Published in final edited form as:

Hum Genet. 2013 April ; 132(4): 385–396. doi:10.1007/s00439-012-1252-5.

Linkage analysis of a large African family segregating stuttering suggests polygenic inheritance

M. Hashim Raza^{1,*}, E. Michael Gertz^{2,*}, Jennifer Mundorff¹, Joseph Lukong¹, Judith Kuster³, Alejandro A. Schäffer², and Dennis Drayna^{1,#}

¹National Institute on Deafness and Other Communication Disorders, National Institutes of Health, Bethesda, MD USA

²National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD USA

³Department of Speech, Hearing, and Rehabilitative Services, Minnesota State University, Mankato, MN

Abstract

We describe a pedigree of 71 individuals from the Republic of Cameroon in which at least 33 individuals have a clinical diagnosis of stuttering. The high concentration of stuttering individuals suggests that the pedigree either contains a single highly penetrant gene variant or that assortative mating led to multiple stuttering-associated variants being transmitted in different parts of the pedigree. No single locus displayed significant linkage to stuttering in initial genome-wide scans with microsatellite and SNP markers. By dividing the pedigree into five sub-pedigrees, we found evidence for linkage to previously reported loci on 3q and 15q, and to novel loci on 2p, 3p, 14q, and a different region of 15q. Using the two-locus mode of Superlink, we showed that combining the recessive locus on 2p and a single-locus additive representation of the 15q loci is sufficient to achieve a two-locus score over 6 on the entire pedigree. For this 2p+15q analysis, we show LOD scores ranging from 4.69 to 6.57, and the scores are sensitive to which marker is chosen for 15q. Our findings provide strong evidence for linkage at several loci.

Introduction

Stuttering is a common disorder affecting the flow of speech, characterized by involuntary repetitions or prolongations of words or syllables, and by interruptions in speech known as blocks. Stuttering typically appears in children at the age of three to five years, where the incidence rate is about five percent. More than 75% of these children recover, either spontaneously or with speech therapy, leading to a prevalence of stuttering of about one percent in the adult population (Craig et al. 2002; Felsenfeld 2002). Twin and adoption studies have suggested high heritability in this disorder (Andrews et al. 1991; Bloodstein 1961; Dworzynski et al. 2007; Fagnani et al. 2011; Felsenfeld et al. 2000; Felsenfeld and Plomin 1997; Godai et al. 1976; Howie 1981; Ooki 2005; van Beijsterveldt et al. 2010). However Mendelian segregation typically does not occur, and a number of traditional linkage studies have produced limited success (Shugart et al. 2004; Suresh et al. 2006; Wittke-Thompson et al. 2007). This is not surprising in view of the characteristics of this disorder, such as the high recovery rate and absence of clear segregation, and supports the model of stuttering as a complex genetic disorder.

[#]To whom correspondence should be addressed at: 5 Research Court, Room 2B-46, Rockville, MD 20850 USA, 301-402-4930, drayna@nidcd.nih.gov.

^{*}These authors contributed equally to the work presented here

To address these impediments to genetic analysis, a number of studies have been done in consanguineous populations, particularly in Pakistan. Studies in this population first revealed a significant linkage to markers on chromosome 12 (Riaz et al. 2005). This linkage was used to subsequently identify mutations in the *GNPTAB* gene, initially in the Pakistani population and subsequently in additional populations (Kang et al. 2010). More recently, strong evidence of linkage to stuttering was found in consanguineous Pakistani families on chromosome 3q (Raza et al. 2010) and chromosome 16q (Raza et al. 2012).

Whole-exome and whole-genome sequencing approaches are being used to identify the mutant genes underlying Mendelian disorders (Ng et al. 2010a; Ng et al. 2010b; Rehman et al. 2010). However, these approaches have so far been much less successful in indentifying the genes that underlie non-Mendelian complex genetic disorders. In such disorders, locus heterogeneity, reduced penetrance, and the frequent occurrence of phenocopies all work to obscure the correlation of a particular genetic variant with stuttering. Thus linkage information, based on co-segregation of a potentially causative variant with the disorder in families, has been used in the successful application of next generation sequencing technologies in disease gene discovery (Rehman et al. 2010; Sobreira et al. 2010).

In pursuit of additional gene loci for stuttering, we ascertained an extended family in Cameroon, West Africa, designated CAMST01 that includes a large number of affected individuals. In contrast to Pakistani families, CAMST01 has no genealogical evidence and no strong marker evidence (e.g., long stretches of homozygous markers) for consanguinity. The existence of multiple affected individuals in several different lineages within the family raised the possibility that assortative mating contributed to the large total number of affected members present in the pedigree. Assortative mating has been shown to play a role in the epidemiology and management of hereditary deafness (Arnos et al. 2008).

Materials and Methods

Family ascertainment and sampling

Family CAMST01 was ascertained via the International Stuttering Awareness Day Online Conference hosted by the Stuttering Home Page (<http://www.mnsu.edu/comdis/kuster/>). The family members and 47 age- and sex-matched normal Cameroonian control subjects were enrolled with written informed consent approved by an IRB of the National Institutes of Health (protocol # 97-DC-0057) and by the Institute of Tropical Medicine IRB, Kumba, South West, Republic of Cameroon. All family members and normal controls came from North West Province, Republic of Cameroon. DNA from peripheral blood was obtained from 51 family members, 48 of whom provided recorded speech samples.

Diagnosis and assignment of affection status

After removing individuals whose genotypes were inconsistent with the reported genealogical relationship, the pedigree contained 71 individuals. Of these 71 individuals, 42 were evaluated using the Stuttering Severity Index, 3rd Edition (SSI-3) (Riley 1994) to quantify stuttering dysfluencies in English. Family members' first language was a native Cameroonian dialect, although they were generally multilingual, with English commonly used. Affected family members had undergone a variety of speech interventions, and displayed persistent stuttering in all languages spoken.

For the analyses shown here, we considered individuals with a disfluency score < 4.07 as unaffected, and there were six such individuals. We considered individuals with a disfluency score ≥ 4.07 affected, and there were 36 such individuals. Six individuals were assigned unknown status (0) because of lack of a recorded sample, although family history suggests that 110, 156, 158 are affected and 141, 191, 205 are unaffected.

Genotyping

Microsatellite genotyping was performed using 50 ng and 25 ng of genomic DNA extracted from peripheral blood for multiplex and single plex PCR reactions, respectively. The initial genome-wide scan was performed using the Marshfield Weber 10 panel set (<http://www.marshfieldclinic.org/research/pages/index.aspx>). From this set, 332 autosomal microsatellite markers passed quality control. We genotyped additional microsatellite markers for fine mapping, including seven on chromosome 2, eight on chromosome 3, three on chromosome 14, and 55 on chromosome 15. PCR amplifications were performed in 10 l reaction volumes using thermocycling programs and the reaction conditions as previously described (Weber and Broman 2001). Capillary electrophoresis was performed on ABI PRISM® 3730 Genetic Analyzer. GeneMapper software was used to call and extract the genotypes from the electropherograms generated by the 3730 Genetic Analyzer. SNP genotyping was performed using the Illumina Infinium II Assay and Human Linkage-12 Panel, which includes 6,090 single-nucleotide polymorphism (SNP) markers chosen from validated HapMap DNA assays. These SNPs are distributed with an average spacing of 0.58 cM. Illumina BeadStudio v3.2 was used for analysis of genotypes as previously described (Raza et al. 2012). A total of 5670 SNPs passed quality control. All base pair positions shown are from human genome build 36/hg18. Haplotypes were constructed using MERLIN (Abecasis et al. 2002).

