Linkage and linkage disequilibrium in chromosome band 1p36 in American Chaldeans with inflammatory bowel disease

Judy H. Cho+, Dan L. Nicolaë2, Richard Ramos, Carter T. Fields, Karen Rabenau, Sarah Corradino, Steven R. Brant3, Rafael Espinosa1, Michelle LeBeau1, Stephen B. Hanauer, Jason Bodzin4 and Denise K. Bonen

The Martin Boyer Genetics Research Laboratories, Gastroenterology and 1Hematology/Oncology sections, Department of Medicine and 2Department of Statistics, The University of Chicago Hospitals, 5841 South Maryland Avenue, MC6084, Chicago, IL 60637, USA, 3The Harvey M. and Lynn P. Meyerhoff Inflammatory Bowel Disease Center, Department of Medicine, The Johns Hopkins University School of Medicine, 918 Ross Research Building, 720 Rutland Avenue, Baltimore, MD 21205, USA, 4DMC/Sinai Hospitals, Detroit, MI, USA

Received 4 February 2000; Revised and Accepted 22 March 2000

The idiopathic inflammatory bowel diseases (IBDs), consisting of Crohn’s disease and ulcerative colitis, are complex genetic disorders involving chronic inflammation of the intestines. Multiple genetic loci have been implicated through genome-wide searches, but refinement of localization sufficient to undertake positional cloning efforts has been problematic. This difficulty can be obviated through identification of ancestrally shared regions in genetic isolates, such as the Chaldean population, a Roman Catholic group from Iraq. We analyzed four multiply affected American Chaldean families with inflammatory bowel disease not known to be related. We observed evidence for linkage and linkage disequilibrium in precisely the same region of chromosome band 1p36 reported previously in an outbred population. Maximal evidence for linkage was observed near D1S1597 by multipoint analysis (MLOD = 3.01, \( P = 6.1 \times 10^{-5} \)). A shared haplotype (D1S507 to D1S1628) was observed over 27 cM between two families. There was homozygous sharing of a 5 cM portion of that haplotype in one family and over a <1 cM region in the second family. Homozygous sharing of this haplotype near D1S2697 and D1S3669 was observed in one individual in a third multiply affected family, with heterozygous sharing in a fourth family. Linkage in outbred families as well as in this genetic isolate indicates that a pathophysiologically crucial IBD susceptibility gene is located in 1p36. These findings provide a unique opportunity to refine the localization and identify a major susceptibility gene for a complex genetic disorder.

INTRODUCTION

The idiopathic inflammatory bowel diseases (IBD), Crohn’s disease (CD) and ulcerative colitis (UC), are chronic inflammatory disorders of unknown cause and have a combined prevalence of 200–300/100 000, representing a major cause of morbidity in young adults, with a peak incidence between 15 and 30 years of age (1–3). Ulcerative colitis involves a continuous inflammation of variable extent limited to the rectal and colonic mucosal layers of the bowel wall. In contrast, CD may involve any part of the gastrointestinal tract (most frequently the terminal ileum and colon) and the bowel inflammation is transmural and discontinuous (4). The differentiation between UC and CD is based on non-specific clinical and histologic patterns; in ~10% of cases confined to the rectum and colon, definitive classification as CD or UC cannot be made and are designated as ‘indeterminate colitis’ (4). Medical therapies for the two disorders are largely overlapping and consist of anti-inflammatory agents such as corticosteroids, aminosalicylates and immunomodulatory agents (5).

There is strong evidence from twin studies (6), familial risk studies (7,8) and complex segregation analyses (9,10) that CD and UC are related, complex genetic disorders. A total of 5–10% of patients have another case of IBD in the family (11). Twin studies, especially for CD, demonstrate higher concordance between monozygotic twins compared with dizygotic twins (6). The population relative risk in first-degree relatives of patients demonstrates a 14-fold increased risk (11). Approximately 75–80% of multiply affected families are concordant for the same disease, with 20% of multiply affected families being mixed, having one member with CD, and another, UC (11). Segregation analyses implicate a major dominant or additive gene for UC with low penetrance, with a major recessive gene with high penetrance for CD (9,10).

