CENTROMERIC LINKAGE IN MAN

A thesis submitted by

Gilbert B. Côté, B.A., B.Sc., M.Sc.

for the degree of Ph.D.

in the

Faculty of Medicine and Dentistry

of the University of Birmingham.

Human Genetics
Department of Social Medicine

December 1974

SUMMARY

The aim of this work is to provide geneticists with appropriate statistical methods and computer programmes for the analysis of human pedigree data in view of mapping genes on the human chromosomes, and discovering the origin of chromosomal abnormalities such as the autosomal trisomies, the 47,XXY Klinefelter's syndrome and the 46,XX men syndrome.

J.H. Edwards' marker algebra is presented in detail as used in his computer programme (MARK III) that analyses linkage with Morton's lod method for normal diploids. The programme is also described with all its specifications.

The cytological mechanisms leading to autosomal trisomy are described to show that the proportion of trisomics carrying three alleles from three of their grandparents is bound to be greater than zero for any locus anywhere on a trisomic chromosome. The use of A.W.F. Edwards' method of support is then demonstrated on various sets of data to definitely exclude the ABO, MN, P, Jk, Gc and Lp loci from chromosome no. 21, and the theory is extended to show that about 40% of 47,XXY men receive an extra X from their fathers and 60% from their mothers, and that in general 46,XX men are more likely to arise from 47,XXY zygotes that lose their Y chromosomes than by an interchange between the X and Y chromosomes of their fathers.

ACKNOWLEDGEMENTS

I wish to express my gratitude to

Professor J.H. Edwards, my supervisor and head of department, for his guidance and most stimulating example,

Dr. B. Chown for all his data and quick, patient answers to my countless questioning letters,

Miss Marion Lewis for being so good at coping with my capricious demands, and for housing and feeding me so well while I was in Winnipeg,

Drs. R.R. Race and Ruth Sanger for kindly making all their blood group data available,

Mrs. L. Irmeli Vale for teaching me how to type this thesis properly, and for typing part of it herself,

Dr. Chitralekha Verma for comprehensively standing my noisy presence in her office for more than two years,

All members of the Infant Development Unit for being such good friends,

The National Research Council of Canada for a Postgraduate Scholarship

and, above all,

My parents for their moral support which I felt over the seas.

TABLE OF CONTENTS

VOLUME I

		Page
I	Chromosome mapping in man	1
II	Methods of analysis	7
	A. The Method of Support	8
	B. Computerization	11
	1. Bit shifting	11
	2. W-arrays	23
	3. Marker algebra	30
	4. Handling of pedigree data	39
III	Linkage in diploids	45
	A. Lods	45
	B. MARK III	48
	1. Historical	48
	2. General features	49
	3. Input	51
	a) The dictionary file	51
	b) Pedigree data	54
	c) Control cards	55
	4. Analysis and output	63
	C. Handling of linkage results	68
	1. Storage and graphical display	68
	2. Accumulation	71
	D. Application	72

		Page
IV	Linkage in autosomal trisomies	73
	A. Model for non-disjunction	74
	B. Literature review	82
	C. Probability models	86
	1. Assumptions	86
	2. Codominant systems	88
	3. Dominant-recessive systems	89
	D. Computerization	93
	E. Data on trisomy-21	95
	F. Results and discussion	100
V	Sex chromosome abnormalities	114
	A. XXY Klinefelter's syndrome	115
	B. The origin of 46,XX males	118
VI	Conclusion	122
Bibliography		123

VOLUME II

APPENDICES

I Listings of computer programmes

II Dictionary file

III Data

IV Results

I CHROMOSOME MAPPING IN MAN

Two loci are said to be closely linked when they do not segregate independently between generations. This happens when they are so near to each other on a chromosome that the space between them is not always the site of at least one cross-over which would cause them to recombine. Recombination fractions between loci can be defined, or estimated, from family data, and used to assign loci to particular chromosomes when one of the loci is directly visible by its staining properties on chromosome preparations. At present, most loci thus recognizable on human chromosomes are situated at the centromeres, hence the term 'centromeric linkage' applied to pedigree studies aimed at mapping specific chromosomes. The current accumulation of such data is so rapid and massive that efficient analysis makes computer processing almost indispensable. Part of the work presented in this thesis has therefore consisted in a contribution to the adapting of a computer programme written for this purpose, and its installation in Canada (Winnipeg) and the Netherlands (Leiden) where it is now fully operational and used for routine analyses. Earlier versions of the programme are also operational in Denmark (Aarhus) and the U.S.A. (New York).

In practice, however, finding informative families depends on good fortune since breeding experiments are incompatible with human rights and standards, and would also be prohibitively expensive. Consequently, other methods have to be devised in laboratory to try and complete man's chromosome map. Certainly the most productive of these has been up to now the hybridization of human and animal cells. The useful hybrids

start off with both chromosome complements in their nuclei but rapidly show a tendency to lose human chromosomes every time they undergo mitosis. It follows that if an exclusively human phenotype constantly disappears when, and only when, a certain human chromosome also does, the locus responsible for the phenotype must be on that chromosome. The process can be accelerated by introducing into the culture medium an anti-metabolite acting against the human product, or the procedure reversed by growing the hybrids in a non-permissive medium where the human product is essential for survival. In the latter case, the locus studied must be on the human chromosome not lost by the surviving hybrids.

