

## Quantitative trait locus for reading disability on chromosome 6.

by Lon R. Cardon, Shelley D. Smith, David W. Fulker, William J. Kimberling, Bruce F. Pennington and John C. DeFries

Interval mapping of data from two independent samples of sib pairs, at least one member of whom was reading disabled, revealed evidence for a quantitative trait locus (QTL) on chromosome 6. Results obtained from analyses of reading performance from 114 sib pairs genotyped for DNA markers localized the QTL to 6p21.3. Analyses of corresponding data from an independent sample of 50 dizygotic twin pairs provided evidence for linkage to the same region. In combination, the replicate samples yielded a [x.sup.2] value of 16.73 ( $P = 0.0002$ ). Examination of twin and kindred siblings with more extreme deficits in reading performance yielded even stronger evidence for a QTL [ $x_{\text{sup.2}} = 27.35$ ,  $P < 0.00001$ ]. The position of the QTL was narrowly defined with a 100:1 confidence interval to a 2-centimorgan region within the human leukocyte antigen complex.

© COPYRIGHT 1994 American Association for the Advancement of Science. Due to publisher request, Science cannot be reproduced until 360 days after the original publication date.

Reading disability (RD), or dyslexia, is a major social, educational, and mental health problem. In spite of average intelligence and adequate educational opportunities, 5 to 10% of schoolchildren have substantial reading deficits(1). Clear evidence for familial transmission has existed for almost a century, and results of recent twin and family studies have shown a substantial genetic component to the disorder(2), with heritable variation estimated at 50 to 70% (3). Mapping QTLs for RD would facilitate identification of the functional genes that cause the disorder and improve risk estimation and early diagnosis.

Several findings indicating possible linkages for RD have been reported. In a study of nine three-generation families selected for a history of specific RD, we previously obtained evidence for a possible linkage on chromosome 15(4). Recently, evidence of linkage has been reported for markers in the Rh region of chromosome 1(5) and with a translocation between 1p and 2q(6). Further research with our kindreds has yielded evidence for a linkage on chromosome 6 in the human leukocyte antigen (HLA) region, but not for linkage on chromosome 15(5, 7). The present report describes results with more informative markers on chromosome 6 in these kindreds and a replication in an independent sample of dizygotic (DZ) twins.

The HLA region was targeted for this study because of a possible association between dyslexia and autoimmune disorders (8). Results of previous studies have suggested that rates of autoimmune diseases are elevated in relatives of dyslexic probands and that the incidence of dyslexia is increased in relatives of probands having autoimmune illnesses(9). Although the causal basis of the association is unknown, the evidence for association from these independent studies points to the HLA complex as a

candidate region.

Our kindred sibling sample comprises 358 individuals from 19 families who were chosen from a variety of sources, including clinics and private schools specializing in RD. Selection criteria included an extended family history of specific RD, as diagnosed by reading performance at least 2 years below expected grade level and in a pattern consistent with autosomal dominant inheritance (10). The twins sample comprises 50 families drawn from the Colorado twin study of RD (3). The twins range in age from 8 to 20 years (mean, 12.16 years). Twin pairs in which at least one member had a positive school history of reading problems were objectively and systematically selected through cooperating school districts. Individuals were administered a battery of psychometric tests including the Peabody individual achievement tests (PIAT) and the WISC-R intelligence test (11). Subjects with verbal or performance IQ of at least 90 were diagnosed on the basis of a composite discriminant score. Discriminant weights for PIAT reading recognition, reading comprehension, and spelling were computed from an independent sample of RD and control nontwin children in order to produce a continuous measure of RD with known psychometric properties. A comparable measure was constructed from the psychometric data obtained on the kindred sample. We refer to this measure as the discriminant score for reading performance.

One advantage of using DZ twins for linkage analysis is that they provide a perfect control for the effects of age. In the kindreds, which span three generations, subjects are not only of different generations but, within sibships, they are of different ages. Compensation for RD in older subjects renders diagnosis difficult. Another advantage of using twin pairs is a broader sampling of families, which increases the informativeness of the markers. Thus, data from a smaller number of DZ pairs may have more power than a greater number of sibships drawn from a few families with greater age variation.