Five genome-wide searches for disease susceptibility genes in IBD have been reported (12–16). For many of the implicated loci, evidence for linkage was observed commonly in CD and

+To whom correspondence should be addressed. Tel: +1 773 702 5375; Fax: +1 773 702 2281; Email: jcho@medicine.bsd.uchicago.edu
UC patients, suggesting the presence of shared susceptibility genes. Significant linkage has been observed among pure CD families at chromosome 16, IBDD (12), a linkage finding which has subsequently been confirmed in both pure CD (17–22) and pure UC (23) families, but not among mixed families. Taken together, these studies indicate that multiple genes contribute to susceptibility to IBD, with at least some of those genes being common to CD and UC.

Our genome-wide screen in 174 American families comprising 297 affected relative pairs (14) identified 17 regions demonstrating nominal [multipoint LOD curve (MLOD) >1.0, \( P < 0.016 \)] evidence for linkage. The greatest evidence for linkage in all families (MLOD = 2.65, \( P = 2 \times 10^{-5} \)) was observed near D1S552. Positive LOD scores were observed in both CD and UC families, but not in mixed families.

For complex disorders, evidence for linkage is typically observed over relatively broad genetic distances (24). The feasibility of refining gene localization through linkage (24) or linkage disequilibrium approaches (25) for complex disorders in large outbred populations is largely unestablished at present. The central difficulty in refining localization for multigenic disorders is that specific recombination events cannot be distinguished reliably from locus heterogeneity. These difficulties can potentially be obviated through studies in genetic isolates (26–28), such as the Chaldean population, where greater genetic homogeneity is predicted. Ancestral sharing around disease loci can be utilized both to provide confirmatory evidence for linkage and to refine localization, to guide subsequent positional cloning efforts.

The Chaldeans consist of Middle Eastern Roman Catholics whose primary language is Aramaic (29). The Chaldean Church was founded in 1445 AD and consists of those parishes officially reconciled with Rome. While not representing a geographic isolate, the Chaldeans, because of religious, cultural and language differences, have remained genetically isolated throughout their history. The present day Chaldean population numbers \( \sim 500,000 \), with \( \sim 80,000 \) Chaldeans in the USA, the majority of whom reside in the metropolitan Detroit area, having immigrated from the early 1960s to the present. It has been recognized anecdotally that the prevalence of IBD among American Chaldeans is high following immigration to the USA (J. Bodzin, personal communication). These observations are concordant with prior studies reporting increased incidence and prevalence of IBD in Asian migrants and their offspring immigrating to more industrialized cities compared with rates in Asia (31,32).

For these reasons, we tested multiply affected Chaldean families for the presence of linkage and linkage disequilibrium in the region of chromosome band 1p36, the same region demonstrating maximal linkage in a genome-wide screen in an outbred population (14). We observed the presence of linkage and linkage disequilibrium in this region among American Chaldean families with IBD, which provides further confirmatory evidence for a susceptibility gene here, as well as critically refines localization to a \( < 1 \) cM region near D1S2697 and D1S3669.

### RESULTS

**Phenotypic characterization**

Family history did not reveal any parental consanguinity or known relationships between any of the tested families after completely defining the ancestral generation preceding the designated parents in Figure 1. With the exception of individual 15 of pedigree A and individual 3 of pedigree D, all affected Chaldeans studied have UC. This contrasts with our outbred population of multiplex families with IBD (14) where the diagnosis of CD predominated. Individual 15 of pedigree A with CD had primarily terminal ileal involvement with development of a fistulous tract, with secondary involvement of the ascending colon.