These techniques were first worked out for the thymidine kinase locus (:TK) by Weiss and Green in 1967, and later used successfully by Migeon and Miller (1968) to assign this locus to a human chromosome of group E (:chromosome 16, 17 or 18). The advent of fluorescence staining techniques for the specific banding of chromosomes allowed Miller et al. (1971) and Ruddle and Chen (1972) to identify the chromosome as no. 17 while the occurrence of a spontaneous translocation to a mouse chromosome provided Boone et al.(1972) with good evidence that the TK locus was on the long arm of no. 17. Finally, McDougall et al. (1973) assigned the locus precisely to the proximal portion of band 17q22 by using the same techniques after having infected the cells with adenovirus 12 which has the particular property of inducing various breaks in that region of the translocated chromosome. Recently, the same band (17q21-22) was shown by the same methods to also carry the locus for galactokinase (Elsevier et al., 1974).

Now that several loci have been placed with certainty on specific human chromosomes (Table 1), gene mapping can be confirmed more rapidly since

it becomes possible to prove the presence or absence of a chromosome segment by determining the concomittant presence or absence of the loci it carries.

Situations equivalent to those observed in hybrid cells occur spontaneously in human beings born with chromosomal deficiencies and duplications, but they rarely seem to yield positive results as far as mapping of autosomes is concerned since only small imbalances are compatible with survival. In fact, the only informative case used to make an assignment later confirmed by other methods is that of a deletion of the short arm of chromosome 2 which was shown to carry the locus for red cell acid phosphatase (Ferguson-Smith et al., 1973). No positive information has ever been given by any of the autosomal deletions known since the first cases were reported in the literature (de Grouchy et al., 1963; Thieffry et al., 1963, Lawler, 1964). Usually, because of chance and the distribution of allele proportions, deleted segments can only be shown not to carry the loci for which affected individuals are heterozygous. Race and Sanger (1968, 1974 in press) give a long list of such exclusions.

Translocations ascertained through defective children are usually unbalanced, although the relative sizes of the deficiency and duplication was generally impossible to assess with any degree of accuracy by earlier techniques (Edwards et al., 1962). Chromosome banding can now clarify such cases.

The chance of a trisomic individual giving any clear positive evidence is also slim, but it has always been obvious that populations of trisomics would show phenotypic distributions different from those of normal

populations if the loci responsible for these phenotypes were on the trisomic chromosomes. Unfortunately, the many population studies carried out so far have remained fruitless for various reasons, including bad luck and weak statistical methods. A definite methodological improvement can be achieved by looking at the problem from a different viewpoint, once it is realized that the proportion of trisomics carrying three alleles from three of their grandparents is bound to be greater than zero for any locus anywhere on a trisomic chromosome. This approach, fully detailed in chapter IV, thus consists in finding linkage between loci and centromeres, and also accounts for the title of this thesis.

It is hoped that the statistical tools herein presented will facilitate future mapping studies. The human chromosomes are already sign-posted with identified sites, only six years after the first autosomal assignment (Donahue et al., 1968), but most discoveries lie ahead if one is to reach the degree of completeness attained for the maps of organisms like the fruit fly and the mouse.

Table 1. List of human loci assigned with certainty to specific autosomes, with Nature of test: A : Agglutination
E : Electrophoresis
H : In situ DNA-RNA hybridization
C : Clinical diagnosis
and Method of assignment:
S : Somatic cell hybridization
F : Family studies
This is a summary of the 2nd international workshop on human gene mapping held in The Netherlands in 1974 (in press).

Chromosome number		Nature of test	Method of assignment
1	Adenylate kinase -2 (AK2)	E	S
	Duffy blood group (Fy)	A	F
	Elliptocytosis - 1	C	F
	Fumarate hydratase (FH)	E	S
	Guanylate kinase (GuK)	E	S
	Pancreatic amylase (Amy-2)	E	F
	Peptidase C	E	S
	Phosphoglucomutase - 1 (PGM_1)	E	F,S
	Phosphopyruvate hydratase (PPH)	E	F,S
	Rh blood group	A	F
	Salivary amylase (Amy-1)	E	F
	Zonular pulverulent cataract	C	F
	5S Ribonucleic acid (5S-RNA)	H	
	6-Phosphogluconate dehydrogenase (6PGD)	E	F,S
2	Red cell acid phosphatase (ACP1)	E	F
	Malate dehydrogenase (MDH1)	E	S
	Isocitrate dehydrogenase (IDH1)	E	S
6	Chido blood group	A	F
	Histocompatibility loci (HL-A)	A	S,F
	Phosphoglucomutase - 3 (PGM_3)	E	S
	Soluble malic enzyme - 1	E	S
	Superoxide dismutase - 2 (SOD_2)	E	S
7	SV40 T antigen	A	
	Mannose phosphate isomerase (MPI)	E	S
10	Glutamate oxaloacetate transaminase (GOT)	E	S
11	Esterase - A4	E	S
	Lactate dehydrogenase - A (LDH-A)	E	S
12	Lactate dehydrogenase B (LDH-B)	E	S
	Peptidase B	E	S
	Triose phosphate isomerase	E	S
13	Ribosomal RNA	H	
14	Nucleoside phosphorylase (NP)	E	S
	Ribosomal RNA	H	

cont'd/..

15	Ribosomal RNA	H	
16	Adenine phosphoribosyl transferase (APRT)	E	S
	Haptoglobin alpha (Hpα)	E	F
	Lecithin cholesterol acetyltransferase (LCAT)	C	F
17	Thymidine kinase (TK)	E	S
	Galactokinase (GaK)	E	S
18	Peptidase A	E	S,F
19	Phosphohexose isomerase (PHI)	E	S
	Polio sensitivity	C	S
20	Adenosine deaminase (ADA)	E	S
21	Anti-viral protein (AVP)		S
	Superoxide dismutase - 1 (SOD_1)	E	S
	Ribosomal RNA	H	
22	Ribosomal RNA	H	

II METHODS OF ANALYSIS

Statistical methods based on large sample theory and tests of significance are too often inappropriate in human genetics where data are often scarce and workers have a tendency to misinterpret the answers given by such tests.