In the original kindred study, four markers on chromosome

## Quantitative trait locus for reading disability on chromosome 6.

6 were genotyped: BF (properdin factor B, a serum protein), GLO (glyoxylase 1, an erythrocyte enzyme), and pTHH157 and 2C5 (restriction fragment length polymorphisms, RFLPs). They are all located on the short arm of chromosome 6 in bands 6p21.31-p21.1 in the region of the HLA loci. Unfortunately, these markers are not very informative, with heterozygosity values ( $H$ ) < 0.30. Subsequently, we used polymerase chain reaction (PCR) to obtain more informative DNA markers in the same region for the kindred sibships and replicated the genotyping in the independent twin sample. Five markers having  $H$  [greater than or equal to] 0.60 were typed: D6S89, D6S109, D6S105, TNFB, and D6S87 (Table 1). The marker TNFB is closest to BF (separated by 0.8 cM) and is also located within the HLA complex. We used these five markers in a two-point interval mapping procedure (12) to analyze the discriminant scores for RD in both the kindred and the twin samples.

[TABULAR DATA OMITTED]

Our method is based on the sib-pair approach of Haseman and Elston(13), but was extended to accommodate interval mapping(14). Conventional sib-pair analysis (the Haseman-Elston method) involves squaring the difference between the phenotypic scores of a pair of sibs ( $Y$ ), and then regressing  $Y$  onto an estimate of the proportion of alleles that sib pairs share identical by descent [ $\pi$ ] at a marker locus:

$$Y = [\alpha] + [\beta] [\pi] \quad (1)$$

The value for  $P$  tests for variation associated with the marker locus(13). The extension to interval mapping involves the use of a pair of [ $\pi$ ] values for adjacent markers to estimate [ $\pi$ ].sub.q], the proportion of alleles shared identical by descent for the putative QTL located somewhere between the markers (15). The value of [ $\pi$ ].sub.q] depends only on the values of [ $\pi$ ] for the two flanking markers and the assumed location of the QTL. By regressing on a range of [ $\pi$ ].sub.q] values, the QTL may be located at the position that provides the best statistical fit of the model to the data. This method also provides increased statistical power over the conventional sibpair method(12).

In place of the statistical model employed in conventional sib-pair analysis, we used two extensions of the regression model of DeFries and Fulker(16). One method involves regression of the score for one sib onto the score for the second sib, the estimate of [ $\pi$ ], and the product of the second sib's score with the estimate of [ $\pi$ ](17). If we assume only additive gene action, the model is

$$C = [B.sub.0] + [B.sub.1]P + [B.sub.2][\pi] + [B.sub.3]P[\pi]$$

(2)

where  $C$  is the phenotype of the first sib and  $P$  is the phenotype of the second sib. In this model, the test for linkage is provided by [B.sub.3]. The statistical procedure involves double entry of data and a corresponding adjustment of the  $t$  value for [B.sub.3](18). A further refinement is to include the effects of dominance in the model(19), when the regression becomes

$$C = [B.sub.0] + [B.sub.1]P + [B.sub.2]([\pi] - 0.5)$$

$$+ [B.sub.3]P([\pi] - 0.5)$$

$$+ [B.sub.4][\text{abs}([\pi] - 0.5) - 0.25]$$

$$+ [B.sub.5]P[\text{abs}([\pi] - 0.5) - 0.25] \quad (3)$$

The coefficient [B.sub.3] provides the overall test for linkage, and [B.sub.5] detects linkage to a QTL with nonadditive gene action. The multiple regression analysis has better statistical properties than the conventional sib-pair approach, particularly when used with small samples and in the presence of possible outliers. In common with other sib-pair methods, it is unnecessary to make restrictive assumptions concerning an appropriate genetic model that most other linkage methods require. This is important when searching for QTLs in complex traits where the mode of transmission is often unknown.

The second method is applicable to the analysis of samples selected for extreme phenotypes. It is based on the idea that, under linkage with a QTL, co-twins of the selected probands should differentially regress back toward the mean of the unselected population according to the proportion of alleles shared with the proband (16, 17). A general test for linkage may be obtained from the [B.sub.2] coefficient when the following model is fit to the data:

$$C = [B.sub.0] + [B.sub.1]P + [B.sub.2][\pi] \quad (4)$$

With nonadditive gene action, the selection model is

$$C = [B.sub.0] + [B.sub.1]P + [B.sub.2]([\pi] - 0.5)$$

$$+ [B.sub.3][\text{abs}([\pi] - 0.5) - 0.25] \quad (5)$$

in which [B.sub.] provides the test for nonadditive effects. With intense selection (for example, 5% or less of the normal distribution), the selected sample approach offers as much as a 10-fold increase in statistical power to detect a QTL over conventional sib-pair methods(19), and the power is further increased when it is used in the context of interval mapping(20).