Linkage analysis of single loci

Preliminary LOD score computations were done with MERLIN. Because MERLIN cannot handle pedigrees as large as CAMST01 and because it was quickly evident that no single locus can explain the stuttering in CAMST01, the pedigree was split into five subpedigrees. Three of these subpedigrees represented the offspring of the three wives of individual 112. The remaining two subpedigrees represented the descendants of different siblings of individual 112 (Figure 1). To balance the risk of multiple testing with the risk of splitting the pedigree too finely, we split the pedigree only once, but analyzed all possible pedigree subsets derived from the split. LOD scores for these analyses were computed with Superlink (Fishelson and Geiger 2002).

We first performed single-marker analysis on sets of subpedigrees. In Figure 1, the overall pedigree is divided into five largely non-overlapping subpedigrees, lettered A through E. Therefore, there are $2^5 - 1 = 31$ ways to choose a subpedigree set that includes at least one of A through E. We analyzed all 31 such subpedigree sets. We computed single-marker LOD scores at all genome scan microsatellite markers, covering the 22 autosomes, and 150 different penetrance functions, later reduced to 30. The penetrance functions represent one of three alternative modes of inheritance: dominant, additive, or recessive. The approach of trying multiple penetrance functions has been studied statistically, (Greenberg et al. 1998; Sham et al. 2000). These studies suggested that the LOD scores generated by such optimization are statistically valid, but one should add 0.3 ($= \log_{10} 2$) or perhaps twice that to thresholds typically used to decide whether a LOD score has genome-wide significance.

Penetrance functions

The trait locus is modeled as having two alleles, which we denote by h(healthy) and d(disease-associated). A penetrance function has three terms: $P(\text{affected} | hh)$, $P(\text{affected} | hd)$, and $P(\text{affected} | dd)$. In these penetrance functions, the first term varied between 0.01 and 0.05, the second term varied from 0.01 to 0.99, and the third term varied from .5 to 0.99. In this notation, dominant inheritance means that $P(\text{affected}|hd) = P(\text{affected}|dd)$, recessive inheritance means $P(\text{affected}|hh) = P(\text{affected}|hd)$, and additive inheritance means that $P(\text{affected}|hh) < P(\text{affected}|hd) < P(\text{affected}|dd)$. We used only combinations in which

$P(\text{affected}|\text{hh})$ $P(\text{affected}|\text{hd})$ $P(\text{affected}|\text{dd})$ on the assumption that additional copies of the trait-associated allele would not reduce the probability of being affected.

Allele frequencies

Trait allele frequencies were generally set at 0.01. Marker allele frequencies were set by averaging in 50:50 proportions the genotypes in the pedigrees and the genotypes in unrelated controls from Cameroon. The use of a 50:50 mixture was a cautious choice that avoids inflating LOD scores for alleles that are rare in controls. Because microsatellite markers are highly polymorphic, it is not unusual for a certain allele to be entirely absent in the control data, necessitating adjustment of the allele frequencies which was done using this mixture. For the sake of consistency, we used a mixture for both the microsatellite and SNP data. We tested the sensitivity of the highest scores to the 50:50 mixture in Results.

Multi-marker analysis and evaluation of loci

Loci of interest were suggested by single marker analysis with microsatellites. Each locus was assigned one “core microsatellite” that was used in multi-marker analysis with nearby markers, either microsatellites or SNPs. Markers were combined in subsets of size two, three, and four with the trait locus moving across the marker map. Marker positions were from the Rutgers map (Matise et al. 2007). For two-locus analysis, the loci on 3p and 3q were modeled as unlinked. Because the pedigree is too large to use the Lander-Green algorithm that could consider all markers e.g. (Abecasis et al. 2002; Ott 1999), we were obliged to use a LOD score computation method such as Superlink in which only a small subset of markers can be considered. Nevertheless, the microsatellites are sufficiently informative that at a true positive locus, one would expect that multi-marker analysis with different marker sets should give similar scores.

The loci we identified scored high in subpedigree sets containing one to four of the subpedigrees A through E. To evaluate these loci in the context of the full pedigree, we computed heterogeneity LOD scores (HLOD) using the program HOMOG (Ott 1999). For example, if a locus scored high in subpedigree set ACD, then we modeled that there are three subpedigrees, ACD, B, and E, and presented three sets of scores to HOMOG. The extent of each locus was determined by the peak LOD – 1 rule (Ott 1999).

Linkage simulations

Linkage simulations were performed using the package FastSLINK (Ott 1989; Schäffer et al. 2011) to estimate power. We used 1000 replicates one or two markers each with five equally frequent alleles, displaying a recombination fraction with the trait locus of 0.05. The primary purpose of the power calculations was to see how observed LOD scores compared to simulated scores, for various subpedigrees and penetrance functions. Because we were interested in “typical behavior” of the maximum and average scores, we used a standard marker heterozygosity of 0.8.

Two-locus analysis

We used the two-locus option of Superlink to determine whether simultaneous analysis of multiple loci would yield a higher score, or yield high scores in larger subpedigrees than analysis of single loci. The settings for two-locus penetrance functions are explained in Results because our usage of Superlink two-locus analysis is context-dependent. The two loci can be either unlinked or linked, and we performed analyses under both conditions. Some Superlink computations were feasible only via the Superlink-online version that uses many computers in a distributed fashion (Silberstein et al. 2006).

Results

We first performed a linkage scan of the autosomes using microsatellites. LOD scores were computed and optimized for mode of inheritance as described in Methods. In the initial scan, twenty-one markers displayed a LOD score of at least 1.0 on the full pedigree (Table 1). Several of these, including the highest scoring markers, were found on chromosomes 2 and 3.

We divided the full pedigree into five subpedigrees labeled A through E, as shown in Figure 1. We then computed LOD scores for all 31 subpedigree combinations. Markers with LOD scores above 2.0 are shown in Table 2. As in Table 1, the scores for Table 2 have been optimized for mode of inheritance, but more coarsely than was done for Table 1.

The data in Table 2 suggested several regions of interest, and markers on chromosomes 2 and 3 again produced the highest scores. Marker D15S659 was also of interest because it segregates nearly perfectly with stuttering in subpedigree E. We then performed genotyping at a total of 5670 SNP loci distributed over all autosomes and additional microsatellite markers for fine mapping at selected locations. Based on multi-marker runs using these data, we found evidence for linkage on the short arms of chromosomes 2 and 3, on the long arm of chromosome 3, and on chromosomes 14 and 15. We present these loci in genome order because they are not easily ordered by strength of evidence. Then we summarize the loci in Table 3, and we compare the evidence among loci. Finally, we carry out two-locus analyses to test if pairs of loci could together explain linkage in a larger set of subpedigrees.

Chromosome 2p

A peak LOD score of 2.81 at D2S405 on the full pedigree first drew our attention to this locus. There was a group of high-scoring markers on the short arm of chromosome 2 extending from about 29 Mbp to 31 Mbp. The subpedigree set ACD generated the highest scores at this locus under a recessive mode of inheritance with incomplete penetrance (0.01;0.01;0.8). The marker D2S405 generated a LOD score of 3.69 in subpedigree set ACDE. Preliminary multi-marker computations suggested that omitting subpedigree E resulted in a modest decrease in single-point LOD scores, but improved multi-marker LOD scores (data not shown). We therefore considered subpedigree set ACD to be optimal.