Odds, on the contrary, are easily understood by anyone and imply the clear formulation of rival hypotheses. Although the odds concept is simple, its handling generally requires the calculation of logarithms and the use of computers, which is enough to prevent a majority of physicians and biologists from using it. But in simple cases where the results of such calculations have already been tabulated, analyses are reduced to quick summations easy to obtain and difficult to misinterpret. Such tables are now widely used in linkage investigations and A.W.F. Edwards' recent systematization (1972) of the approach into the Method of Support has facilitated its study by the non-initiated. Computerization of complex cases has also been greatly simplified by the set of computer programmes written by J.H. Edwards with the use of his Marker Algebra (Edwards, 1972).

The present chapter will describe the method of support and the marker algebra used extensively in this study.

A. THE METHOD OF SUPPORT

Faced with more than a single hypothesis to explain the occurrence of a set of data, this method aims at measuring the support to be given to the various hypotheses by comparing their likelihoods, the highest likelihood deserving highest support.

The likelihood $L(H|D)$ of hypothesis H, given data D, is defined as the probability $P(D|H)$ of observing data D, given hypothesis H. $L(H|D)$ and $P(D|H)$ are thus described by the same equation, or probability model, but whereas P is a function of the data for a fixed hypothesis, the likelihood L is regarded as a function of the hypothesis, given fixed data. The various hypotheses to be considered are obtained by varying one or more of the parameters of the probability model. The data D are said to support hypothesis H_1 better than H_2 if $L(H_1|D)$ is greater than $L(H_2|D)$, or in other words, if the likelihood ratio $L(H_1|D)/L(H_2|D)$ is greater than 1. This is the odds ratio, otherwise written as $L(H_1,H_2|D)$.

Support is defined as the natural logarithm of the likelihood ratio. It is obviously positive or negative according to whether the ratio is greater or smaller than 1. A difference of support of more than 2 between rival hypotheses is fairly good grounds to question the validity of the less supported one, but does not constitute an absolute criterion of rejection. In large samples, it corresponds to the support given to the mean against a value two standard deviations away from this mean:

$$\ln \frac{(2\pi)^{-\frac{1}{2}}}{(2\pi e^{2^2})^{-\frac{1}{2}}} = 2.$$

This is the value used as a limit for the rejection or acceptance of

hypotheses in two-tailed significance tests (at the 0.05 level) on the mean of a Normally distributed variable.

The method of support can be used on very little data to describe the small amount of information gained nevertheless about rival hypotheses, none of which can be rejected. The gradual accumulation of evidence is facilitated by the fact that support pertaining to the same hypotheses is additive over various sets of data. Similarly, prior and experimental supports are simply added together to give posterior support.

In addition to the word 'support', A.W.F. Edwards (1972) has introduced a new nomenclature for functions used, and statistics obtained specifically by this method. It is adopted in the following chapters and consists of the following terms:

Support function: The natural logarithm of the likelihood function.

Evaluate: A shorter name for 'maximum likelihood estimate'. It is the best supported value of a parameter and is used when a point estimate is required.

Formation: In large samples, this is the equivalent of the classical variance. Geometrically, it is the radius of curvature of the support curve and is obtained by taking the inverse of the information --i.e. the inverse of minus the second derivative of the support function. When taken at the evaluate, it is known as the observed formation.

Span: The square root of the observed formation. It is a measure of the width of the support curve near the maximum, and the equivalent of the standard deviation in large samples.

M-unit support limits: These are the two values of a parameter for which the support is M units below the maximum. The support curve of a parameter can be conveniently summarized by its evaluate quoted with its 2-unit support limits in parentheses. In the limiting case of large

samples, the support curve is a parabola with 2-unit support limits two standard deviations away from the mean.

M-unit support region: For a number of parameters, this is the region in the parameter space bounded by the curve on which the support is M units less than the maximum.

Evaluator: The explicit solution, in algebraic form, for the evaluate of a parameter. It can be obtained by setting the first derivative of the support function equal to zero, but this is not always possible, and computer iteration is often the easiest way out to find the evaluate(s). The next section describes a computer approach designed to make life easier for programmers dealing with genetical problems.

B. COMPUTERIZATION

Most of the subprogrammes described in this section were originally written by Karen P. Glenn under the supervision of J.H. Edwards who presented them in his article on marker algebra in 1972. In the course of years, they were mainly updated by Karen P. Glenn and Eleidi Chautard-Freire-Maia to accommodate new machines and changing needs. I have then adapted them to the IBM 370, a computer with words of 32 bits, and made all FORTRAN routines compatible with the WATFIV, FORTRAN G and FORTRAN H compilers. What is presented here is the end product, together with J.H. Edwards' original theory.

1. BIT SHIFTING

In genetics, the word 'marker' designates any system situated at one specific locus on a chromosome, and possessing two or more alleles showing dominance, recessivity or codominance. Such systems are expressed in phenotypes which can be completely determined by the 'yes' or 'no' answers to one or more tests or questions. When these answers are coded as '1' or '0', their sequence can be seen as binary numbers which can be most efficiently processed by computers since these machines store all their information in words composed of between 16 and 48 bits each of which is given the value '1' or '0'.

For instance, a phenotype at the ABO locus can be described by two bits representing responses to tests A and B, the four possible phenotypes O, A, B and AB being 00, 10, 01 and 11 respectively. Similarly, the MNSs phenotypes can be described by four bits each. This coding can be extended to

haplotypes, diplotypes and genotypes. A diplotype is a pair of haplotypes where the order is important; a genotype is an unordered diplotype. Thus, for the ABO system there are

3 haplotypes: O, A, B,

9 diplotypes: OO, OA, AO, AA, OB, BO, BB, AB, BA,

6 genotypes: OO, OA, AA, OB, BB, AB,

and 4 phenotypes: O, A, B, AB.

