Am. J. Hum. Genet. 42:498–505, 1988

Efficient Computations in Multilocus Linkage Analysis

G. M. Lathrop* and J.-M. Lalouel*,†

*Howard Hughes Medical Institute and †Department of Human Genetics, University of Utah, Salt Lake City

Summary

This paper describes efficient methods for likelihood calculations and maximum-likelihood estimation in multilocus linkage analysis of reference families and general disease pedigrees, and it documents their performance as implemented in the LINKAGE programs. This information should be of considerable value in determining computing needs for linkage investigations, and in evaluating the merits of alternative algorithms.

Introduction

In recent years, genes responsible for several human diseases, including Huntington disease (Gusella et al. 1982), cystic fibrosis (Wainwright et al. 1985; White et al. 1985*b*), and von Recklinghausen neurofibromatosis (Barker et al. 1987), have been mapped to specific chromosome regions through linkage to polymorphic DNA markers. The localization of a disease gene usually requires the detection of linkage in extended pedigrees or families. Computer programs (Ott 1974; Lathrop and Lalouel 1984) are necessary for the analysis of human linkage data for several reasons: genotypes may be missing for some members of the pedigree; most marker loci are not completely informative; allelic phase is frequently unknown; and penetrance is often incomplete.

After the preliminary detection of linkage, other markers in the region of the disease gene can be used to refine its localization (Lathrop et al. 1988). Mapping of adjacent marker loci in a reference panel of normal families greatly increases the precision of localization (Lathrop et al. 1985). Gene order and recombination rates can be estimated, by multilocus analysis, from linkage information provided by several loci simultaneously, both for construction of the genetic map and for application to disease families (Lathrop et al. 1984, 1985). Computer implementation of multilocus analysis is available in the LINKAGE program package (Lathrop and Lalouel 1984).

Construction of the map and localization of a disease gene impose different requirements on multilocus analysis. Marker loci are most efficiently mapped by typing in reference families chosen for a structure that is highly informative for linkage analysis. Nuclear families with large sibships (and including grandparents for phase information) provide an ideal reference panel (White et al. 1985*a*). The fixed family structure and the availability of DNA from most family members in the reference panel simplify the construction of the map, but to build extensive maps it may be necessary to consider a large number of loci simultaneously. In contrast, disease pedigrees have a highly variable structure and often contain many members with unknown genotypes; fewer loci will usually be examined, however, as only those in the immediate vicinity of the gene are of interest.

Although statistical considerations (Lathrop et al. 1984; Thompson 1984; Lathrop et al. 1985) and practical experience (Drayna and White 1985; Leppert et al. 1986; Lathrop et al. 1988) have shown multilocus analysis to be the most effective method of constructing genetic maps and localizing disease genes, some investigators have argued that the computing times required are prohibitive (Morton et al. 1986). Modified computational algorithms designed

Received July 17, 1987; revision received October 7, 1987.

Address for correspondence and reprints: G. M. Lathrop, Howard Hughes Medical Institute, 603 Wintrobe Building, University of Utah Medical Center, Salt Lake City, UT 84132.

© 1988 by The American Society of Human Genetics. All rights reserved.

0002-9297/88/4203-0013$02.00

to improve computing efficiency have been developed for nuclear pedigrees (nuclear families with grandparents) and implemented in the LINKAGE programs (Lathrop et al. 1986). Other computing techniques have been proposed for multilocus X-linked data (Clayton 1986) and, more recently, for nuclear pedigrees (Lander and Green 1987).

The purpose of this paper is to describe efficient methods of multilocus linkage analysis and to document their performance for reference families and for general disease pedigrees as implemented in the LINKAGE programs. The information provided here should be of value in determining computing needs for linkage investigations and in evaluating the merits claimed for alternative algorithms.

Description of the Programs

Earlier versions of the LINKAGE programs have been described by Lathrop and Lalouel (1984). Likelihoods given phenotypic data on a pedigree are obtained using a modification of a basic recursive algorithm (Elston and Stewart 1971). In addition to codominant marker loci, disease loci with variable penetrance as well as quantitative measurements can be considered. Sex-specific recombination and mutation rates are included as options.