Supplemental Table 1 and supplemental figure 2 show LOD scores using markers on chromosome 2p. Single-marker LOD scores are shown at recombination fraction $\theta = 0$. For multi-marker analyses, one marker was allowed to vary; those that were fixed are so indicated. Multi-marker LOD scores represent either the peak LOD scores with the trait locus located within the interval defined by the markers (those labeled *inner*), or located outside the interval (those labeled *outer*). Outer LOD scores were unidirectional, computed only for the end that was bounded by the marker that was allowed to vary. Combinations not calculated are marked with a “---”.

The SNP rs2272386 (28.72 Mbp) appeared to be the telomeric boundary of this linkage interval. Single-marker scores suggested that the interval may extend as far as rs305175 (36.22 Mbp), but multi-marker runs suggested that the interval extends no further than rs1054889 (31.41 Mbp). The peak LOD score using markers rs11127193, D2S405 and rs7560152, which are all firmly within the linkage interval, was 3.86. To better understand the significance for this score, we performed simulations in SLINK. These simulations compared the score on our observed data with the empirical cumulative distribution function of scores on simulated data. This score of 3.86 had a percentile of 94.7 with two flanking markers, and thus it exceeded 94.7 percent of the scores obtained in these simulations, suggesting that it is a true positive.

Chromosome 3p

A locus on the short arm of chromosome 3 contains D3S2432, the marker that attained the highest score on the full pedigree. The interval with high linkage scores extended from about 29 to about 38.5 Mbp. The optimal subpedigree set for this locus was ACE under a recessive mode of inheritance, using the penetrance function 0.01;0.01;0.8. The marker D3S2432 attained a LOD score of 2.97 on the full pedigree with the penetrance function 0.03;0.03;0.6. D3S2432 attained a LOD score of 3.69 on subpedigree set ACDE with the penetrance function 0.01;0.01;0.8. However, ACE had higher two-marker LOD scores using D3S2432 and its neighboring microsatellite markers D3S3727 (LOD score 2.71) and D3S3518 (LOD score 2.98), so we take ACE as the optimal subpedigree set for this locus.

LOD scores for markers near D3S2432 are shown in Supplemental Table 2 and supplemental figure 3. The marker rs1381397 (28.74 Mbp) appeared to be the telomeric boundary of the linkage region, and rs762318 (38.49 Mbp) appeared to be the centromeric boundary. The highest multi-marker LOD score was attained for the markers rs304838, D3S2432 and D3S3518 with a peak score of 3.18 between rs304838 (30.78 Mbp) and D3S2432 (32.14 Mbp). The score of 3.18 has a percentile of 97.4 in SLINK simulations with two flanking markers, suggesting that it is a true positive.

Chromosome 3q

We observed a high-scoring locus on the long arm of chromosome 3 spanning the interval from approximately 196 to 199 Mbp. The optimal subpedigree set for this locus was ADE under a recessive mode of inheritance using the penetrance function 0.01;0.01;0.99. On the full pedigree, microsatellite D3S1311 attains a peak LOD score of 1.42 with a recessive penetrance function 0.01;0.01;0.6. The optimal subpedigree set for this marker is ADE, where it attained a LOD score of 2.90 under the penetrance function 0.01;0.01;0.99. Two-marker scores using neighboring SNPs (Table 2 and 3) and preliminary three-marker scores (data not shown) were promising, so we genotyped the nearby microsatellite D3S1305.

Supplemental Table 3 and supplemental figure 4 show three-marker LOD scores calculated by keeping D3S1311 and D3S1305 fixed and varying the third marker. LOD scores above 3.4 were common from this analysis, with the highest scoring combination being rs711995, D3S1311, and D3S1305, which produced a peak LOD of 3.47 between rs711995 and D3S1311. The score of 3.47 has a percentile of 98.2 in SLINK simulations with two flanking markers, suggesting that it is a true positive. The SNP rs7627589 (195.33 Mbp) appeared to be the centromeric boundary of this linkage region, and rs718501 (198.57 Mbp) appeared to be the telomeric boundary. The most centromeric marker listed in Table 3 q, D3S2418, is less than 2Mbp from the marker D3S3054, which is at the center of a locus identified in a study of stuttering in the Hutterite population (Wittke-Thompson et al. 2007). Thus, our linkage at 3q in CAMST01 is a replication of this finding.

Chromosome 14

We observed a high-scoring locus on chromosome 14 extending from approximately 65.5 to 70.7 Mbp. The optimal subpedigree set is ABDE with a recessive penetrance function 0.01;0.01;0.99. The microsatellite D14S588 attains a LOD score of 3.31 on ABDE (Supplemental Table 4 and supplemental figure 5). Scores for two-marker runs in which marker D14S588 was fixed were highly variable. Three-marker runs fixing D14S588 and rs987579 gave higher LOD scores (data not shown), but the addition of the intermediate marker rs8688 in four-marker analyses lowered the LOD score. Based on multi-marker analyses, the marker D14S125 (65.45 Mbp) marked the proximal boundary and rs221924 (70.65 Mbp) marked the distal boundary. The highest LOD score was 3.45, obtained using rs975232, D14S588, rs8688, and rs987579. The score of 3.45 has a percentile of 87.5 in

SLINK simulations with two flanking markers, which is not as high as the percentiles for the best multipoint scores on 2p, 3p, or 3q.

Chromosome 15

There were two weakly linked loci on chromosome 15. One locus came to our attention because the marker D15S659 (44.16 Mbp) segregates perfectly with stuttering in subpedigree E in a dominant manner, yielding a LOD score of 2.01, above the 2.0 threshold used for Table 2. We genotyped additional markers in the region, several of which also generated a LOD score of 2.01 on this subpedigree, which is the maximum attainable for the settings we used of a penetrance function 0.01;0.99;0.99 and a trait allele frequency 0.01; see Supplemental Table 5. This linkage region extends from rs1426932 (43.47 Mbp) to D15S962 (54.36 Mbp).

We investigated nearby markers on chromosome 15 on other subpedigrees for two reasons. First, there was some evidence of linkage to D15S822 in subpedigree set CD. Second, a prior study had suggested a locus on chromosome 15 at approximately 23 cM (Suresh et al. 2006). Investigations of the centromere-proximal portions of chromosome 15 revealed that the marker GATA50C03 (34.78 Mbp, 34.58 cM) had a single-point LOD score of 1.71 on subpedigree B with a dominant mode of inheritance. This score was comparable to the peak scores derived by SLINK; for subpedigree B, LOD scores are sensitive to marker allele frequencies. Several nearby markers also had LOD scores above 1, and multi-marker runs yielded scores above 2; see Supplemental Table 6 and supplemental figure 6. The large region consistent with linkage in subpedigree B extended from rs2703955 (25.85 Mbp) to rs276855 (37.32 Mbp). We ignored the low score at D15S1232 when calculating this extent, because the scores at nearby markers were much higher. This encompasses and thus replicates the previously suggested locus. The promising region on subpedigree B was near the region optimal for subpedigree E, but did not overlap (Supplemental Tables 5 and 6; supplemental figure 6).