**The evidence for linkage in the Chaldean families overlaps the region of linkage observed in outbred American families**

Figure 1 demonstrates the pedigree structures of three of the Chaldean families studied. The fourth family is an affected aunt (CD)/niece (UC) pair. Table 1 demonstrates the two-point, non-parametric LOD scores obtained using the exponential model (32,33). The maximal two-point LOD score was observed at D1S199 (LOD 3.74, \( P = 6.1 \times 10^{-5} \)), which maps (\( \theta = 0 \)) closely to D1S552, the screening marker demonstrating the maximal evidence for linkage in the outbred population. The absence of significant evidence for two-point linkage at D1S552 (Table 1) stems from its lack of informativeness,

<table>
<thead>
<tr>
<th>Marker</th>
<th>Distance</th>
<th>LOD score</th>
<th>( P )-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1S468</td>
<td>0</td>
<td>0.35</td>
<td>0.10</td>
</tr>
<tr>
<td>D1S1612</td>
<td>12</td>
<td>1.13</td>
<td>0.012</td>
</tr>
<tr>
<td>D1S1597</td>
<td>26</td>
<td>2.35</td>
<td>0.0011</td>
</tr>
<tr>
<td>D1S507</td>
<td>30</td>
<td>1.50</td>
<td>0.0050</td>
</tr>
<tr>
<td>D1S1193</td>
<td>31</td>
<td>2.11</td>
<td>0.0013</td>
</tr>
<tr>
<td>D1S436</td>
<td>33</td>
<td>1.23</td>
<td>0.011</td>
</tr>
<tr>
<td>D1S2697</td>
<td>33</td>
<td>1.26</td>
<td>0.0093</td>
</tr>
<tr>
<td>(CA)20B21</td>
<td>33</td>
<td>0.13</td>
<td>0.21</td>
</tr>
<tr>
<td>D1S3669</td>
<td>33</td>
<td>0.074</td>
<td>0.25</td>
</tr>
<tr>
<td>D1S1592</td>
<td>35</td>
<td>0.62</td>
<td>0.049</td>
</tr>
<tr>
<td>D1S2826</td>
<td>38</td>
<td>1.18</td>
<td>0.012</td>
</tr>
<tr>
<td>D1S2644</td>
<td>40</td>
<td>0.86</td>
<td>0.022</td>
</tr>
<tr>
<td>D1S552</td>
<td>41</td>
<td>0.17</td>
<td>0.79</td>
</tr>
<tr>
<td>D1S199</td>
<td>41</td>
<td>3.74</td>
<td>( 6.10 \times 10^{-5} )</td>
</tr>
<tr>
<td>D1S2647</td>
<td>41</td>
<td>0.79</td>
<td>0.027</td>
</tr>
<tr>
<td>D1S2864</td>
<td>46</td>
<td>0.0095</td>
<td>0.39</td>
</tr>
<tr>
<td>D1S2734</td>
<td>49</td>
<td>0.47</td>
<td>0.074</td>
</tr>
<tr>
<td>D1S1676</td>
<td>51</td>
<td>0.55</td>
<td>0.054</td>
</tr>
<tr>
<td>D1S2749</td>
<td>51</td>
<td>0.61</td>
<td>0.049</td>
</tr>
<tr>
<td>D1S1622</td>
<td>53</td>
<td>0.065</td>
<td>0.29</td>
</tr>
<tr>
<td>D1S247</td>
<td>54</td>
<td>0.33</td>
<td>0.11</td>
</tr>
<tr>
<td>D1S1161</td>
<td>57</td>
<td>0.41</td>
<td>0.081</td>
</tr>
<tr>
<td>D1S201</td>
<td>58</td>
<td>0.28</td>
<td>0.12</td>
</tr>
<tr>
<td>D1S1628</td>
<td>59</td>
<td>0.43</td>
<td>0.081</td>
</tr>
<tr>
<td>D1S1190</td>
<td>60</td>
<td>0.71</td>
<td>0.033</td>
</tr>
</tbody>
</table>
especially within pedigree A, which accounts for a significant portion of the observed evidence for linkage in this region (Fig. 1). Both parents (A-2 and A-3) are homozygous for allele 3 at D1S552, and therefore no linkage is observed here for this family by two-point analysis (LOD = 0), despite all siblings clearly having identity-by-descent of two. For the marker D1S199, all of the siblings in the three represented families have identity-by-state of two; however, the multipoint analysis reveals that individuals 1 in pedigree B and 1 in pedigree C have identity-by-descent of only one near D1S199 with the other affected siblings in their respective families (Fig. 1). Therefore, the two-point LOD score at D1S199 is overestimated, with the MLOD score providing a more accurate estimate of evidence for linkage. Figure 2 demonstrates the MLOD (32,33). The maximum MLOD score of 3.01 ($P = 6.1 \times 10^{-5}$) was observed near D1S1597. The two-point LOD score for D1S1597 is 2.35. Note that the $P$-values for the maximal two-point and multipoint LOD scores are equal even though the LOD scores are different. This is a consequence of the discrete distribution of the LOD score (under complete information) which is used in assessing the significance of the observed LOD scores. The highest value the complete information LOD score can take for this data set is 4.21 and, under no linkage assumptions, it occurs with probability $1/2^{14}$; the