Maximum-likelihood estimates are obtained by numerical maximization with the GEMINI program (Lalouel 1979), which implements a variable metric algorithm that involves an efficient, recursive approximation to second derivatives of the likelihood function (Hessian matrix). The mathematic and numerical properties of this algorithm with finite-length arithmetic, as performed by computers, are well documented and supported by published benchmarks on difficult standard test functions (Lalouel 1979). Several likelihood evaluations are needed at each iteration for a line search and numerical approximations to derivatives.

Several methods have been implemented for optimizing the evaluation of the likelihood for general pedigrees. Recoding alleles reduces the number of genotypes at marker loci without changing the linkage information (Ott 1978; Braverman 1985). A reduced set of possible genotypes for untyped individuals is obtained by eliminating genotypes at each locus that are incompatible with data on other members of the pedigree; the algorithm is similar to that described in Lange and Boehnke (1983). Phase information from each nuclear family can also be used to reduce the number of haplotype combinations that need be considered. When a parental combination of multilocus genotypes is found to have zero probability during the first evaluation of the likelihood, it is eliminated from further calculations.

For the analysis of rare dominant diseases, a further reduction in computing time is often possible by assuming that homozygotes do not occur at the disease locus. Because the homozygote frequency is much smaller than the heterozygote frequency for rare genes, this assumption generally has a negligible effect on the likelihood. Similarly, one can usually assume that the dominant disease gene has been introduced by only one ancestor even when the affected status of founding members of the population is unknown. However, since pedigrees including one or more homozygotes, or heterozygote × heterozygotes matings, are occasionally ascertained, the approximations covering these conditions are options in the LINKAGE programs.

Analysis of the program shows that array accesses account for the largest expenditure of central processing unit (CPU) time. Array access occurs primarily during the phase of the recursion in which genotype probabilities for an individual are calculated conditional on the information from the spouse and their children. Let G_i stand for a multilocus genotype, composed of haplotypes H_{i1} and H_{i2}, for individual i. If individual 1 is a founding member of the pedigree, with spouse 2, the final stage of the recursive calculation can be represented as

$$P(X_1, Y_1|G_1) = P(X_1|G_1) \times \sum_{G_2} P(X_2, Y_2|G_2) \times \left[\prod_j \sum_{G_j} P(X_j, Y_j|G_j)P(G_j|G_1, G_2)\right], \quad (1)$$

where X_i represents the multilocus phenotype of individual i and Y_i represents the multilocus phenotypes of i's relatives in parts of the pedigree that have already been traversed. The product is taken over all children of individuals 1 and 2. For further details see Elston and Stewart (1971) and Cannings et al. (1978).

Certain simplifications can be introduced to reduce the number of array accesses needed for the calculation of equation (1); for example, haplotype combinations in the children can be grouped by recombination class, i.e., the pattern of recombination events

between adjacent loci. All members of the same recombination class have equal probability. For given parental genotypes, each recombination class can contain as many as four different haplotype combinations (denoted $H_{11}H_{21}$, $H_{11}H_{22}$, $H_{12}H_{21}$, and $H_{12}H_{22}$). If the parental phase is changed, the same haplotype combinations are found in a different recombination class. Hence, for parental genotypes containing the same alleles but having different phase, the terms $P(X_j, Y_j|H_{11}, H_{21}) + P(X_j, Y_j|H_{11}, H_{22}) + P(X_j, Y_j|H_{11}, H_{22}) + P(X_j Y_j|H_2, H_{12})$ recur together in the inner summation in equation (1), where they are weighted by segregation probabilities depending on the phase combination. Performing this summation prior to the application of equation (1) substantially increases the efficiency of the calculation.

The last step of the calculation is the formation of the sum $P(G_1|X_1, Y_1)$. The terms in this sum are posterior probabilities of the genotypes for individual 1, conditional on all the phenotype information in the pedigree. Typically, only a few of these probabilities contribute significantly to the total likelihood. Terms that make a small contribution to the likelihood can be identified by their magnitude relative to the most probable genotype. Once they are identified, these terms can be ignored in subsequent evaluations if the parameter values are not greatly changed, e.g., when calculating derivatives of the likelihood function by finite difference. As the corresponding conditional probabilities for the offspring and other relatives in equation (1) need not be calculated for a term that is ignored in the likelihood, computing times for large numbers of loci may be greatly reduced through this approximation. As the relative sizes of the posterior genotype probabilities may change considerably as a function of the recombination rates, this approximation cannot be used in a line search.