The potentially distinct linkage regions observed in B and E suggested an analysis of these subpedigrees together. On chromosome 15, the highest scoring mode of inheritance for the BE subpedigree is additive (Supplemental Table 7). Scores at intermediate markers, however, were inconsistent (data not shown), so additive inheritance was not an entirely satisfactory explanation. The peak multi-marker LOD score for the microsatellites D15S514, D15S537, and D15S659 was 1.55.

Comparing the evidence for different loci

Table 3 summarizes the evidence for each locus and adds HLOD values (see Materials and Methods) and SLINK percentiles, which represent the rank of observed score compared to the simulated scores. The SLINK replicates were generated and analyzed under the same single-locus penetrance model, while the observed data represent likely polygenic inheritance analyzed under an oversimplified single-locus model.

Among the four recessive loci, the evidence for linkage on chromosomes 2p, 3p, and 3q is stronger than the evidence for linkage on chromosome 14. The chromosome 14 locus has the lowest SLINK percentile and multi-marker scores do not give a smooth peak (Supplemental Table 4 and supplemental figure 5). Support for the other three recessive loci is stronger. The locus on chromosome 2 has the highest LOD and HLOD, but is narrow. The locus on 3p is the broadest of the recessive loci and has a high SLINK percentile (97.4%). The locus on 3q has the strengths that it replicates a previously reported linkage (whereas linkages on 2p and 3p are novel), has the highest SLINK percentile (98.2%) and the multi-marker scores show an ideal, flat peak (Supplemental Table 3 and supplemental figure 4). For a selection

of high-scoring markers, we tested the sensitivity of the choice of a 50:50 mixture, compared to 1:99, 25:75, 75:25 or 99:1 mixtures, and report the outcome of this test in Supplemental Table 8. Except at the chromosome 14 locus, for which the evidence is weakest, the high LOD scores vary little.

Considered as single loci, the three possible loci on chromosome 15 in Table 3 are much weaker than the four recessive loci based on the total LOD scores and the HLODs, but we report them partly because they perform well in two-locus analyses.

Two-locus analyses

After recognizing that various loci showed evidence of linkage in different subpedigree sets, we investigated how these loci could segregate in pairs. This is feasible via the two-locus analysis option of Superlink-online (Silberstein et al. 2006). Unfortunately, analysis of more than two loci simultaneously is not implemented. Due to computational limitations, we used only one microsatellite per locus. We used penetrance functions that are a composite of the functions that give high scores at single loci as suggested by (Strauch et al. 2003) but unlike their approach, we use different penetrance classes for different parts of the pedigree.

We used two types of penetrance functions. The first type of penetrance function was used when the optimal subpedigree sets for the two loci did not overlap (e.g., ACD and B). For this type, we included two penetrance classes in our model because there was not compelling evidence that trait alleles for more than one locus entered either of the disjoint sets. Individuals were assigned to a class based on their position in a pedigree. Individuals in the subpedigree set optimal for the first locus were assigned to class 1. Similarly, individuals in the subpedigree set optimal for the second locus were assigned to class 2. Within each class, the corresponding trait-associated allele was assumed to explain the phenotype; Table 4 shows the penetrance function used for analysis of chromosome 2p (marker D2S405) with chromosome 15 (marker D15S537) under additive inheritance.

The second type of penetrance function was used when the optimal subpedigree sets for the two loci were not disjoint. When the subpedigree sets were not disjoint, this suggested that: in the subpedigrees unique to locus 1, only trait-associated alleles for locus 1 entered, in the subpedigrees unique to locus 2, only alleles for locus 2 entered, and in the subpedigrees shared by the two loci, alleles for both loci entered. Therefore, we included three penetrance classes.

The resulting penetrance functions are similar to that shown in Table 5, which shows the penetrance function used for chromosomes 2p and 3q. Although the best model for each single locus was recessive, we had limited information regarding the best choice for the middle value in the table, corresponding to individuals that inherit one trait-associated allele at each locus. This value is low in the simpler no-epistasis model that gave better scores. Thus, in Table 5, the middle value for class 3 is low and that was used in the analyses shown.

Table 6 shows LOD scores generated by combining pairs of high-scoring loci. Next to each LOD score is the subpedigree set used. Analyzing the data on chromosome 15 as two weakly linked loci yields a LOD score of 3.42 on subpedigree set BE. This score is more than 1 LOD unit better than any single-locus explanation we tested, suggesting that there are two distinct stuttering loci on chromosome 15 (one previously described and one new). When we analyze chromosome 15 in conjunction with a locus on another chromosome, we are forced to represent chromosome 15 by a single additive locus. Despite this limitation, the 2p locus also has notably high LOD scores when combined with loci on 15. As explained in the Discussion, the 2p+15 two-locus LOD score is sensitive to the marker chosen to

represent 15. The highest score we obtained was 6.57, combining the highest scoring marker in Supplemental Table 7 (D15S537, with a score of 2.99) with D2S405 representing 2p. However, if we use instead a marker that scores much lower with the additive model on 15q, such as D15S1039 (scores 1.12 in Supplemental Table 7), then the two-locus score drops to 4.69. Tests with several other markers demonstrated a pattern that the 2p+15 two-locus score was approximately 3.57 plus the additive 15q score in Supplemental Table 7.

Discussion

Although twin studies have indicated a high heritability for stuttering, further genetic studies have been hampered by a number of features of the disorder, including unequal rates of occurrence in males and females, a high recovery rate, and a general lack of Mendelian transmission in families (Drayna et al. 1999; Bloodstein and Ratner 2011). Studies in consanguineous families have generated a number of clear linkage loci for this disorder (Raza et al. 2012; Raza et al. 2010) and one such locus on chromosome 12 (Riaz et al. 2005) has led to the discovery of causative genes for this disorder (Kang et al. 2010). However the known loci can explain only a fraction of familial stuttering.

This study differs from previous genetic studies of stuttering in that we studied individuals from Africa, all of whom are multilingual. There have been a variety of studies on multilingualism and stuttering (reviewed in Van Borsel et al. 2001 and Shenker 2011). Most, but not all multilingual stutterers stutter in all languages spoken, though sometimes in varying degrees. A more controversial question is whether multilingual individuals have a higher incidence of stuttering; a positive answer would give a cautionary note to genetic studies since it would identify one environmental factor contributing to stuttering. Van Borsel et al. state that “The belief that stuttering is more prevalent in bilinguals than in monolinguals seems to be widespread indeed” (Van Borsel et al. 2001). Yet, ten years later, based on extensive clinical experience in bilingual Québec, Shenker (2011) concluded that “there is no credible research to support or refute the idea that bilingualism increases the risk of stuttering or to justify asking a child to become monolingual.”

The results of our study of family CAMST01 are consistent with previous genetic findings in stuttering. We have replicated two previously suggested linkage loci on chromosomes 15q and 3q. We have also generated substantial evidence for novel loci on chromosomes 2p, 3p, and 15q, together with suggestive evidence for a novel locus on chromosome 14, with peak single-locus LOD scores ranging from 3.18 to 3.86. Our statistical evidence for linkage at these single loci is strengthened by the results of our two-locus analyses. Model-based analysis of two loci has been possible for some time through either TLINKAGE (Schork et al. 1993) or GENEHUNTER-TWOLOCUS (Dietter et al. 2004), but the CAMST01 pedigree is too large for these methods. Therefore, we used the two-locus capability of Superlink (Silberstein et al. 2006), including the capability to analyze two weakly linked loci on chromosome 15. While Superlink has proven highly useful (<http://cbl-hap.cs.technion.ac.il/superlink-snp/successStories.php>), published Superlink studies have not previously used the two-locus feature. Our results indicate that such two-locus analysis may be useful for other large pedigrees of complex traits, especially in the presence of assortative mating.