## Quantitative trait locus for reading disability on chromosome 6.

Conventional sib-pair analysis of the discriminant measure of reading performance in the original kindred sibships involving the four markers 2C5, BF, pTHH157, and GLO indicated a possible QTL in the BF region(7). Results of an interval mapping analysis (model 3) of the same four loci are shown in Fig. 1, where there is a sharp peak at the BF marker (position 25.1 cM) that is within the HLA complex ( $t = 2.84$ ,  $P = 0.0027$ ). There is no evidence of a QTL at any other location.

Results of interval mapping using the five more-informative markers are shown in Fig. 2A. For the sibling sample, there is a sharp peak located at 6p21.3 between markers D6S105 and TNFB ( $t = 1.75$ ,  $P = 0.04$ ). For the twins, there is a more pronounced peak between the same two markers ( $t = 3.69$ ,  $P = 0.0003$ ), only 0.4 cM away from that of the kindred siblings. There are other secondary peaks that reach levels of statistical significance, but the confidence intervals fail to distinguish them from the major peak.

The significant results in Fig. 2A correspond to broad heritability in our model. For the twin sample, the parameter relating specifically to nonadditive gene action (dominant or recessive) was also highly significant ( $[B.sub.5], t = 2.78$ ;  $P = 0.004$ ). In the kindred sample, this parameter failed to reach significance ( $t = 0.90$ ,  $P = 0.19$ ), probably reflecting the reduced power of the test for nonadditivity over that for total heritability in this sample. The finding in the twins indicates a departure from additivity and suggests that the putative QTL has either a recessive or dominant mode of expression. Dominant transmission was also suggested in our segregation analysis of data from the Colorado family reading study (21).

The evidence for linkage is further enhanced when the results are combined for the independent samples (Fig. 2B)(22). The peak between D6S105 and TNFB provides a  $[x.sup.2]$  value of 16.73 ( $P = 0.0002$ ), the location support interval is extremely narrow, and the remaining peak between TNFB and D6S87 appears to be a ghost image of the main peak because its support interval extends into the position of the primary peak.

Analysis of data from individuals with more extreme deficits in reading performance provides even stronger evidence for a QTL in this region. Individuals in the kindred and twin families having discriminant scores 2 SDs or more below the mean of the unselected population were designated as probands and analyzed with the selected sample regression procedure (model 5). The original markers in the kindred siblings indicated a single peak at the BF marker with a  $t$  value of 3.55 ( $P = 0.0003$ ) (Fig. 3). This peak is located at the same position as in the

unselected analysis, but with a greater level of statistical significance. Analyses of the more informative DNA markers in the kindred and twin samples indicate a QTL in the same region, with maximal peaks between markers D6S105 and TNFB (Fig. 4A) (twin:  $t = 5.12$ ,  $P < 0.00001$ ; kindred:  $t = 1.52$ ,  $P = 0.066$ ). The combined kindred and twin samples reinforce this finding, providing even stronger support for a QTL in this region ( $[x.sup.2] = 27.35$ ,  $P < 0.00001$ ). The peak in this interval is very narrowly defined, with a 1000:1 location support interval contained within the 2-cM distance separating these two markers(23) (Fig. 4B).

Results of the three QTL analyses described in this report are highly consistent (24). Interval analysis of the BF and associated markers in the sibling sample suggested a possible QTL in the HLA region. The analysis of the more informative markers that were subsequently obtained on the same subjects confirmed the initial finding. Although scores from the same subjects were included in the two analyses, a new pair of markers (D6S105 and TNFB) very close to BF also yielded significant results. The third analysis is a true replication, involving data from an independent sample of DZ twin pairs. Thus, the combined results of these three analyses provide compelling evidence for a QTL in the HLA region that influences RD(25).