![Figure 1. Haplotype analysis, three Chaldean families. The common, at-risk haplotype between pedigrees is boxed. The haplotype shared among affected individuals in pedigree C is italicized. D1S552, D1S2647 and D1S199 have a genetic recombination rate of zero and the listed order is non-definitive. Dashed lines represent missing genotypes.](image-url)
next largest possible value when there is complete identity-by-descent information is 2.98. Under a parametric model (with penetrances of 0.95, 0.05 and 0.001 for homozygous risk allele, heterozygous and homozygous wild-type genotypes, respectively) the LOD score is 2.97.

**Haplotype analysis demonstrates an extended shared haplotype between two Chaldean families**

Haplotype analysis (Fig. 1) in pedigrees A and B demonstrates an extended shared haplotype of 27 cM between D1S1193 and D1S1628 in seven of eight affected individuals. The centromeric shared boundary of the remaining affected individual (pedigree A, individual 5) is D1S2864, representing a shared haplotype of 15 cM. Furthermore, in pedigree A, homozygous sharing of the common haplotype from D1S507 to D1S2826 (5 cM) is observed among the four affected siblings.

In the region bounded by D1S1193 and D1S1592, in pedigree B, two of the three affected sisters are homozygous for the markers D1S436, D1S2697, D1S3669 and D1S2826, with heterozygous sharing observed at D1S1592. The precise localization of D1S1592 (demonstrating heterozygous sharing) relative to the flanking markers (demonstrating homozygous sharing) can provide critical refinement of localization information if one assumes a recessive model. To more rigorously establish marker order in this region, radiation hybrid (RH) mapping and fluorescence in situ hybridization (FISH) analysis was performed using P1 artificial chromosome (PAC) clones containing each of these markers.

**RH analysis suggests that D1S1592 maps centromeric to D1S3669 and p-telomeric to D1S552**

RH mapping using the Stanford G3 panel was performed on D1S436, D1S2697, D1S3669, D1S2826 and D1S552. The retention frequency for D1S3669 was low, and significant placement order could not be obtained. D1S3669 is present in a sequenced (AC004824) PAC clone, 20B21, which, by electronic PCR, contains the sequence-tagged site, SHGC-1466. RH mapping analysis established the most likely order (p-telomeric to centromeric): D1S436–D1S2697–SHGC1466 (and, by extension, D1S3669–D1S1592–D1S2826–D1S552. With the exception of the D1S436–D1S2697 ordering, the markers mapped to separate bins, providing >1000:1 odds support. To provide complementary support for these findings, FISH analysis was performed using PAC clones in the region.