To understand why the two loci on chromosome 15 appear to be separate and why the scores for a single additive locus are variable from marker to marker, we computed haplotypes. Supplemental Figure 1 shows two possible haplotype assignments for chromosome 15 in subpedigree set BE over a region spanning both loci on chromosome 15. In light of the large pedigree, it was surprising that MERLIN would choose different haplotype assignments depending which individuals or markers were included.

Supplemental Figure 1, panel a) illustrates the evidence that the two chromosome 15 loci are separate. The 8-5-10-3-7 (haplotype 2) haplotype segregates perfectly in subpedigree E, except that there is a crossover, so that affected individual 200 has only the lower 10-3-7 subhaplotype, partly explaining the high scores on E further down the chromosome. Affected individuals in B appear to share a 3-3-11 (haplotype 1) haplotype, but the lower part is not shared by individuals 184 and 185 in B.

The more telomeric markers have a highly unusual pattern of allele sharing. Either or both of two haplotypes, namely 5-6-4-13-3-1 (haplotype 1) and 7-9-2-11-3-6 (haplotype 3), that are inferred to occur in the ungenotyped individual 112 can also be inferred to occur in the children of 127 in subpedigree B. The pedigree structure, however, suggests that the children of 127 can share at most one of these haplotypes identical-by-descent with 112. This results in ambiguity in the haplotypes, illustrated by comparing panel b with panel a in Supplemental Figure 1. Analysis using all 37 available microsatellites between D15S971 and D15S1032 failed to resolve this ambiguity, suggesting that both haplotypes could be present in the children of 127.

The ambiguity in the haplotypes becomes more interesting when one notices that six out of seven affected children of 112 in E inherited the 5-6-4-13-3-1 (haplotype 1) haplotype from 112, even though it is the 8-5-10-3-7 (haplotype 2) haplotype inherited from 111 that segregates nearly perfectly with the trait in E. Since 111 is the affected ancestor in E and the marriage of 111 and 112 connects subpedigree B to subpedigree E, one would not expect haplotypes from 112 to recur so often among the affected individuals in both subpedigrees. The unusual inheritance pattern of haplotypes suggests that the pedigree as drawn in Figure 1 may be missing at least one additional ancestral relationship that connects subpedigrees B and E. The markers that produce the highest score when analyzing BE as a single additive locus (Supplemental Table 7) are those such as D15S971 and D15S537, at which the trait-associated allele in B arises on a haplotype that is not associated with the trait in subpedigree E. The large number of high-scoring markers in our additive BE analysis provides additional evidence for a common ancestor between subpedigrees B and E (Supplemental Table 7).

The evidence of two shared chromosome 15 haplotypes and many high-scoring markers spread around chromosome 15 is the only evidence we see for a founder effect, in which either multiple haplotypes or multiple copies of the same haplotype from an unknown distant founder segregate in the pedigree. At recessive loci, such as the four loci we described above on 2q, 3p, 3q, and 14q, a founder effect would typically manifest as homozygosity for a haplotype that is passed identical by descent via two paths; we see no evidence of homozygosity at these loci or on chromosome 15. However, the evidence of two shared haplotypes on chromosome 15 does suggest that pedigrees B and E have a common founder. If that is so, one mating would have to be consanguineous and one should ask why we do not see homozygosity for some chromosome 15 marker interval in affected individuals in subpedigrees B and E. There are three possible answers that are not mutually exclusive. First, the preferred mode of inheritance is dominant, so there is no need for homozygosity on chromosome 15 to obtain positive evidence of linkage on chromosome 15. Second, the consanguineous mating could involve only ungenotyped parents and offspring (e.g., 107, 108, 127); in that scenario, the homozygosity by descent exists, but we do not see the homozygosity because it exists only in individuals who are ungenotyped. Third, the common ancestor is so many generations back that the resulting homozygous intervals are too small to detect at the moderate density of our genotyping.

An important issue in two-locus analysis is what LOD scores to treat as significant. (Schork et al. 1993) suggested adding 0.5 to the LOD score threshold. A more general approach was

suggested by (Ott 1999), which is to add 0.3 for each extra free parameter. In doing two-locus analysis we are choosing pedigree subsets for each locus and a mode of inheritance for each locus. We consider this to involve four choices, so it is appropriate to add 1.2 to the LOD score threshold one would use for standard single-locus analysis. Since the current typically used threshold is 3.0–3.3, this suggests that a threshold of 4.2–4.5 should be used for significance. Our two-locus LOD scores of 4.55–6.57 (Table 6) are thus strong support for linkage.

Family CAMST01 was ascertained in large part because it includes a large number of affected individuals. While large families with many members who stutter have been described in outbred populations (Macfarlane et al. 1991), such families are rare and most family clusters of stuttering contain a modest number of cases (Canhetti-Oliveira and Richieri-Costa 2006; Shugart et al. 2004; Suresh et al. 2006; Viswanath et al. 2004; Wittke-Thompson et al. 2007). Since we could find no substantial evidence of consanguinity in family CAMST01, two hypotheses were possible. One hypothesis was that stuttering in this family was due to an allele at a single locus with an unusually large effect that was distributed across the family by polygamous marriages. An alternative hypothesis was that multiple alleles at different loci gave rise to the many observed cases in this family. Our analysis demonstrates that the latter is the case, and thus to date, single highly penetrant alleles that exert an effect large enough to generate very large families that contain many affected individuals have rarely been observed.

Preliminary analyses (data not shown) did not show a substantial difference when different parameters were used for males and females, which differs from results found in a previous linkage study (Suresh et al. 2006). CAMST01 may have unique genetic and environmental causes of stuttering, not subject to the gender differential observed in other samples.

Our findings give rise to the question of how such multiple alleles came into this single extended family. Stuttering in Cameroon has not been shown to occur at a higher rate than elsewhere. For the loci with a recessive mode of inheritance, most affected individuals are not homozygous at the highest scoring markers. Thus, the genotype evidence does not support the hypothesis that the recessive trait alleles originated in a common founder. There is an unusual pattern of haplotype sharing between subpedigrees B and E for the dominant locus of 15q, suggesting an unknown common ancestor, but this haplotype sharing is not sufficient to explain the trait locus in the entire family. One possible hypothesis is assortative mating, in which stuttering individuals non-randomly marry other stuttering individuals. Assortative mating became common in hereditary deafness, another communication disorder, when non-random mating was associated with the attendance of large numbers of affected individuals at schools or universities for the deaf (Arnos et al. 2008). Factors that could contribute to such assortative mating in this population are unknown.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This study was supported by the Intramural Research Program of the NIH, NIDCD (Z01-000046-11) and the NLM (LM000097), and by the Stuttering Foundation. We are especially grateful to the members of family CAMST01 for their participation. We thank Bailey Levis for genotyping during the early stages of this project, and Drs. Thomas Friedman and Robert Morell for suggestions that improved the manuscript.