In order to save core space in the computer, both determinants of diplotypes and genotypes can be stored in the same word by dividing each word into parts M and N for the paternal and maternal determinants:

Paternal	Maternal
M	N

For linkage studies each half can be further subdivided to accommodate the determinants of two loci A and B. Each word is thus divided into four parts labelled I, J, K and L:

Paternal		Maternal	
Locus A	Locus B	Locus A	Locus B
I	J	K	L

In practice, it is useful to reserve the left-most bit for the sign, and designate unknown determinants by a negative number. On the IBM 370, this leaves seven bits for each of the parts I, J, K and L.

Formation of a diplotype from two haplotypes is simply a problem of shifting bits into the appropriate parts of a word, while production of a phenotype

from a diplotype or genotype can be achieved by performing a boolean OR operation on the responses to each test. The four possibilities are

0 or 0 = 0

0 or 1 = 1

1 or 0 = 1

1 or 1 = 1

and show, for instance, that haplotype A of the ABO system - i.e. 10 - orred with haplotype B - i.e. 01 - will give phenotype AB - i.e. 11.

Thirteen subprogrammes have been written to execute a wide range of operations concerned with bit shifting and boolean algebra. Each of these is described below by

a) its name and the list of its parameters, all of which are integers:
   i) array names end with the letter A,
   ii) a parameter giving the dimension of an array starts with the letter I followed by the first letter(s) of the array name,
   iii) all other parameters start with the letter N;
b) the output parameter(s) of SUBROUTINES, or the mention FUNCTION for FUNCTION subprogrammes; (The other parameters are used for input only and remain unchanged.)
c) a description of the operation performed in computer terms;
d) the name of other subprogrammes required, if any.

The subprogrammes are given in alphabetical order. Their computer listings can be found in Appendix I.

BINX (IL, LA, N)

Output: N.

Parameters:

LA: Integer array of IL words.

N: Integer word.

Operation:

The routine is used to obtain the denary value of a boolean string of 0's and 1's which are individually stored in the integer words of array LA. The string is reduced into a string of low order bits in the word N.

No requirements.

ICOMP (NL, NMASK)

This is a FUNCTION.

Parameters:

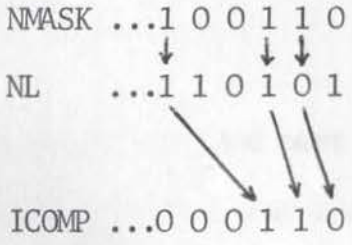
NL: Integer with bits to be compressed.

NMASK: Integer containing masking instructions to be carried out on word NL.

Operation:

The return function is initially cleared, and the operation consists in copying certain bits of word NL into the low order bits of the return function. The shifting is carried out from low order to high order on those bits of NL having an order identical to the '1' bits of NMASK. Bits of NL corresponding to the '0' bits of NMASK are ignored.

Example:



Requirement:

ZBIT.

ILEQ (NP)

This is a FUNCTION.

Operation:

Parts I and L of integer NP are copied into parts K and L of the return function which is initially cleared.

$$\underset{\text{NP}}{|A|B|C|D|} \rightarrow \underset{\text{ILEQ}}{|b|b|A|D|}$$

Requirement:

ZBIT.

IQSH (NP, NQ)

This is a FUNCTION.

Operation:

Part L of integer NP, and part L of integer NQ are copied into parts K and L of the return function which is initially cleared.

$$\underset{\text{NP}}{|A|B|C|D|} + \underset{\text{NQ}}{|E|F|G|H|} \rightarrow \underset{\text{IQSH}}{|b|b|D|H|}$$

Requirement:

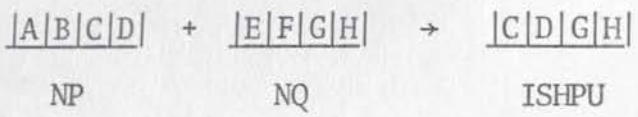
ZBIT.

ISHPU (NDEX, NP, NQ)

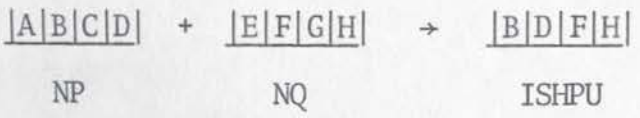
This is a FUNCTION.

Operation:

If NDEX = 1, parts K and L of integer NP, and parts K and L of integer NQ are respectively copied into parts I, J, K, L of the return function.



If NDEX > 1, parts J and L of integer NP, and parts J and L of integer NQ are respectively copied into parts I, J, K, L of the return function.



Requirements:

NEQ, ZBIT.

KJEQ (NP)

This is a FUNCTION.

Operation:

Parts J and K of integer NP are copied into parts K and L of the return function which is initially cleared.

Requirement:

ZBIT.

LMARK (NP, NQ)

This is a FUNCTION.

Operation:

If NP or NQ or both are negative, LMARK = -1.

If not, LMARK = IQSH (NP, NQ).

Requirement:

IQSH.

MEQ (NP)

This is a FUNCTION.

Operation:

Part M of integer NP is copied into part N of the return function which is initially cleared.

Requirement:

ZBIT.

MORN (NP)

This is a FUNCTION.

Operation:

The boolean OR operation is performed on parts M and N of integer NP, and the result stored in part N of the return function which is initially cleared.

Requirements:

MEQ, NEQ, ZBOOL.

NEQ (NP)

This is a FUNCTION.

Operation:

Part N of integer NP is copied into part N of the return function which is initially cleared.