**FISH analysis establishes that D1S1592 is centromeric to D1S3669 and represents the centromeric boundary for the disease gene**

The radiation hybrid map information was used to choose telomeric (D1S2697) and centromeric anchors (D1S552) for FISH analysis. All five PAC clones mapped to 1p36.1. Interphase FISH analysis established the following order: (telomeric)–D1S2697 (164M8)–D1S3669 (20B21)–D1S1592 (207B10)–D1S2826 (304)–D1S552 (93P18)–(centromeric).

Within the homozygous 5 cM region between D1S507 and D1S1592 defined by pedigree A, two of three affected sisters in pedigree B demonstrate homozygosity for the same alleles from D1S436 to D1S3669. Taken together, haplotype analysis of pedigrees A and B localizes the IBD gene to a <1 cM region containing D1S436 and D1S3669.

**Two additional Chaldean pedigrees demonstrate sharing in the D1S2697 and D1S3669 region**

In this region, pedigree C demonstrates excess identity-by-descent sharing, with a maximum non-parametric linkage score of 1.97 (32). To further test the region near D1S3669, an additional short tandem repeat marker, (CA)20B21, was developed and typed from the PAC clone containing D1S3669. Haplotype analysis of individual 2 demonstrates a most likely haplotype of 1-6-10-3-9 [D1S2697–(CA)20B21–D1S3669–D1S1592–D1S2826], which is common to the at-risk haplotype defined by pedigrees A and B. A fourth Chaldean pedigree, (pedigree D) representing an aunt/niece affected relative pair was typed in this region with identity-by-state of one in this region inferred (Fig. 3). Two additional single nucleotide polymorphisms (SNPs), SNP3 and SNP1, contained within the ephrin A2 gene, were identified between D1S2697 and D1S3669 and sequenced. The allele frequencies for the minor allele (allele 2) were 0.23 and 0.30 in a control group for SNP3 and SNP1, respectively. Of note is that individual 1 of pedigree D (Fig. 3) is homozygous for the at-risk haplotype in the same region near D1S2697 and D1S3669. It is significant that the at-risk haplotype in this region includes the rarer alleles (i.e. allele 2) at SNPs 1 and 3 and thus provides strong evidence of ancestral sharing in all Chaldean families with IBD. Specifically, of the four separate Chaldean families multiply affected with IBD, three have at least one individual homozygous for the presently defined at-risk haplotype with a fourth family (pedigree C) having one individual heterozygous for the at-risk haplotype. Taken together, these data implicate the region near D1S2697 and D1S3669 as the most likely region containing a disease gene contributing to susceptibility to IBD. D1S3669 is 7cM p-telomeric (Marshfield screening set 8 map) to D1S552, the marker demonstrating the maximal evidence for linkage in our prior genome-wide screen (14). Our maximal MLOD score previously was 2.65 at D1S552, with an MLOD = 1.70 at D1S3669. Marker informativeness, and marker density will...
greatly affect the precise location of the peak of the MLOD curve and probably accounts for the 7 cM shift observed between the two studies. The present localization is within the one-LOD confidence interval reported previously (14), and represents confirmatory evidence for the linkage reported previously on chromosome 1p.

Of note is that two individuals (pedigree A, individual 15 and pedigree B, individual 1) were homozygous at D1S2697 for allele 1. Typing of additional SNPs immediately flanking D1S2697 failed to demonstrate homozygosity in these two individuals (data not shown).

The chromosome 1p36 region represents a unique region of linkage and linkage disequilibrium in American Chaldeans with IBD

There is no evidence of homozygous sharing demonstrated by screening set markers elsewhere in the genome. There is only one region (chromosome 9) with a LOD score >1.5 and most of the affected individuals are heterozygous in the region of linkage.

DISCUSSION

We have previously reported the results of a genome-wide screen in 174 American families with IBD and demonstrated the maximal region of linkage to be at 1p36, near D1S552 (14). The present report describes the presence of both linkage and linkage disequilibrium in American Chaldean families with IBD over an extended haplotype overlapping the previously defined linkage region. Taken together, these results indicate that a pathophysiologically crucial disease susceptibility gene for IBD is located at 1p36 and refines localization to a <1 cM region near D1S2697 and D1S3669.