In reference families with codominant loci, further simplifications of the likelihood calculations are impossible. The grandparents provide information on the phase in the parents; as this is independent of the recombination rates, information from grandparents can be used just once to eliminate possible haplotypes in the parents. The likelihood is

$$L = \sum_{G_1,G_2} P(G_1|H_{11},H_{12})P(G_2|H_{21},H_{22}) \times \prod_j \sum_{H_{j1},H_{j2}} P(H_{j1}|H_{11},H_{12})\, P(H_{j2}|H_{21},H_{22}) \,, \quad (2)$$

where the inner sum is taken over all haplotypes H_{ji} that are compatible with the multilocus phenotype of offspring j and can be formed with or without recombination from the parental haplotypes H_{i1} and H_{i2}. The outer sum in equation (2) is taken over all haplotypes compatible with the maternal and paternal phenotypes (accounting also for the grandparental information). Since each combination of possible parental genotypes is included in the sum, the segregation probabilities $P(H_{k1}|H_{11},H_{12})$ and $P(H_{k2}|H_{21},H_{22})$ reappear in several different terms. Therefore, we obtain these values prior to calculating the summations and products in equation (2). The probabilities of the parental genotypes, conditional on the haplotypes, are either 1 or 0 if all genotypes are known.

Applying transformation and factorization rules to simplify the likelihood calculations, as described by Lathrop et al. (1986), greatly improves the computing efficiency for nuclear pedigrees. These rules allow replacement of data on each family with modified genotypes on a series of new families having the same size and structure as the original. The new families contain fewer loci than the original, but, with appropriate corrections for the recombination rates in adjacent chromosome segments, the total likelihood is the same for the transformed and the original data.

Benchmark Calculations

Benchmark calculations were performed using two sets of reference data described below (available on request). To evaluate the programs in different settings, the calculations were repeated on three computers: (1) a VAX 8650 with 80 Mbytes of memory operating under VMS 4.5 with the DEC PASCAL Compiler V3.5, (2) a Micro-VaxII with 9 Mbytes of memory also under VMS 4.5 with the DEC PASCAL Compiler V3.5, and (3) an IBM-AT at 6 Mhz with 640 Kbytes of memory, TURBO PASCAL V3.01A (non-8087 version). All calculations were performed in single precision.

Reference Families

As an example of computation for the mapping of genetic loci in a reference panel, we considered 50 typical markers from a data base of genotypes characterized in 30 families with large sibships (416 individuals); 28 of the families are part of the Centre d'Etude du Polymorphisme Humain (CEPH) reference panel. Some genotypic data were missing from

Table 1

CPU Times and Number of Function Evaluations Required for Convergence to the Maximum-Likelihood Estimates of Recombination Rates, on Three Different Computers, for Linkage Analysis of Markers in 30 Reference Families

No. of Loci	CPU Time for a Single Likelihood Evaluation (S): VAX 8650 Without Rules	VAX 8650 With Rules	Micro-VaxII	IBM-AT	No. of Evaluations to Convergence: Function Change Criterion	Normalized Gradient Criterion
5	0.05	0.05	0.45	7.52	30[a]	71
10	0.60	0.20	1.64	37.40	271	271
15	4.17	0.43	3.48	83.10	536	583
20	68.78	0.56	4.60	110.51	584[b]	1074
50	ND	6.34	54.48	ND	3463	4262

Note.—Convergence was judged either by reaching a tolerance on the normalized gradient as specified in the GEMINI program or by successive function evaluations with a likelihood difference $<10^{-2}$. The calculations were made after the application of transformation and factorization rules. CPU times on the VAX 8650 for the original data without the application of rules are included for the purposes of comparison. ND = Not done because of memory limitations.

[a] Stopped prematurely at 11.70 units from the maximum (2 ln likelihood scale).

[b] Stopped prematurely at 20.33 units from the maximum (2 ln likelihood scale).

one or both parents in 12 of the families. Genotypes were available for at least one grandparent in 26 of the families.