A	B	C	D	→	b	b	C	D
NP					NEQ			

Requirement:

ZBIT.

XBIN (N, NX, NPOS, LA)

Output: NPOS, LA.

Parameters:

N: Number of low order bits to be considered in N.

NX: Integer word.

NPOS: Number of '1' bits in N.

LA: Integer array of 28 words.

Operation:

The routine is used to dissect an integer into bits by giving the value of each of the N low order bits of word NX to a complete high-index word in the zeroized array LA.

No requirements.

ZBIT (NP, NX, NY, NR, NZ)

Output: NR.

Parameters:

All integers.

NX, NY, NZ represent bit order.

Operation:

When the sign bit is designated as the first bit of a word, the operation consists in taking the NXth to NYth bits inclusively from word NP, and moving them into the NZth to (NZ + NY - NX)th bits of NR, leaving the other bits of NR unchanged.

No requirements.

The routine exits in two forms:

1) In FORTRAN for use in conjunction with calling programmes run on the WATFIV compiler,
2) In IBM-360/70 ASSEMBLER for use in conjunction with calling programmes run on the FORTRAN G and FORTRAN H compilers.

ZBOOL (NP, NQ, NR, NDEX)

Output: NR.

Parameters: All integers.

Operation:

The result of a boolean operation performed on the bits of NP and NQ is stored in NR.

The operation is determined by NDEX:

NDEX = 1 for AND

NDEX = 6 for Exclusive OR

NDEX = 7 for OR.

No requirements.

The routine exits in two forms:

1) In FORTRAN for use in conjunction with calling programmes run on the WATFIV compiler,
2) In IBM-360/70 ASSEMBLER for use in conjunction with calling programmes run on the FORTRAN G and FORTRAN H compilers.

2. W-ARRAYS

The previous routines all deal with single words containing the binary representation of responses - or 'qualities' - of genetic entities such as haplotypes, diplotypes, genotypes and phenotypes. But the complete specification of such entities requires both these qualities and their frequencies - or quantities.

Each entity can thus be fully described by a dyad of two computer words, the first being its qualitative element, and the second its quantitative one. The set of dyads describing a marker can be grouped in a double integer vector of qualities and quantities called W-array (J.H. Edwards, 1972). For practical purposes, a head of four words is attached to each array so that the array contains

a) in the first word of its head: the number of words it occupies,
b) in the second word: the scaling factor by which its quantitative elements have been multiplied to give an integer value,
c) in the third word: its alphanumeric identity (in letters and/or ciphers),
d) in the fourth word: the number of responses - or bits - used to describe each quality.

The head allows the easy reference to W-arrays and their dyads. Basic operations such as the finding, copying and updating of W-arrays are done by simple calls to W-routines, six of which are documented below in the pattern already used for the bit-shifting routines.

It should be noted, in addition, that names of W-arrays in the parameter lists always end with the letter W and are integer arrays, irrespective

of their first letter. Qualitative and quantitative elements of a dyad are called JB and JN respectively. Owing to the addition or deletion of dyads from a W-array by some of the W-routines, the dimension of the array in the parameter list must sometimes be greater than the value stored in the first word of the head. Indeed, the value stored in the head refers to the number of relevant words present in the array, i.e. twice the number of dyads, plus four. The operation performed by each routine is summarized in general terms under the title 'Usage'. Computer listings can be found in Appendix I.

ADDEW (SW, IS, JB, JN)

Input-output: SW.

Parameters:

SW: Array of IS words.

JB: Quantitative element of a dyad.

JN: Qualitative element of a dyad.

Operation:

Words JB and JN are added to the tail of the array SW to form a new dyad. 2 is added to the value of SW(1).

No requirements.

Usage:

To add a new dyad to a W-array.

EQW (IAW, II, IBW)

Output: IBW.

Parameters:

IAW and IBW: Arrays of NI words each.

Operation:

The array IAW is copied into the array IBW.

No requirements.

Usage:

To copy a W-array into another.

FNDWS (HWS, IH, NAME, SW, IS)

Output: SW.

Parameters:

HWS: String of W-arrays totalling a maximum of IH words.

NAME: Integer containing the alphanumeric identity of a W-array to be found amongst all W-arrays forming HWS.

SW: Array of IS words.

Operation:

The W-array stored in the string HWS of W-arrays, and the identity of which is identical to that in NAME, is copied into SW, an array of IS words.

No requirements.

Usage:

To retrieve a W-array from a string of W-arrays.

NINBW (SW, IS, JB, JN)

Input-output: SW.

Parameters:

SW: Array of IS words.

JB: Qualitative element of a dyad.

JN: Quantitative element of a dyad.

Operation:

When a dyad in SW is found to have a qualitative element identical to JB, its quantitative element is replaced by JN which is assumed to be correctly scaled.

No requirements.

Usage:

To update the quantitative element of a dyad in a W-array.

RNKBW (SW, IS)

Input-output: SW.

Parameter:

SW: Array of IS words.

Operation:

The dyads of array SW are ordered according to the denary values of their qualitative elements, highest values coming first. If two dyads have identical qualitative elements, the second is deleted after its quantitative element has been added to the first, and the value of SW(1) is reduced by 2.

No requirements.

Usage:

To rank a W-array.

UPDWS (HWS, IH, NAME, SW, IS)

Input-output: HWS.

Parameters:

HWS: String of W-arrays totalling a maximum of IH words.

NAME: Integer containing the alphanumeric identity of a W-array of IS words, which is to be updated.

SW: Array of IS words containing updated information.

Operation:

The W-array with identity NAME, stored in the string HWS of W-arrays, is replaced by the array SW of same dimension.

No requirements.