The multipoint LOD score (32,33) and significance value (MLOD = 3.01, \( P = 6.1 \times 10^{-5} \)) do not reflect the presence of the extended (27 cM) shared haplotype between pedigrees A and B, as well as the homozygous sharing (5 cM in pedigree A, 2 cM in pedigree B) observed. Mis-specification of allele frequencies can potentially falsely increase the evidence for linkage (34). However, in these families, with the dense map used and complete family structures obtained, mis-specification of allele frequencies is unlikely to have falsely increased the evidence for multipoint linkage. Therefore, the present findings provide complementary, confirmatory evidence for an IBD susceptibility gene in this region (35).

A major obstacle to gene identification in complex genetic disorders is that precise localization through genetic linkage is problematic because of locus heterogeneity (36). Positive linkage results are typically observed over very broad regions (40–50 cM) for genes of major effect (12–16). For multigenic disorders, recombination events cannot be definitively distinguished from locus heterogeneity. The presence of more than one susceptibility gene on a chromosome will further complicate refinement efforts (14). Estimates of the number of informative affected relative pairs required to definitively refine localization through linkage approaches may be prohibitively large (24) for loci conferring modest risk (e.g. 2-fold increased relative risk), which will probably be the case for most susceptibility loci for complex disorders (37).

Studies in genetic isolates, especially with multiplex families, can potentially obviate this problem by identifying regions of linkage disequilibrium, as the present study demonstrates (26–28). The value of using isolated populations to search for linkage disequilibrium for complex disorders has been largely unproven (28). It is well established that increased linkage disequilibrium is present around rare disease mutations, but for more common diseases (25), it is hypothesized that many of the underlying susceptibility alleles are common gene variants associated with little selective pressure (37). Whether linkage disequilibrium is increased around common disease variants in isolated populations has, for the most part, been undetermined. The demographic history of most populations, including the Chaldeans, is not known in sufficient detail to make precise theoretical calculations (28). Therefore, the empiric results from the present study provide a unique opportunity to refine localization for a complex genetic disorder.
The homozygous sharing observed in most of the UC patients from pedigrees A, B and D suggests that, in these families, the disease gene is operating in a recessive, or an additive fashion. This is not unexpected as linkage approaches possess greater power to identify recessive loci. This contrasts with complex segregation analyses in outbred populations suggesting an autosomal dominant or additive major gene with low penetrance in UC (7).

In the outbred population, positive evidence for linkage was observed in pure CD (MLOD = 2.17, P = 7.9 × 10⁻⁵) and pure UC (MLOD = 1.06, P = 0.09) families, with little evidence observed in mixed families (one member with CD, another with UC, MLOD = 0.0074) (14). This is analogous to the observations with respect to IBD1 on chromosome 16, where positive linkage has been observed in pure CD (12,17–22) and UC families (23), but not mixed families. While the highest recurrent risk to first degree relatives is for the same disease (7), the cross-disease relative risk data and results from genome-wide searches (13–16) in IBD suggest the presence of at least some shared susceptibility genes.

In pedigree A, individual 15 with CD shares at least one haplotype (and thus specific disease gene polymorphism) with the remaining affected individuals in the family, all with UC. One possibility is that individual 15 (pedigree A) will have a different polymorphism on the unshared chromosome, directly accounting for phenotypic differences between CD and UC. This could partially account for the relative absence of evidence for linkage in mixed families compared with pure CD or UC families in the outbred population (14) if different polymorphisms in the same gene are required for one family member to have CD, and another, UC. Alternatively, this individual may eventually be discovered to share two copies of the same pathogenic polymorphism here, with the phenotypic difference resulting from separate genetic (e.g. IBD1 on chromosome 16) or environmental factors. For example, it has been postulated that tobacco use (associated with CD, protective against UC) can partially account for the phenotypic differences in mixed families with one member who have UC, and the other, CD (38). Of note is that individual 15 (pedigree A) does not have a history of tobacco use. Further studies may provide significant insight into the molecular basis of the phenotypic differences between CD and UC.

