himmelbauer H, Gillespie GAJ, Somlo S, Wirth B, Barton N, et al (1992) The gene for autosomal dominant polycystic kidney disease lies in a 750-kb CpG-rich region. Genomics 13:144–151

Green ED, Olson MV (1990) Systematic screening of yeast artificial-chromosome libraries by use of the polymerase chain reaction. Proc Natl Acad Sci USA 87:1213–1217

Harris PC, Barton NJ, Higgs DR, Reeders ST, Wilkie AOM (1990) A long-range restriction map between the α-globin complex and a marker closely linked to the polycystic kidney disease (PKD1) locus. Genomics 7:195–206

Harris PC, Thomas S, MacCarthy AB, Stallings RL, Breuning MH, Jenne DE, Fink TM, et al. A large duplicated area in the polycystic kidney disease 1 (PKD1) region of chromosome 16 is prone to rearrangement (submitted)

Harris PC, Thomas S, Ratcliffe PJ, Breuning MH, Coto E, Lopez-Larrea C (1991) Rapid genetic analysis of families with polycystic kidney disease by means of a microsatellite marker. Lancet 338:1484–1487

Hill WG (1975) Linkage disequilibrium among multiple neutral alleles produced by mutation in finite population. Theor Popul Biol 8:117–126

Hirst MC, Rack K, Nakahori Y, Roche A, Bell MV, Flynn G, Christadoulou Z, et al (1991) A YAC contig across the fragile X site defines the region of fragility. Nucleic Acids Res 19:3283–3288

Huntington's Disease Collaborative Research Group (1993) A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. Cell 72:971–983

Jeffreys AJ, Royle NJ, Wilson V, Wong Z (1988) Spontaneous mutation rates to new length alleles at tandem-repetitive hypervariable loci in human DNA. Nature 332:278–281

Kerem B-S, Rommens JM, Buchanan JA, Markiewicz D, Cox TK, Chakravarti A, Buchwals M, et al (1989) Identification of the cystic fibrosis gene: genetic analysis. Science 245:1073–1079

Kimberling WJ, Fain PR, Kenyon JB, Goldgar D, Sujansky E, Gabow PA (1988) Linkage heterogeneity of autosomal dominant polycystic kidney disease. N Engl J Med 319:913–918

Kimberling WJ, Kumar S, Gabow PA, Kenyon JB, Connolly CJ, Somlo S (1993) Autosomal dominant polycystic kidney disease: localization of the second gene to chromosome 4q13-q23. Genomics 18:467–472

Murray JC, Shiang R, Carlock LR, Smith M, Buetow KH (1987) Rapid RFLP screening procedure identifies new polymorphisms at albumin and alcohol dehydrogenase loci. Hum Genet 76:274–277

Nelson DL, Ledbetter SA, Corbo L, Victoria MF, Ramirez-Solis R, Webster TD, Ledbetter DH, et al (1989) Alu polymerase chain reaction: a method for rapid isolation of human-specific sequences from complex DNA sources. Proc Natl Acad Sci USA 86:6686–6690

Ott J (ed) (1985) Analysis of human genetic linkage. Johns Hopkins University Press, Baltimore

Peral B, San Millan JL, Hernandez C, Valero A, Lathrop GM, Beckmann JS, Moreno F (1993) Estimating locus heterogeneity in autosomal dominant polycystic kidney disease (ADPKD) in the Spanish population. J Med Genet 30:910–913

Peters DJM, Sandkuijl LA (1992) Genetic heterogeneity of polycystic kidney disease in Europe. In: Breuning MH, Devoto M, Romeo G (eds) Contributions to nephrology: polycystic kidney disease. Karger, Basel, pp 128–139

Peters DJM, Spruit L, Saris JJ, Ravine D, Sandkuijl LA, Fossdal R, Boersma J, et al (1993) Chromosome 4 localization of a second gene for autosomal dominant polycystic kidney disease. Nature Genet 5:359–362

Pound SE, Carothers AD, Pignatelli PM, Macnicol AM, Watson ML, Wright AF (1992) Evidence for linkage disequilibrium between *D16S94* and the adult onset polycystic kidney disease (PKD1) gene. J Med Genet 29:247–248

Reeders ST, Breuning MH, Davies KE, Nicholls RD, Jarman AP, Higgs DR, Pearson PL, et al (1985) A highly polymorphic DNA marker linked to adult polycystic kidney disease on chromosome 16. Nature 317:542–544

Romeo G, Costa G, Catizone L, Germino GG, Weatherall DJ, Devoto M, Roncuzzi L, et al (1988) A second genetic locus for autosomal dominant polycystic kidney disease. Lancet 2:8–10

Schlötterer C, Tautz D (1992) Slippage synthesis of simple sequence DNA. Nucleic Acids Res 20:211–215

Snarey A, Thomas S, Schneider MC, Pound SE, Barton N, Wright AF, Harris PC, et al. Linkage disequilibrium in the region of the autosomal dominant polycystic kidney disease gene (PKD1) (submitted)

Somlo S, Wirth B, Germino GG, Weinstat-Saslow D, Gillespie GAJ, Himmelbauer H, Steevens L, et al (1992) Fine genetic localization of the gene for autosomal dominant polycystic kidney disease (PKD1) with respect to physically mapped markers. Genomics 13:152–158

Stallings RL, Torney DC, Hildebrand CE, Longmire JL, Deaven LL, Jett JH, Doggett NA, et al (1990) Physical mapping of human chromosomes by repetitive sequence fingerprinting. Proc Natl Acad Sci USA 87:6218–6222

Thompson AD, Shen Y, Holman K, Sutherland GR, Callen DF, Richards RI (1992) Isolation and characterization of $(AC)_n$ microsatellite genetic markers from human chromosome 16. Genomics 13:402–408

Weber JL (1990) Informativeness of human (dC-dA)n (dG-dT)n polymorphisms. Genomics 8:524–530

Weir BS, Cockerham CC (1978) Testing hypotheses about linkage disequilibrium with multiple alleles. Genetics 88:633–642