References

- Abecasis GR, Cherny SS, Cookson WO, Cardon LR. Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet.* 2002; 30(1):97–101.10.1038/ng786 ng786 [pii] [PubMed: 11731797]
- Andrews G, Morris-Yates A, Howie P, Martin NG. Genetic factors in stuttering confirmed. *Arch Gen Psychiatry.* 1991; 48 (11):1034–1035. [PubMed: 1747019]
- Arnos KS, Welch KO, Tekin M, Norris VW, Blanton SH, Pandya A, Nance WE. A comparative analysis of the genetic epidemiology of deafness in the United States in two sets of pedigrees collected more than a century apart. *Am J Hum Genet.* 2008; 83(2):200–207. S0002-9297(08)00398-4 [pii]. 10.1016/j.ajhg.2008.07.001 [PubMed: 18656178]
- Bloodstein O. Stuttering in families of adopted stutterers. *J Speech Hear Disord.* 1961; 26:395–396. [PubMed: 13870118]
- Bloodstein, O.; Ratner, NB. A handbook on stuttering. 6. Thomson Delmar Learning; 2011. illustrated edn
- Canhetti-Oliveira CM, Richieri-Costa A. A study of familial stuttering. *Am J Med Genet A.* 2006; 140(19):2139–2141.10.1002/ajmg.a.31320 [PubMed: 16752384]
- Craig A, Hancock K, Tran Y, Craig M, Peters K. Epidemiology of stuttering in the community across the entire life span. *J Speech Lang Hear Res.* 2002; 45 (6):1097–1105. [PubMed: 12546480]
- Dietter J, Spiegel A, Mey DA, Pflug HJ, Al-Kateb H, Hoffmann K, Wienker TF, Strauch K. Efficient two-trait-locus linkage analysis through program optimization and parallelization: application to hypercholesterolemia. *Eur J Hum Genet.* 2004; 12(7):542–550.10.1038/sj.ejhg.5201196 [PubMed: 15100714]
- Drayna D, Kilshaw J, Kelly J. The sex ratio in familial persistent stuttering. *Am J Hum Genet.* 1999; 65(5):1473–1475. S0002-9297(07)62159-4 [pii]. 10.1086/302625 [PubMed: 10521318]
- Dworzynski K, Remington A, Rijdsdijk F, Howell P, Plomin R. Genetic etiology in cases of recovered and persistent stuttering in an unselected, longitudinal sample of young twins. *Am J Speech Lang Pathol.* 2007; 16(2):169–178. 16/2/169 [pii]. 10.1044/1058-0360(2007/021) [PubMed: 17456895]
- Fagnani C, Fibiger S, Skytthe A, Hjelmberg JVB. Heritability and environmental effects for self-reported periods with stuttering: A twin study from Denmark. *Logoped Phoniatr Vocol.* 2011; 36(3):114–120.10.3109/14015439.2010.534503 [PubMed: 21080843]
- Felsenfeld S. Finding susceptibility genes for developmental disorders of speech: the long and winding road. *J Commun Disord.* 2002; 35 (4):329–345. [PubMed: 12160352]
- Felsenfeld S, Kirk KM, Zhu G, Statham DJ, Neale MC, Martin NG. A study of the genetic and environmental etiology of stuttering in a selected twin sample. *Behav Genet.* 2000; 30 (5):359–366. [PubMed: 11235981]
- Felsenfeld S, Plomin R. Epidemiological and offspring analyses of developmental speech disorders using data from the Colorado Adoption Project. *J Speech Lang Hear Res.* 1997; 40 (4):778–791. [PubMed: 9263943]
- Fishelson M, Geiger D. Exact genetic linkage computations for general pedigrees. *Bioinformatics.* 2002; 18(Suppl 1):S189–198. [PubMed: 12169547]
- Godai U, Tatarelli R, Bonanni G. Stuttering and tics in twins. *Acta Genet Med Gemellol (Roma).* 1976; 25:369–375. [PubMed: 1036386]
- Greenberg DA, Abreu P, Hodge SE. The power to detect linkage in complex disease by means of simple LOD-score analyses. *Am J Hum Genet.* 1998; 63(3):870–879. S0002-9297(07)61390-1 [pii]. 10.1086/301997 [PubMed: 9718328]
- Howie PM. Concordance for stuttering in monozygotic and dizygotic twin pairs. *J Speech Hear Res.* 1981; 24 (3):317–321. [PubMed: 7197739]
- Kang C, Riazuddin S, Mundorff J, Krasnewich D, Friedman P, Mullikin JC, Drayna D. Mutations in the lysosomal enzyme-targeting pathway and persistent stuttering. *N Engl J Med.* 2010; 362(8): 677–685. NEJMoa0902630 [pii]. 10.1056/NEJMoa0902630 [PubMed: 20147709]
- Macfarlane WB, Hanson M, Walton W, Mellon CD. Stuttering in five generations of a single family: A preliminary-report Including evidence supporting a sex-modified mode of transmission. *Journal of Fluency Disorders.* 1991; 16 (2–3):117–123.