Within the region near D1S2697 and D1S3669 are a number of known genes and expressed sequence tags which, by expression profiles and established function, may provide reasonable candidates for IBD-associated disease genes. Ephrin A2 is a tyrosine protein kinase receptor which has been implicated in contributing to intestinal epithelial migration and barrier function (39). Other known genes in the region include RNU1A (RNA U1 small nuclear), microfibrillar associated protein, isofrom 2 and eukaryotic translation initiation factor, 4C.

It is unlikely that the data on these families can give a better resolution and we are confirming the positional cloning effort using this information in the outbred data set initially demonstrating linkage in this region (14). Also, further mapping studies with additional singly affected and more distantly related multiplex Chaldean families are currently underway. However, even within this relatively homogeneous population, allelic and locus heterogeneity is predicted. In this regard, homozygosity mapping in individuals 1 (pedigree B) and 2 (pedigree C) will be of greatest value in further localization.

Inclusion of individuals 15 (pedigree A) and 3 (pedigree D) is less reliable because of the diagnosis of CD. Whether the IBD susceptibility gene at 1p36 will be common to sporadic cases of IBD in the Chaldean population, as well as larger, outbred populations has yet to be determined. However, the simultaneous presence of maximal evidence for linkage in both outbred, multiplex American families and Chaldean Americans, suggests that a crucial, common IBD susceptibility gene will ultimately be found in both populations.

MATERIALS AND METHODS

Family ascertainment

The Chaldean families with IBD were identified (J.B.) and informed consent for a molecular genetic study was obtained. Extended family pedigrees were obtained by phone. Additional unrelated Chaldeans were recruited to obtain estimates of allele frequencies. The study protocol was approved by the institutional review board of the University of Chicago. Confirmation of diagnoses for idiopathic, chronic inflammatory bowel disease, i.e. CD, UC or indeterminate colitis, was obtained from primary review of endoscopic, radiologic and pathologic data (4).

Genotyping

Peripheral blood samples were taken from family members after informed consent was obtained. Genomic DNA was extracted and PCR-based genotyping performed using markers in the region of IBD1 using kinase-labeled oligonucleotide primers obtained from Research Genetics (Huntsville, AL). Weber 8.0 screening set markers were genotyped by the Mammalian Genotyping Service (National Heart, Lung, and Blood Institute) at the Marshfield Center for Medical Genetics. Additional short tandem repeat markers were typed radioactively using kinase-labeled markers with at least a 2–3 cM map density over a 45 cM region spanning the region of linkage on chromosome 1 (4) in order to identify shared haplotypes. High-throughput genomic sequence (AC004824) from the PAC clone 20B21 obtained from Genome Systems (St Louis, MO) (which contains D1S3669) was used to identify a polymorphic CA repeat. Primers were synthesized by Life Technologies (Grand Island, NY) (forward primer: 5'-gaggttgtgcagacaaa-3'; reverse primer: 5'-gtgtgctcctcgctgc-3') to genotype this polymorphic repeat, designated (CA)20B21. PCR amplification was performed using either a three-step or two-step (94°C denaturation for 30 s, 60°C anneal for 30 s, 72°C for 30 s) procedure. PCR products were separated on a 5% polyacrylamide denaturing gel and results analyzed by autoradiography. Alleles were genotyped by two independent scorings and discrepancies resolved.

Data analysis

Linkage analysis was performed using Allegro (39), an extension of Genehunter (32). Genehunter utilizes an inheritance vector to characterize meiosis events and calculates, at each location, a probability distribution on the space of inheritance vectors. Identity-by-descent sharing test statistics can be estimated from this probability distribution. The Spairs scoring function was used, and the significance of the corresponding
test statistic was assessed using the exponential model (33). Allele frequencies were estimated from 20 Chaldean founders. The four families were assigned equal weight on the linkage tests (33). \( P \)-values were calculated from the distribution of the LOD score in the complete information case (perfect-data approximations). Haplotype reconstruction was performed using Allegro (40).