Estimates of the recombination rates between adjacent loci were obtained after application of the factorization and transformation rules to subsets of 5, 10, 15, 20, and 50 loci on three different computers. Although it is unlikely that 50-locus analysis will be required in practical applications, we have included these results for illustrative purposes. CPU times for a single-likelihood calculation, and the number of function evaluations needed for convergence to the maximum-likelihood solution using the GEMINI program, are shown in table 1. The computing time for a single likelihood evaluation is <1 on the VAX 8650, and the total analysis for a fixed order takes <10 min with 20 loci. Although the IBM-AT is almost 200 times slower than the VAX-8650 in the configurations we used, analysis of as many as 20 loci is possible. The approximation to the likelihood for second-derivative calculations was not necessary, because the exact function evaluations were sufficiently rapid.

To investigate the effects of the factorization and transformation rules, we calculated the distribution of the number of informative loci (heterozygotes) in parents by family in the 50-locus problem. Because the loci in each parent were counted separately, the maximum number is 100. In the original data, the mean number of informative loci in the 30 families was 42 (range 11–55). After applying the rules, we obtained 207 families with a mean number of 7.8 informative loci (range 2–28). The distributions of informative loci before and after transformation are shown in figure 1. To show the effects of this reduction on the likelihood calculations, the data analysis

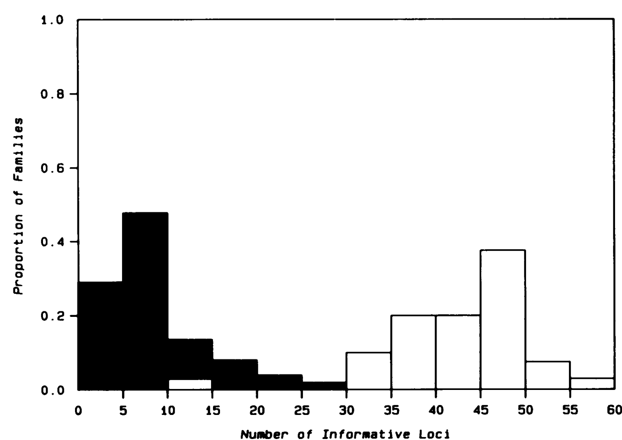

Figure 1 The distribution of number of informative loci per family in 30 reference nuclear pedigrees before (white bars) and after (black bars) transformation and factorization.

was repeated on the VAX 8650, using the same programs but without transformation and factorization. As shown in table 1, the calculations for the 20-locus problem are more than 100 times slower without the rules than with them.

We have adopted very stringent criteria for judging convergence to a maximum-likelihood solution. As discussed by Lalouel (1979), convergence in GEMINI is based on tolerance on a normalized gradient. As shown in table 1, a large number of likelihood evaluations may be needed to achieve convergence with that criterion. This can be reduced if convergence is judged by a small change in the likelihood function between successive iterations. We have illustrated this in table 1 by reporting the number of function evaluations needed to achieve a difference of $\leq 10^{-2}$ in two successive function evaluations. In two of the six examples, however, this stopping rule resulted in premature termination far from the maximum-likelihood solution.

The number of function evaluations needed to achieve convergence depends on the values chosen for the recombination rates at the initial iteration. In the examples given here, the initial recombination rates between adjacent loci were all 1. Other methods of choosing the initial values, e.g., using the estimates from two-locus analysis, may reduce considerably the total computing time.

A Disease Pedigree

To test the effectiveness of the LINKAGE programs for general pedigrees, we made linkage calculations with the 46-member disease pedigree shown in figure 2. We assumed both a rare dominant with incomplete penetrance and no sporadics. The gene frequency and the penetrance were assumed to be known; therefore, only recombination rates were estimated. Five marker loci linked to the disease were included for analysis; .1 was chosen for the initial recombination rates between adjacent loci. Both the pedigree and the distribution of known and unknown genotypes were similar to those encountered in linkage studies in our laboratory.

Computing times for the calculation of the likelihood with the disease locus, and for one to five markers, are shown in table 2. On the VAX-8650 the calculation time for all six loci is ~48 s/likelihood calculation if disease homozygotes and heterozygote