### REFERENCES AND NOTES

- [1.] B. S. Brown, in *Dyslexia: An Appraisal of Current Knowledge*, A. L. Benton and D. Pearl, Eds. (Oxford Univ. Press, Oxford, 1978), pp. v-vii.
- [2.] J. C. DeFries and J.J. Gillis, in *Nature, Nurture and Psychology*, R. Plomin and G. E. McClearn, Eds. (APA Books, New York, 1993), pp. 121-145; J. C. DeFries, R. K. Olson, B. F. Pennington, S. D. Smith, in *The Reading Brain: The Biological Basis of Dyslexia*, D. D. Duane and D. B. Gray, Eds. (York, Parkton, MD, 1991), pp. 53-88; B. F. Pennington, *J. Child Psychol. Psychiatry* 31, 193 (1990).
- [3.] J. C. DeFries, D. W. Fulker, M. C. LaBuda, *Nature* 329, 537 (1987).
- [4.] S. D. Smith, W. J. Kimberling, B. F. Pennington, H. A. Lubs, *Science* 219, 1345 (1983).
- [5.] M. Rabin et al., *Lancet* 342, 178 (1993).
- [6.] U. Froster, G. Schulte-Korne, J. Hebebrand, H. Remschmidt, *ibid.*, p. 178.
- [7.] S. D. Smith and W. J. Kimberling, *Reading Disabilities: Genetic and Neurological Influences*, B. Pennington, Ed. (Kluwer Academic, Boston, 1991).
- [8.] N. Geschwind and P. O. Behan, *Proc. Natl. Acad. Sci. U.S.A.* 79, 5097 (1982); B. F. Pennington, S. D. Smith, W. J. Kimberling, P. A. Green, M. M. Haith, *Arch. Neurol.* 44, 634 (1987); D. K. Urion, *Ann. Neurol.* 24, 267 (1988); J. W. Gilger, B. F. Pennington, P. Green, S. M. Smith, S. D. Smith, *Neuropsychologia* 30, 209 (1992).
- [9.] K. Hugdahl, B. Synnevag, P. Satz, *Neuropsychologia* 28, 673 (1990); R. G. Lahita,

## Quantitative trait locus for reading disability on chromosome 6.

Psychoneuroendocrinology 13, 1 (1988); O. Hansen, J. Nerup, B. Holbeck, Hereditas 105, 165 (1986). [10.] S. D. Smith, B. F. Pennington, W. J. Kimberling, P. S. Ing, J. Am. Acad. Child Adolesc. Psychiatry 29, 204 (1990). [11.] L. M. Dunn and F. C. Markwardt, Examiner's Manual: Peabody Individual Achievement Test (American Guidance Service, Circle Pines, MN, 1970); D. Wechsler, Manual for the Wechsler Intelligence Scale for Children - Revised (The Psychological Corporation, New York, 1974). [12.] D. W. Fulker and L. R. Cardon, Am. J. Hum. Genet. 54, 1092 (1994). [13.] J. K. Haseman and R. C. Elston, Behav. Genet 2, 3 (1972). [14.] E. S. Lander and D. Botstein, Genetics 121, 185 (1989). [15.] For flanking markers located at positions [L.sub.1] and [L.sub.2] and a putative QTL located at [L.sub.q], denote the recombination fractions between [L.sub.1] and [L.sub.q], [L.sub.q] and [L.sub.2], and [L.sub.1] and [L.sub.2] as  $[\theta]_{sub.1}$ ,  $[\theta]_{sub.2}$ , respectively. The proportions of alleles corresponding to [L.sub.1], and [L.sub.2],  $[\pi]_{sub.1}$  and  $[\pi]_{sub.2}$ , are used to estimate the average number of alleles shared at the putative trait locus:  $[\pi]_{sub.q} = [\alpha] + [\beta]_{sub.1}[\pi]_{sub.1} + [\beta]_{sub.2}[\pi]_{sub.2}$ , with solutions for the regression coefficients given as

[Mathematical Expression Omitted]

and

$$[\alpha] = (1 - [\beta]_{sub.1} - [\beta]_{sub.2})/2.$$

The region between [L.sub.1] and [L.sub.2] is divided into N intervals with  $[\lambda]_{sub.k} = [k \lambda]_{sub.12}/N$ , and  $[\lambda]_{sub.2} = [\lambda]_{sub.12} - [\lambda]_{sub.1}$  for  $k = 0, 1, \dots, N$ , and  $[\theta]_{sub.1}$  and  $[\theta]_{sub.2}$  are determined with Haldane's mapping function  $[\theta] = 0.5[1 - \exp(-2[\lambda])]$ , where  $[\lambda]_{sub.1}$  and  $[\lambda]_{sub.12}$  are map distances corresponding to the recombination fractions. Sib-pair regressions are then fitted for each of the N + 1 intervals (12). [16.] J. C. DeFries and D. W. Fulker, Behav. Genet. 15, 467 (1985); Acta Genet. Med. Gemellol. 37, 205 (1988). [17.] D. W. Fulker et al., Read. Writ. Interdisc. J. 3, 299 (1991). [18.] In unselected samples, one sibling is arbitrarily designated a proband and the other a cosib, and then pairs are double-entered in opposite categories. The procedure does not affect the estimated regression coefficients, but the standard errors of the estimates must be rescaled by the square root of 2 in order to obtain unbiased t statistics (16, 17). [19.] G. Carey and J. Williamson, Am. J. Hum. Genet. 49, 786 (1991). [20.] L. R. Cardon and D. W. Fulker, ibid. 55, 825 (1994). [21.] B. F. Pennington et al., J. Am. Med. Assoc. 266, 1527 (1991). [22.] In large samples t values are asymptotically distributed as normal deviates, and squared deviates,