Usage:

To update a W-array stored in a string of W-arrays.

3. MARKER ALGEBRA

Given all the routines presented in the last two subsections, the next step made by J.H. Edwards was to develop a special algebra using operators capable of acting jointly but differently on the two types of elements forming the dyads.

For instance, a single operator can be devised that will OR the qualities of two haplotypes, and multiply their quantities to produce a dyad containing the resulting phenotype. When subroutines perform such operations on all dyads of W-arrays, they are called W-operators. Eight W-operators are described below. Operations are explained in computer terms, and their genetical meaning given under 'usage'.

ABCW (HAW, IHA, HBW, IHB, HABW, IHAB)

Output: HABW.

Parameters:

HAW: Haplotype array of IHA words for locus A.

HBW: Haplotype array of IHB words for locus B.

HABW: Haplotype array of IHAB words for loci A and B.

IHAB: ((IHA-4)(IHB-4)/2)+4

Operation:

For every possible pair of dyads --one from HAW with qualititative element IP, and the other from HBW with quantitative element IQ-- the operation IR = ISHPU (2, IP, IQ) is performed, and the quantitative elements multiplied and rescaled. The results become dyads of a new array HABW.

Requirement:

ISHPU

Usage:

To generate a set of haplotypes at two loci from two sets of haplotypes at one locus each.

ALIKW (EXW, IEX, OBW, IOB)

Output:

Function ALIKW, the likelihood value.

Parameters:

EXW: Array of IEX words representing expected phenotypes and their proportions.

OBW: Array of IOB words representing observed phenotypes and their frequencies.

Operation:

When a qualitative element appears in both OBW and EXW, its quantitative element in EXW is raised to the power of its quantitative element in OBW. This is done for all dyads in OBW, and the results are multiplied together. The return function takes the value of the product.

No requirements.

Usage:

To calculate the variable part of a likelihood function, given the expected and observed phenotypes.

CHKDW (DW, ID, NPHEN, DRW)

Output: DRW.

Parameters:

DW: Diplotype array of ID words.

NPHEN: Integer containing a phenotype in denary form.

DRW: Reduced diplotype array of ID words.

Operation:

Dyads of DW are copied into DRW with the exception of those with diplotypes incompatible with phenotype NPHEN. The relevant length of array DRW is changed accordingly, but the quantitative elements of the dyads stay the same.

Requirements:

EQW, MANN, MEQ, MORN, RNKBW, ZBIT, ZBOOL.

Usage:

To find out all diplotypes that can give a specific phenotype at one or two loci, given all possible diplotypes at that or those loci in the population.

DHW (HW, IH, DW, ID, HAW)

Output: HAW.

Parameters:

HW: A population haplotype array of IH words.

DW: An individual's diplotype array of ID words.

HAW: An individual's haplotype array of IH words.

Requirements:

MEQ, NEQ, RNKBW.

Usage:

To find out all of an individual's possible haplotypes at two loci, given his possible diplotypes and all haplotypes present at these two loci in the population.

HHDW (HAW, IHA, HBW, IHB, DABW, IDAB)

Output: DABW.

Parameters:

HAW: Haplotype array of IHA words for parent A.

HBW: Haplotype array of IHB words for parent B.

DABW: Diplotype array of IDAB words for children born to parents A and B.

IDAB: ((IHA-4)(IHB-4)/2)+4

Operation:

For every possible pair of dyads --one from HAW with qualititative element IP, and the other from HBW with quantitative element IQ-- the operation IR = ISHPU (1, IP, IQ) is performed, and the quantitative elements multiplied and rescaled. The results become dyads of a new array DABW.

Requirement:

ISHPU.

Usage:

To generate a set of diplotypes from two sets of haplotypes.

ORPW (HPW, HMW, PW)

Output: PW.

Parameters:

HPW and HMW: Haplotypes arrays of twelve words each.

PW: Phenotype array of 36 words, the OR product of HPW and HMW.

Operation:

For every possible pair of dyads --one from HPW and the other from HMW-- the qualitative elements are orred, and the quantitative elements multiplied and rescaled. The results become dyads in a new array PW which is ranked before being output.

Requirements:

RNKBW, ZBOOL.

Usage:

To generate a set of possible phenotypes from two sets of possible parental haplotypes.

PPRDW (HABW, IHAB, JIND, JFAT, JMOT, DW, ID)

Output: DW.

Parameters:

HABW: Population haplotype array of IHAB words for two loci.

JIND: An individual's phenotype in denary form.

JFAT: Father's phenotype in denary form.

JMOT: Mother's phenotype in denary form.

DW: Diplotype array of ID words for individual with phenotype JIND.

Requirements:

CHKDW, DHW, EQW, HHDW.

Usage:

To determine an individual's possible diplotypes, given his phenotype and those of his parents if they are known.

SEGW (JDIPL, NTHETA, HAPLW)

Output: HAPLW.

Parameters:

JDIPL: Integer containing the denary value of a diplotype at two loci.

NTHETA: 100 times the value of the recombination fraction θ.

HAPLW: Haplotype array of twelve words.

Operation:

Given integers NTHETA and JDIPL,

$$\underline{|A|B|C|D|}$$
JDIPL

the four dyads of array HAPLW are formed and consist of the qualitative elements

$$\underline{|b|b|A|B|}_{\text{HAPLW(5)}}\ ,\ \underline{|b|b|A|D|}_{\text{HAPLW(7)}}\ ,\ \underline{|b|b|C|D|}_{\text{HAPLW(9)}}\ \text{and}\ \underline{|b|b|B|C|}_{\text{HAPLW(11)}}\ ,$$

paired with the quantitative elements 100-NTHETA, NTHETA, 100-NTHETA and NTHETA respectively.