- Matisse TC, Chen F, Chen W, De La Vega FM, Hansen M, He C, Hyland FCL, Kennedy GC, Kong X, Murray SS, Ziegler JS, Stewart WCL, Buyske S. A second-generation combined linkage physical map of the human genome. *Genome Res.* 2007; 17(12):1783–1786. gr.7156307 [pii]. 10.1101/gr.7156307 [PubMed: 17989245]
- Ng SB, Bigham AW, Buckingham KJ, Hannibal MC, McMillin MJ, Gildersleeve HI, Beck AE, Tabor HK, Cooper GM, Mefford HC, Lee C, Turner EH, Smith JD, Rieder MJ, Yoshiura K, Matsumoto N, Ohta T, Niikawa N, Nickerson DA, Bamshad MJ, Shendure J. Exome sequencing identifies *MLL2* mutations as a cause of Kabuki syndrome. *Nat Genet.* 2010a; 42(9):790–793. ng.646 [pii]. 10.1038/ng.646 [PubMed: 20711175]
- Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, Huff CD, Shannon PT, Jabs EW, Nickerson DA, Shendure J, Bamshad MJ. Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet.* 2010b; 42(1):30–35. ng.499 [pii]. 10.1038/ng.499 [PubMed: 19915526]
- Ooki S. Genetic and environmental influences on stuttering and tics in Japanese twin children. *Twin Res Hum Genet.* 2005; 8(1):69–75. 10.1375/1832427053435409 [PubMed: 15836814]
- Ott J. Computer-simulation methods in human linkage analysis. *Proc Natl Acad Sci U S A.* 1989; 86(11):4175–4178. [PubMed: 2726769]
- Ott, J. Analysis of human genetic linkage. 3. Johns Hopkins Univ. Press; Baltimore: 1999.
- Raza MH, Amjad R, Riazuddin S, Drayna D. Studies in a consanguineous family reveal a novel locus for stuttering on chromosome 16q. *Hum Genet.* 2012; 131(2):311–313. 10.1007/s00439-011-1134-2 [PubMed: 22205390]
- Raza MH, Riazuddin S, Drayna D. Identification of an autosomal recessive stuttering locus on chromosome 3q13.2–3q13.33. *Hum Genet.* 2010; 128(4):461–463. 10.1007/s00439-010-0871-y [PubMed: 20706738]
- Rehman AU, Morell RJ, Belyantseva IA, Khan SY, Boger ET, Shahzad M, Ahmed ZM, Riazuddin S, Khan SN, Friedman TB. Targeted capture and next-generation sequencing identifies *C9orf75*, encoding taperin, as the mutated gene in nonsyndromic deafness DFNB79. *Am J Hum Genet.* 2010; 86(3):378–388. S0002-9297(10)00034-0 [pii]. 10.1016/j.ajhg.2010.01.030 [PubMed: 20170899]
- Riaz N, Steinberg S, Ahmad J, Pluzhnikov A, Riazuddin S, Cox NJ, Drayna D. Genomewide significant linkage to stuttering on chromosome 12. *Am J Hum Genet.* 2005; 76(4):647–651. S0002-9297(07)62875-4 [pii]. 10.1086/429226 [PubMed: 15714404]
- Riley, GD. Stuttering severity instrument for children and adults, third edition (SSI-3). PRO-ED, Inc; 8700 Shoal Creek Blvd., Austin, TX: 1994. p. 78758-6897.
- Schäffer AA, Lemire M, Ott J, Lathrop GM, Weeks DE. Coordinated conditional simulation with SLINK and SUP of many markers linked or associated to a trait in large pedigrees. *Hum Hered.* 2011; 71(2):126–134. 000324177 [pii]. 10.1159/000324177 [PubMed: 21734403]
- Schork NJ, Boehnke M, Terwilliger JD, Ott J. Two-trait-locus linkage analysis: a powerful strategy for mapping complex genetic traits. *Am J Hum Genet.* 1993; 53(5):1127–1136. [PubMed: 8213836]
- Sham PC, Lin M-W, Zhao JH, Curtis D. Power comparison of parametric and nonparametric linkage tests in small pedigrees. *Am J Hum Genet.* 2000; 66(5):1661–1668. S0002-9297(07)62995-4 [pii]. 10.1086/302888 [PubMed: 10762550]
- Shenker RC. Multilingual children who stutter: Clinical issues. *J Fluency Disord.* 2011; 36(3):186–193. [PubMed: 22118395]
- Shugart YY, Mundorff J, Kilshaw J, Doheny K, Doan B, Wanyee J, Green ED, Drayna D. Results of a genome-wide linkage scan for stuttering. *Am J Med Genet A.* 2004; 124A:2–135. 10.1002/ajmg.a.20347
- Silberstein M, Tzemach A, Dovgolevsky N, Fishelson M, Schuster A, Geiger D. Online system for faster multipoint linkage analysis via parallel execution on thousands of personal computers. *Am J Hum Genet.* 2006; 78(6):922–935. S0002-9297(07)63915-9 [pii]. 10.1086/504158 [PubMed: 16685644]
- Sobreira NL, Cirulli ET, Avramopoulos D, Wohler E, Oswald GL, Stevens EL, Ge D, Shianna KV, Smith JP, Maia JM, Gumbs CE, Pevsner J, Thomas G, Valle D, Hoover-Fong JE, Goldstein DB. Whole-genome sequencing of a single proband together with linkage analysis identifies a

- Mendelian disease gene. *PLoS Genet.* 2010; 6(6):e1000991.10.1371/journal.pgen.1000991 [PubMed: 20577567]
- Strauch K, Fimmers R, Baur MP, Wienker TF. How to model a complex trait - 2. Analysis with two disease loci. *Human Heredity.* 2003; 56(4):200–211.10.1159/000076394 [PubMed: 15031621]
- Suresh R, Ambrose N, Roe C, Pluzhnikov A, Wittke-Thompson JK, Ng MC-Y, Wu X, Cook EH, Lundstrom C, Garsten M, Ezrati R, Yairi E, Cox NJ. New complexities in the genetics of stuttering: significant sex-specific linkage signals. *Am J Hum Genet.* 2006; 78(4):554–563. S0002-9297(07)63695-7 [pii]. 10.1086/501370 [PubMed: 16532387]
- van Beijsterveldt CEM, Felsenfeld S, Boomsma DI. Bivariate genetic analyses of stuttering and nonfluency in a large sample of 5-year-old twins. *J Speech Lang Hear Res.* 2010; 53(3):609–619. 1092-4388_2009_08-0202 [pii]. 10.1044/1092-4388(2009/08-0202) [PubMed: 20029049]
- Van Borsel J, Maes E, Foulon S. Stuttering and bilingualism. *J Fluency Disord.* 2001; 26(3):179–205.
- Viswanath N, Lee HS, Chakraborty R. Evidence for a major gene influence on persistent developmental stuttering. *Hum Biol.* 2004; 76 (3):401–412. [PubMed: 15481675]
- Weber JL, Broman KW. Genotyping for human whole-genome scans: past, present, and future. *Adv Genet.* 2001; 42:77–96. [PubMed: 11037315]
- Wittke-Thompson JK, Ambrose N, Yairi E, Roe C, Cook EH, Ober C, Cox NJ. Genetic studies of stuttering in a founder population. *J Fluency Disord.* 2007; 32(1):33–50. S0094-730X(06)00098-2 [pii]. 10.1016/j.jfludis.2006.12.002 [PubMed: 17276504]

Figure 1.

Pedigree of Cameroonian family CAMSTO1. Five sub-pedigrees (A through E) are shown by drawing four solid lines across the pedigree. Asterisks indicate the genotyped individuals. Black symbols designate the affecteds, unaffecteds are shown with blank symbols and the individuals with unknown phenotypes are denoted with grey symbols. Based on speech disfluency scores and the speech history obtained from family, phenotypes of few individuals were changed for some analysis that are not pointed out in this figure but in text.

Table 1

High-scoring markers in single-marker analysis of the entire CAMST01 pedigree

Marker	Chr	Mbp	cM	Penetrance	θ	LOD
D3S2432	03	32.14	56.76	0.02:0.02;0.7	0.0	2.97
D2S405	02	29.33	50.20	0.03:0.03;0.6	0.0	2.81
D2S1334	02	136.17	148.10	0.05:0.05;0.5	0.0	2.19
D2S1788	02	36.11	58.64	0.02:0.02;0.6	0.05	2.08
D3S1768	03	34.60	59.46	0.02:0.02;0.5	0.0	2.00
D2S2374	02	35.36	57.98	0.01:0.01;0.7	0.09	1.86
D1S1594	01	238.85	263.36	0.01:0.9;0.9	0.0	1.54
D3S1305	03	196.79	219.85	0.01:0.01;0.7	0.070	1.54
D2S2283	02	31.29	53.71	0.05:0.05;0.5	0.00	1.47
D2S367	02	34.29	56.88	0.02:0.02;0.5	0.09	1.46
D12S372	12	3.46	8.92	0.03:0.03;0.6	0.0	1.46
D3S1311	03	198.50	222.81	0.01:0.01;0.6	0.13	1.42
D3S3727	03	30.65	54.36	0.02:0.02;0.5	0.06	1.35
D18S858	18	53.05	81.99	0.03:0.03;0.6	0.06	1.27
D19S245	19	38.79	56.39	0.01:0.01;0.5	0.12	1.24
PAH	12	---	109.47	0.05:0.05;0.99	0.09	1.21
D2S2170	02	25.01	46.63	0.05:0.05;0.5	0.0	1.21
D2S2247	02	27.16	48.16	0.05:0.05;0.5	0.02	1.20
D2S1400	02	11.52	27.14	0.01:0.01;0.8	0.15	1.18
D2S2168	02	24.94	46.57	0.05:0.05;0.5	0.03	1.18
D2S1328	02	125.90	135.77	0.04:0.04;0.5	0.02	1.18
D22S689	22	27.19	35.21	0.05:0.05;0.5	0.01	1.17
D14S588	14	69.29	65.14	0.01:0.01;0.99	0.17	1.12
D8S373	08	144.30	167.45	0.01:0.01;0.6	0.13	1.04
D10S1239	10	103.19	120.67	0.05:0.99;0.99	0.0	1.03
D2S1360	02	17.36	37.92	0.05:0.05;0.5	0.01	1.02
D3S3518	03	33.66	59.01	0.02:0.02;0.5	0.10	1.02