[Z.sub.2], follow a  $[x]_{sup.2}$  distribution with 1 degree of freedom (df). For the two independent samples,  $[Z]_{sub.1} + [Z]_{sub.2}$  is approximately distributed as  $[x]_{sup.2}$  with 2 df. [23.] When data from selected samples are subjected to multiple regression analysis, information pertaining to both gene action and effect size can also be obtained. With regard to gene action, the relation between cosib means and  $[\pi]$  will differ for dominant and recessive QTLs. If a rare dominant QTL results in lower scores, the average scores of cosibs with a  $[\pi]$  of 0.5 will be intermediate to those with a  $[\pi]$  of 1.0 or 0.0. In contrast, if homozygosity for a rare recessive QTL results in lower performance, cosibs with a  $[\pi]$  of 0.5 or 0.0 will both manifest a similar regression to the mean. Data from the twin sample included in the present study revealed a pattern of increasing cosib means as a function of decreasing  $[\pi]$ , thereby indicating either dominant or additive gene action for the putative QTL. The effect size of a QTL (a) may be defined as half the difference between the phenotypes of two homozygous genotypes, for example, AA and aa. When Eq. 5 is fitted to data from selected sibling pairs,  $[B]_{sub.2}$  provides an estimate of 2a. The average effect size estimated from analyses of the DNA markers in the sibling and twin samples is  $0.86 \pm 0.18$ , suggesting that the putative QTL decreases reading performance by over 1.5 SDs, on average. [24.] Although the gender ratio (males/females) in clinic and referred samples of RD children is often 3:1 or greater, the ratio in the present twin sample is only 1.96:1, and the linkage results obtained from the twin and kindred samples do not differ as a function of gender. When the regression models were extended to include main effects and interactions involving gender (17), no significant gender effects or interactions were obtained in either sample. Because of the multifactorial etiology of dyslexia and the possibility of specific subtypes, it is possible that several major loci contribute to RD. Although the primary objective of this study was to test for a QTL on chromosome 6, we have also genotyped the two samples for a number of additional DNA markers covering most of chromosomes 12 (9 markers from 12p13.3 to 12q24.33) and 15 (20 markers, 15cen to 15qter), and the Rh region of chromosome 1. Interval mapping analyses of these markers yielded no evidence for linkage at any location that replicated in the two samples. We also found no evidence for allelic association with any of our markers, including those within the HLA complex. [25.] Results of a recent association study provide tentative evidence for a QTL in this region that may influence IQ [R. Plomin et al., Behav. Genet 24, 107 (1994)]. Because dyslexia is associated with somewhat lower IQ, this could mean that the QTL we have located contributes to low IQ. Alternatively, it is possible that their marker is associated with verbal ability rather than general cognitive ability. To address this issue, we reanalyzed our twin data using measures of verbal,

## Quantitative trait locus for reading disability on chromosome 6.

performance, and full-scale IQ in place of discriminant reading scores. There was no indication of a QTL for any measure of IQ at any location on chromosome 6. This outcome suggests that the QTL identified in the present study relates to specific RD and not to low IQ. [26.] Support intervals define the boundaries for unit changes in lod scores (logarithm of the likelihood ratio for linkage). Lander and Botstein (14) have shown that lod scores are asymptotically distributed as  $1/2([\log.\text{sub}.10]e)^{[Z.\text{sup}.2]}$ , where Z is a normal deviate. For the independent samples in the present application, the lod transformation is  $1/2([\log.\text{sub}.10]e)^{([t.\text{sub}.1.\text{sup}.2] + [t.\text{sub}.2.\text{sup}.2])}$ . [27.] C. S. Haley and S. A. Knott, *Heredity* 69, 315 (1992). [28.] We thank J. Gillis Light for assistance with data preparation. Supported in part by program project and center grants from the National Institute of Child Health and Human Development (NICHD) (HD-11681 and HD-27802) to J.C.D., National Center for Human Genome, Research grant HG-00085 to L.R.C., NICHD grants HD-18426 and HD-07289 and National Institute of Mental Health (NIMH) grant MH43889 to D.W.F., and by awards from NIMH (MH00419 and MH38820), the March of Dimes (12-135), and the Orton Dyslexia Society to B.F.P.