Requirements:

ILEQ, KJEQ, MEQ, NEQ, RNKBW.

Usage:

To obtain the description and proportions of the four possible haplotypes derived from the segregation of two loci into gametes, given the original diplotype at these two loci, and θ, their recombination fraction.

4. HANDLING OF PEDIGREE DATA

In practice, the handling of large kindreds creates computing problems that can be alleviated by calling subroutine NUMAB which allocates a serial integer number to each individual of the kindred. All information concerning a particular individual can then be stored in matrix rows or columns of that number. When parental serial numbers are also stored, sibships and nuclear families can be rapidly reconstructed after a call to subroutine FAMGAL.

When the number of tests used to describe a diplotype requires more bits than are available in the four parts of a computer word, routines TRIMW and PRUNW will eliminate from the phenotypes of a family those bits that carry no information on linkage, till the number of bits is small enough, i.e. seven bits per phenotype on the IBM 370.

Some coding and punching errors can be detected by subroutine ILLEG checking on legitimacy.

These five routines are described hereunder. Their listings can be found in Appendix I.

FAMGAL (MREL, IRO, NIND)

Input-output: MREL.

Parameters:

MREL: Matrix of dimensions (IRO, 3) containing the father's and mother's serial numbers in columns 1 and 2 respectively, for up to IRO members of a kindred.

NIND: Number of individuals in a particular kindred. NIND $\leq$ IRO.

Operation:

If in a kindred of NIND individuals, no. I is found to have the same parents as no. J (where J > I), the value J is stored in MREL(I, 3). If there is no sib following no. I, MREL(I, 3) = -1.

No requirements.

Usage:

To identify sibships when parental identities are given.

ILLEG (NLOC, NAME, MAR, IRO, ICOL, NIND, NL)

Input-output: MAR.

Parameters:

NLOC: No. of the column where the phenotypes of locus NAME are stored in matrix MAR.

NAME: Alphanumeric name of locus no. NLOC.

MAR: Matrix of dimensions (IRO, ICOL) containing all phenotypes at all loci of a nuclear family.

NIND: Number of individuals forming the nuclear family stored in MAR. NIND $\leq$ IRO.

NL: Number of loci analysed. NL $\leq$ ICOL.

Operation:

All of a child's phenotypes are compared to the parents'. If a child is positive for a test negative in both parents, the child is declared illegitimate in a printed message, and all of his phenotypes are blanked out. This is done for the whole sibship in one call of the routine. The operation is not carried out if one of the parents is unknown.

Requirements:

ZBIT, ZBOOL.

Usage:

To check on legitimacy.

NUMAB (NA, NB)

This is a FUNCTION.

Parameters:

NA and NB: Two integer-words containing an individual's alphanumeric identity.

Operation:

An individual with identity NA and NB is given a unique serial number. Further entries of the same individual in the routine give out the same number. The routine must first be initialized by a call where NA is equal to the alphanumeric string 'ZER'. The maximum number of individuals to be serialized is 150. Up to eight alphanumeric characters can be stored in NA and NB to define an individual.

No requirements.

Usage:

To give each member of a kindred a unique integer serial number for use within core.

PRUNW (SW, IS, NMASK)

Input-output: SW.

Parameters:

SW: Array of IS words.

NMASK: Integer containing elimination instructions to be carried out on all qualitative elements of AW.

Operation:

The operation IP = ICOMP (IP, NMASK) is carried out an all qualitative elements of the array SW which is ranked before being output.

Requirements:

ICOMP, RNKBW.

Usage:

To eliminate uninformative tests from loci which would otherwise require more than a word for the description of a diplotype.

TRIMW (MAR, IRO, ICOL, NIND, NLOC, NMASK, N)

Input-output: MAR, NTESTS.

Output: NMASK.

Parameters:

MAR: Matrix of dimensions (IRO, ICOL) containing all phenotypes of a nuclear family.

NIND: Number of individuals forming the nuclear family stored in MAR.

NLOC: No. of the column of matrix MAR where the phenotypes of the locus to be trimmed are stored.

NMASK: Integer containing trimming instructions.

N: Number of tests considered for locus no. NLOC.

Operation:

The bits of same order in the children's phenotypes are copied into N strings of 0's and 1's, and this is also done for the parents. If two strings are identical or exactly opposite, one of the orders they represent is eliminated from the phenotypes of all members of the family --that order where both parents have a '1' bit, if this is observed. Similarly, if all bits of a specific order are 1's or all are 0's in parents and children alike, they are eliminated. Elimination is carried out till the number of tests is equal to or less than 7, and the eliminated orders are identified by a '1' bit in that order in NMASK.

Requirements:

ICOMP, ZBIT, ZBOOL.

Usage:

To identify and eliminate uninformative bits from the phenotypes of loci which would otherwise require more than a word for the description of a diplotype.

III LINKAGE IN DIPLOIDS

A. LODS

Given the powerful tools presented in the last chapter, it becomes relatively easy to write a computer programme to study the segregation pattern of loci, or pairs of loci, observed in pedigree data. The likelihood method generally employed for the linkage analysis of human pedigrees differs from the method of support in that it uses logarithms to the base 10, and talks of 'lods' --an acronym for 'logarithm of the odds ratio' (Barnard, 1949)-- rather than 'support'. It was presented in a series of papers by Morton (Morton, 1955, 1956, 1957; Steinberg and Morton, 1956), and was based on the earlier works of Wald (1947), Haldane and Smith (1947), Barnard (1949) and Smith (1953). Given the pedigree data D on two loci, and their likelihood $L(\theta|D)$ in terms of the recombination fraction θ, it consists in comparing various hypotheses about θ with the hypothesis of no linkage. If two loci are not linked, or far away on the same chromosome, they recombine in 50% of cases on average, and the evaluate of θ equals 0.5. Otherwise, they are closely linked and the lod for θ_1 (where $0.0 \leq \theta_1 < 0.5$) is