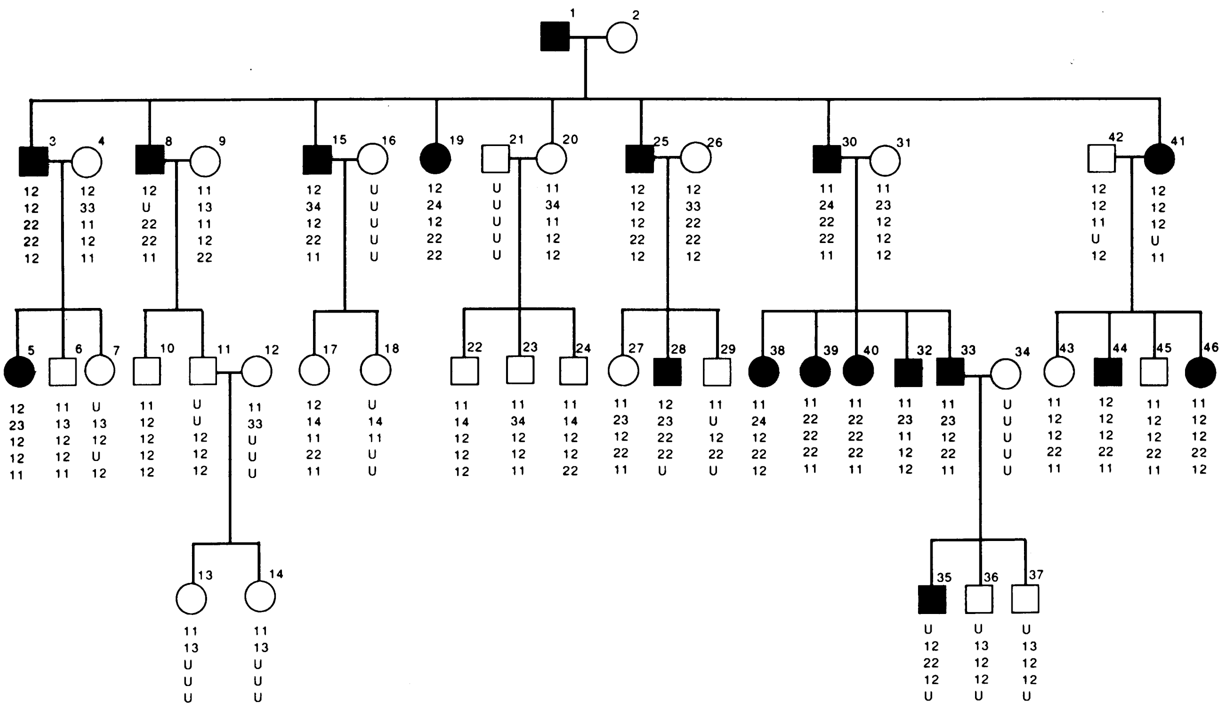

Figure 2 A dominant disease pedigree and five marker loci used for benchmarks in general pedigrees. Affected members of the pedigree are indicated by dark shading. U = Unknown genotypes.

Table 2

CPU Times and Number of Function Evaluations Needed for Convergence to the Maximum-Likelihood Recombination Estimates for Data on the Extended Disease Pedigree Shown in Figure 2

	CPU Time for a Single Likelihood Evaluation (S)				
	VAX 8650				
No. of Loci	Without Approximation	With Approximation	Micro-VaxII	IBM-AT	No. of Evaluations to Convergence
2	0.40	0.11	0.65	1.75	45
3	0.57	0.18	1.36	5.93	41
4	2.86	0.65	4.52	33.51	39
5	29.40	4.91	34.56	413.00	93
6	326.07	48.85	337.04	ND	127

NOTE.—Calculations were performed excluding disease homozygotes and heterozygote × heterozygote matings. CPU times for the calculations without this approximation on the VAX 8650 are included for comparison. Convergence was judged using the tolerance on the normalized gradient as contained in the GEMINI program. ND = Not done because of memory limitations.

[a] The six-locus calculation was not possible.

× heterozygote matings are eliminated. Without this approximation, the time per likelihood calculation may increase substantially depending on the number of marker loci considered (table 2). The likelihood values with and without the approximation are identical to the second decimal place in all the calculations. With this pedigree, calculations with as many as five loci were possible within the memory limits of the IBM-AT.

To decrease the computing time required to meet the GEMINI convergence criteria in the six-locus calculation, we used the approximations to the likelihood for calculating second derivatives, as discussed above. Individual 1 was chosen for the final summation. After all the pedigree information is accounted for, there are 144 possible genotypes, or haplotype combinations, for this individual. The posterior probabilities of these genotypes are equal when all the recombination rates are .5; otherwise, they differ.