Marker	Chr	Mbp	cM	Penetrance	θ	LOD
D14S608	14	27.92	21.54	0.01;0.2;0.5	0.0	1.02
D14S290	14	62.60	60.07	0.01;0.2;0.99	0.0	1.01
D18S1163	18	7.46	26.85	0.03;0.03;0.99	0.13	1.01
D18S1357	18	56.97	89.38	0.03;0.03;0.99	0.16	1.00

Table 2

High-scoring markers in subpedigree sets.

Subped.	Marker	Chr	Mbp	cM	Penetrance	θ	LOD
ACDE	D2S405	02	29.33	50.20	0.01;0.01;0.8	0.0	3.69
ACDE	D3S2432	03	32.14	56.76	0.01;0.01;0.8	0.0	3.52
ABCE	D3S1768	03	34.60	59.46	0.02;0.02;0.7	0.0	3.36
ABDE	D14S588	14	69.29	65.14	0.01;0.01;0.99	0.0	3.31
BCD	D2S1334	02	136.17	148.10	0.02;0.02;0.5	0.01	3.14
CD	D15S822	15	24.97	11.84	0.01;0.01;0.99	0.07	2.90
ADE	D3S1311	03	198.50	222.81	0.01;0.01;0.99	0.0	2.90
BCD	D1S1594	01	238.85	263.36	0.01;0.99;0.99	0.0	2.45
ACE	D3S3727	03	30.65	54.36	0.01;0.01;0.5	0.0	2.41
AB	D8S262	08	3.66	7.36	0.01;0.01;0.9	0.0	2.39
AE	D18S1163	18	7.46	26.85	0.01;0.8;0.99	0.0	2.34
ACD	D2S1788	02	36.11	58.64	0.01;0.01;0.6	0.09	2.31
ADE	D3S1305	03	196.79	219.85	0.01;0.01;0.99	0.0	2.25
ABDE	D3S1764	03	140.67	146.88	0.01;0.01;0.99	0.0	2.18
ABDE	D14S290	14	62.60	60.07	0.01;0.01;0.6	0.0	2.18
BD	ATA52D02	05	---	189.23	0.02;0.02;0.99	0.0	2.16
AE	D3S1309	03	142.21	148.58	0.01;0.01;0.99	0.0	2.11
ACE	D3S3518	03	33.66	59.01	0.01;0.01;0.5	0.01	2.09
ACE	D18S858	18	53.05	81.99	0.02;0.02;0.6	0.0	2.02
E	D15S659	15	44.16	44.34	0.01;0.99;0.99	0.0	2.01
BE	GATA138C05	16	---	81.15	0.01;0.01;0.99	0.0	2.01

Summary of loci considered reportable for one of three reasons: LOD score > 3, perfect segregation in a subpedigree, or replication of a locus in another study. The composite locus on chromosome 15 is needed later for two-locus analyses, as a way to include both dominant loci on 15 together.

Table 3

Chr.	Position	Penetrance	LOD full	Subped	LOD	HLOD	SLINK
2	28.72–31.41 Mbp	0.01;0.01;0.8	2.35	ACD	3.86	3.22	94.7%
3	28.74–38.49 Mbp	0.01;0.01;0.8	2.00	ACE	3.18	2.51	97.4%
3	195.33–198.57 Mbp	0.01;0.01;0.99	1.57	ADE	3.47	2.70	98.2%
14	65.45–70.65 Mbp	0.01;0.01;0.99	1.21	ABDE	3.45	2.85	87.5%
15	43.47–54.36 Mbp	0.01;0.99;0.99	unlinked	E	2.01	2.39	100.0% ^{**}
15	25.85–37.32 Mbp	0.01;0.99;0.99	unlinked	B	2.25	1.50	56.0% ^{**}
15	Not Applicable	0.01;0.6;0.99	unlinked	BE	1.55	ND	60.4%

The SLINK column is the percentile of the observed score among the empirical cumulative distribution of simulated scores for 1000 replicates of a linked pedigree and two-marker analysis, except for the row marked

* ND: not done.

**

For the perfectly segregating locus on 15E, 10.8% of the SLINK single-marker replicates also achieve the highest observed score of 2.01; due to the way SLINK is used, the perfect segregation is not evident if two-marker analysis is used. Similarly, for the locus on chromosome 15 in subpedigree B, the low percentile of 56.0% is due to using two-marker analysis with ungenotyped founders. In the observed data, the locus segregates perfectly in B, but scores below the 2.0 threshold used for Table 2.

Table 4

An example of a penetrance function with two penetrance classes.

First Locus\Second Locus	Zero Trait Alleles	One Trait Allele	Two Trait Alleles
Penetrance Class One			
Zero Trait Alleles	0.01	0.01	0.01
One Trait Allele	0.01	0.01	0.01
Two Trait Alleles	0.8	0.8	0.8
Penetrance Class Two			
Zero Trait Alleles	0.01	0.6	0.99
One Trait Allele	0.01	0.6	0.99
Two Trait Alleles	0.01	0.6	0.99

Table 5

An example of a penetrance function with three penetrance classes.

First Locus\Second Locus	Zero Trait Alleles	One Trait Allele	Two Trait Alleles
Penetrance Class One			
Zero Trait Alleles	0.01	0.01	0.01
One Trait Allele	0.01	0.01	0.01
Two Trait Alleles	0.8	0.8	0.8
Penetrance Class Two			
Zero Trait Alleles	0.01	0.01	0.99
One Trait Allele	0.01	0.01	0.99
Two Trait Alleles	0.01	0.01	0.99
Penetrance Class Three			
Zero Trait Alleles	0.01	0.01	0.99
One Trait Allele	0.01	0.01	0.99
Two Trait Alleles	0.8	0.8	0.99

Results for pairwise analysis of loci. For the chromosome 2p (ACD)/chromosome 15 (BE)/chromosome 2p (ACD) pair, the score is approximately 3.57 plus the score in Supplemental Table 7 for the marker chosen to represent 15. The range of scores 1.12–2.99 in Supplemental Table 7 therefore, leads to the range of two-locus scores 4.69–6.57 in the lower right Table entry.

Table 6

	2p (ACD)	3p (ACE)	3q (ADE)	15 (B)	15 (E)	15 (BE)
2p (ACD)	-	-	-	-	-	-
3p (ACE)	4.75 ACDE	-	-	-	-	-
3q (ADE)	4.10 ACDE	3.99 ACDE	-	-	-	-
15 (B)	5.28 ABCD	4.98 ABDE	4.55 ABDE	-	-	-
15 (E)	5.59 ACDE	3.50 ACE	2.90 ADE	3.42 BE	-	-
15 (BE)	4.69–6.57 full	5.54 ABCE	4.76 ABDE	-	-	-