**Radiation hybrid mapping**

Primers for D1S436, D1S2697, D1S2826, D1S3669, D1S1592, D1S552 and SHGC-1466 were used to screen the Stanford G3 radiation hybrid (RH) panel from Research Genetics by PCR. PCR products were analyzed by agarose gel electrophoresis. Results were analyzed using SAMini available from the Stanford Genome Center (http://www-shgc.stanford.edu).

**FISH**

PAC clones were obtained by PCR screening of the PAC pools obtained from Genome Systems for D1S2697 (164M8), D1S2826 (304), D1S3669 (20B21), D1S1592 (207B10) and D1S552 (93P18), and analyzed by FISH to determine order. Human metaphase and interphase cells were prepared from phytohemagglutinin-stimulated peripheral blood lymphocytes. Multi-color FISH was performed as described previously (41). Labeled PAC probes were prepared by nick-translation using Bio-11-dUTP from Roche Molecular Biochemicals (Indianapolis, IN). Biotin-labeled probes were detected with FITC-conjugated avidin, and the digoxigenin-labeled probes were detected with rhodamine-conjugated sheep anti-digoxigenin antibodies from Vector Laboratories (Burlingame, CA). Chromosomes were identified by staining with 4,6-diamidino-2-phenylindole-(DAPI) dihydrochloride. The slides were examined using an FITC/rhodamine double-band pass filter obtained from Chroma Technology (Brattleboro, VT). Pairwise combinations of two differentially labeled probes were hybridized to metaphase cells prepared from mitogen-stimulated lymphocytes. To determine the order of probes in interphase nuclei, combinations of three probes, two labeled with biotin and one with digoxigenin, or vice versa, were co-hybridized. The signal pattern in 20 metaphase cells or interphase nuclei was scored for each combination of probes. Probe order was designated when ≥75% of signals showed the same order.

**Development of additional SNPs in the region between D1S2697 and D1S3669**

Additional genetic markers in this region were developed through establishment of a P1 contig surrounding D1S2697. SNP1 and SNP3 contained within the ephrin A2 gene were mapped to this contig using direct sequencing and PCR. Map order was confirmed using the TNG RH from Research Genetics and the Sanger Center contig maps (http://www.sanger.ac.uk/HGP/). Primers for SNP3 were (forward) 5′-CTTCGCAGGCCTCAGCTCCTTC-3′ and (reverse) 5′-GGGGAGGAAAGAACGTTAAGATAA-3′ with the variant sequence cgcat-C/T-gccta (the more common C allele is designated allele 1 with the T allele being rarer). Sequences for the adjacent SNPs were in complete linkage disequilibrium with the more common haplotype (G-G) designated allele 1, and the rarer haplotype (A-A), allele 2.

**Acknowledgements**

We gratefully acknowledge the invaluable contributions of Joseph B. Kirsner, Graeme I. Bell, Nancy J. Cox, Francine Jackson, James Weber and Eugene B. Chang. This project was funded partially by the Gastrointestinal Research Foundation, the Reva and David Logan Center for Gastrointestinal Research (J.H.C., S.B.H.), the Mazza Foundation, the Circle of Service Foundation, P30 DK 42086 (Digestive Disease Research Center Core) the National Institutes of Health/ National Center for Research Resources GCRC RR00055, the Mammalian Genotyping Service (National Heart, Lung, and Blood Institute), the Crohn’s and Colitis Foundation of America (Career Development Award, J.H.C.), and the NIDDK Clinical Scientist Award (K08 DK 02560-01 ZDK1-SRC-C) NIDDK R01DK5731 (J.H.C.).

**REFERENCES**