To approximate the derivatives of the likelihood function, we eliminated all genotype possibilities having a posterior probability <1/100 of the most probable genotype for the recombination rates at the corresponding iteration. With this approximation, the number of haplotype combinations considered for individual 1 ranged from two to 42 during the gradient calculations. The final solutions with and without the approximation were identical to the second decimal place, but the total calculation time was reduced by >25% (6,540 s vs. 4,800 s on the VAX-8650).

Discussion

The initial versions of the LINKAGE programs were primarily designed to investigate the scientific merits of multilocus analysis in human genetics. Methods of data transformation to simplify the likelihood calculations for nuclear pedigrees make the analysis of large numbers of loci possible for the construction of human gene maps (Lathrop et al. 1986). The proven usefulness of multilocus methods has led to the consideration of other techniques to improve computational efficiency (Clayton 1986; Lander and Green 1987). As shown here, extensive multilocus analysis is possible on a wide variety of computers through optimization of the original LINKAGE algorithms. With these optimizations, computing times should not be a major limitation to the application of multilocus methods. Other modifications of the algorithm are presently under investigation; these are expected to give further improvements in performance, especially for disease pedigrees.

In the LINKAGE programs, maximum-likelihood estimates are obtained via numerical optimization with the GEMINI program. One advantage of this approach over other techniques for numerical maximization,

such as the EM-algorithm implemented by Lander and Green (1987), is the ease with which it can be adapted to situations in which parameters other than recombination rates are estimated. Thus, new penetrance functions, quantitative phenotypes with interactions or residual familial correlations, or new models of sex effect on recombination can be introduced rapidly with a minimum of alterations to the program.

The number of function evaluations required to reach a solution critically depends on the nature of the likelihood surface being examined and the criteria used to judge convergence. The former is generally unknown; because our benchmarks were performed on an arbitrarily selected set of loci, actual performance on a particular application may be significantly better or worse than in the examples reported here.

The choice of convergence criteria is a difficult issue. Because computers perform calculations with a fixed number of significant digits, analytical results alone are insufficient to establish the convergence properties of a numerical algorithm. The optimal properties of the variable metric methods implemented in GEMINI are supported by extensive numerical analysis (see references cited in Lalouel 1979).

Assessing convergence, i.e., implementing stopping rules for the iterative process, is more difficult than is usually appreciated. The simple consideration of differences in function values between two successive iterations can reduce the time required to obtain a solution, as illustrated in table 1. However, the risk of premature termination in flat areas of the likelihood surface is an important factor when this criterion is used. In GEMINI the convergence criteria are based on the magnitude of the gradient appropriately weighted by the curvature of the likelihood function, which gives the true numerical momentum of the iterative process. While the law of diminishing returns could be invoked at some point, the gain in certainty usually outweighs other considerations even at later iterations.

Multilocus linkage analysis provides a plenitude of possible statistical tests for application to a particular set of data. Elsewhere, we have presented strategies for the selection of appropriate tests to efficiently construct a genetic map of marker loci and to localize disease genes (Lalouel et al. 1986). Efficient computation algorithms will permit the adoption of such strategies as routine tools for linkage analysis.

Acknowledgments

We thank R. White for helpful discussions and encouragement. J.-M.L. is an investigator for the Howard Hughes Medical Institute.

References

Barker, D., E. Wright, K. Nguyen, L. Cannon, P. Fain, D. Goldgar, D. T. Bishop, J. Carey, B. Baty, J. Kivlin, H. Willard, J. S. Waye, G. Grieg, L. Leinwand, Y. Nakamura, P. O'Connell, M. Leppert, J. M. Lalouel, R. White, and M. Skolnick. 1987. Gene for von Recklinghausen neurofibromatosis is in the pericentromeric region of chromosome 17. Science **236**:1100–1102.

Braverman, M. S. 1985. An algorithm to improve the computational efficiency of genetic linkage analysis. Comp. Biomed. Res. **18**:24–36.

Cannings, C., E. A. Thompson, and M. H. Skolnick. 1978. Probability functions on complex pedigrees. Adv. Appl. Probability **10**:26–61.

Clayton, J. 1986. A multipoint linkage analysis program for X-linked disorders, with the example of Duchenne muscular dystrophy and seven DNA probes. Hum. Genet. 73:68–72.

Drayna, D., and R. White. 1985. The genetic linkage map of the human X chromosome. Science 230:753–758.

Elston, R. C., and J. Stewart. 1971. A general model for the analysis of pedigree data. Hum. Hered. **21**:523–542.

Gusella, J. F., N. S. Wexler, P. M. Conneally, S. L. Naylor, M. A. Anderson, R. E. Tanzi, P. C. Watkins, K. Ottima, M. R. Wallace, A. Y. Sakaguchi, A. B. Young, I. Shoulson, E. Bonilla, and J. B. Martin. 1982. A polymorphic DNA marker genetically linked to Huntington's disease. Nature 306:234–238.

Lalouel, J. M. 1979. GEMINI: a computer program for optimization of general non-linear functions. Tech. Rep. 14. Department of Medical Biophysics and Computing, University of Utah, Salt Lake City.

Lalouel, J.-M., G. M. Lathrop, and R. White. 1986. Construction of human genetic linkage maps. II. Methodological issues. Cold Spring Harbor Symp. Quant. Biol. 51:39–48.

Lander, E. S., and P. Green. 1987. Construction of multilocus genetic linkage maps in humans. Proc. Natl. Acad. Sci. USA **84**:2363–2367.

Lange, K., and M. Boehnke. 1983. Extensions to pedigree analysis. V. Optimal calculation of Mendelian likelihoods. Hum. Hered. 33:291–301.

Lathrop, G. M., M. Farrall, P. O'Connell, B. Wainwright, M. Leppert, Y. Nakamura, N. Lench, H. Kruyer, M. Dean, M. Park, G. Vande Woude, J.-M. Lalouel, R. Williamson, and R. White. 1988. Refined linkage map of

chromosome 7 in the region of the cystic fibrosis gene. Am. J. Hum. Genet. **42**:38–44.

Lathrop, G. M., and J.-M. Lalouel. 1984. Easy calculations of lod scores and genetic risks on small computers. Am. J. Hum. Genet. **36**:460–465.

Lathrop, G. M., J.-M. Lalouel, C. Julier, and J. Ott. 1984. Strategies for multilocus linkage analysis in humans. Proc. Natl. Acad. Sci. USA **81**:3443–3446.

———. 1985. Multilocus linkage analysis in humans: detection of linkage and estimation of recombination. Am. J. Hum. Genet. **37**:482–498.

Lathrop, G. M., J.-M. Lalouel, and R. L. White. 1986. Calculation of human linkage maps: likelihood calculations for multilocus linkage analysis. Genet. Epidemiol. **3**:39–52.

Leppert, M., W. Cavanee, P. Callahan, T. Holm, P. O'Connell, K. Thompson, G. M. Lathrop, J.-M. Lalouel, and R. L. White. 1986. A primary genetic map of chromosome 13. Am. J. Hum. Genet. **39**:425–437.

Morton, N. E., C. J. MacLean, R. Lew, and S. Yee. 1986. Multipoint linkage analysis. Am. J. Hum. Genet. **38**: 868–883.

Ott, J. 1974. Estimation of the recombination fraction in human pedigrees: efficient computation of the likelihood for human linkage studies. Am. J. Hum. Genet. **26**:588–597.

———. 1978. A simple scheme for the analysis of HLA linkage in pedigrees. Ann. Hum. Genet. **42**:255–257.

Thompson, E. A. 1984. Information gain in joint linkage analysis. IMA J. Math. Appl. Med. Biol. **1**:31–49.

Wainwright, B. J., P. J. Scambler, J. Schmidtke, E. A. Watson, H. Y. Law, M. Farrall, H. J. Cooke, H. Eiberg, and R. Williamson. 1985. Localization of cystic fibrosis locus to human chromosome 7cen-q22. Nature **318**:384–385.

White, R., M. Leppert, T. Bishop, D. Barker, J. Berkowitz, C. Brown, P. Callahan, T. Holm, and L. Jerominski. 1985*a*. Construction of linkage maps with DNA markers for human chromosomes. Nature **313**:382–384.

White, R., S. Woodward, M. Leppert, P. O'Connell, Y. Nakamura, M. Hoff, J. Herbst, J.-M. Lalouel, M. Dean, and G. Vande Woude. 1985*b*. A closely linked genetic marker for cystic fibrosis. Nature **318**:382–384.