$$\log_{10} \left(L(\theta = \theta_1|D) \; / \; L(\theta = 0.5|D) \right)$$

and positive since the value for θ is more likely θ_1 than 0.5. When the phase (coupling or repulsion) of the two loci is known, lods of approximately 1, 2 and 3 are obtained for $\theta_1 = 0.00$ in double backcrosses with 3, 7 and 10 children, all of whom are non-recombinants. For all matings, a lod of 0.869 at any value of θ_1 is the equivalent of a difference of support of 2, but geneticists tend to wait till the lods reach a cons-

ervative absolute value of about 3.0 (or support of 6.9) before concluding in favour of the evaluate.

The problem of calculating lods can be greatly simplified by using the marker algebra detailed in the preceding chapter. Subroutine ALNKW was written especially for this purpose. Given the arrays of possible parental diplotypes at two loci, it calls routine SEGW (see page 38) to generate for each parent the qualities and quantities of the four gametes derived from each diplotype when the recombination fraction has some particular value θ_1. (θ can be different for the two parents, but at present ALNKW uses the same value of θ for both sexes.) It then calls ORPW (see page 36) to obtain the W-array of phenotypes for all the possible children of these parents, and finally finds the likelihood of the actual children by calling ALIKW (see page 32).

One of the input parameters of ALNKW is an index that specifies to the routine which of the parents is informative for linkage. Finding out the value of this index and the possible parental diplotypes is the task of the linkage programme known locally as MARK III in which are incorporated all the routines described in the last chapter, as well as others, including ALNKW. (All routines to be mentioned in this chapter are listed in Appendix I. The specifications for ALNKW are given below.)

ALNKW (DPW, IDP, DMW, IDM, PCW, IPC, NTH, NDEX)

Output:

Function ALNKW, the log-likelihood value.

Parameters:

DP: Paternal diplotype array of IDP words.

DM: Maternal diplotype array of IDM words.

PCW: Array of IPC words containing the phenotypes of observed children.

NTH: Recombination fraction θ multiplied by 100.

NDEX: Index designating the type of parental information.

NDEX = 2^{21} when only the father is informative for linkage.

= 2^{22} when only the mother is informative.

= $2^{21} + 2^{22}$ when both parents are informative.

= 0 when NTH = 50.

Requirements:

ALIKW, ORPW, SEGW.

Usage:

To calculate the variable part of the log-likelihood of a sibship, given the children's phenotypes at two loci, their parents' possible diplotypes, and a particular value for the recombination fraction θ between the two loci.

B. MARK III

1. HISTORICAL

MARK III is the most recent in a series of computer programmes designed by J.H. Edwards for linkage investigations. The MARK I was written in 1968 using a less generalized approach for an 8K IBM 1130 (Falk and Edwards, 1970; J.H. Edwards, 1972b) and has been considerably extended by Cathy T. Falk at the New York Blood Center where it is in regular use.

The MARK II is based on the marker algebra, was mainly written by Karen P. Glenn and used on a variety of computers in England and the U.S.A. for the analysis of data gathered at the international workshop on HL-A (Edwards et al., 1972), and the Brazilian data of N.E. Morton and his colleagues (Chautard-Freire-Maia, 1974). It is no longer operational in England. Its most advanced version is now in Aarhus, Denmark, where it has been operational since 1972. It was extended by L.U. Lamm and Karen P. Glenn to accept complex phenotypes and reduce them to minimal informative boolean strings (see PRUNW and TRIMW, pages 43 and 44), a feature which can be a great advantage with the HL-A and Gm loci.

The MARK III is my adaptation of MARK II to the requirements of the Winnipeg Rh Laboratory and the capacity of the University of Winnipeg computer, an IBM 370 supporting the WATFIV, FORTRAN G and FORTRAN H compilers. Its major differences with MARK II concern the input format, the printed output and the elimination of some errors which, however, did not seriously affect the answers of MARK II in the absence of close linkage. Some computing improvements have resulted in an appreciable

reduction in the amount of core used by the programme. The version I shall describe is that which is operational at the Rijksuniversiteit te Leiden in the Netherlands, and almost identical to that of Winnipeg. It is run in Leiden on an IBM 370/158 computer and uses 204K bytes of core at execution time in FORTRAN H. Tests have shown that on average it produces two lods per second on this machine.

2. GENERAL FEATURES

In one run, MARK III can calculate five different lods for all pairs of as many as 70 loci. It can accommodate kindreds of as many as 150 persons, with sibships of up to 20 individuals. A locus can have as many as ten haplotypes in the population, and there is a maximum of 40 diplotypes per pair of loci.

The programme analyses any sibship that has two parents. When present, one, two, three or four grandparents are also taken into account to remove impossibilities from the sets of parental diplotypes at two loci. When the two loci have less than six haplotypes each, the lods are even calculated for sibships with only one known parent. For reasons of economy, more complex loci are only analysed when both parents are present, because they are then generally informative about linkage, and the small proportion of information gained from incomplete families does not seem to justify the time and money necessary to calculate all possibilities of segregation in an unknown person. The same reason applies to all other pairs of loci for sibships without known parents. Tests have shown that the programme would take minutes, and even hours, to compute all possibilities for a few pairs of loci. It was therefore decided to eliminate these families from routine data processing. Also, the lods are not

calculated for some pairs of loci where enough information has already been accumulated to show that they were not closely linked (New Haven Conference, 1973):

ABO: Fy, Kell, Rh, P, MNSs, Hp

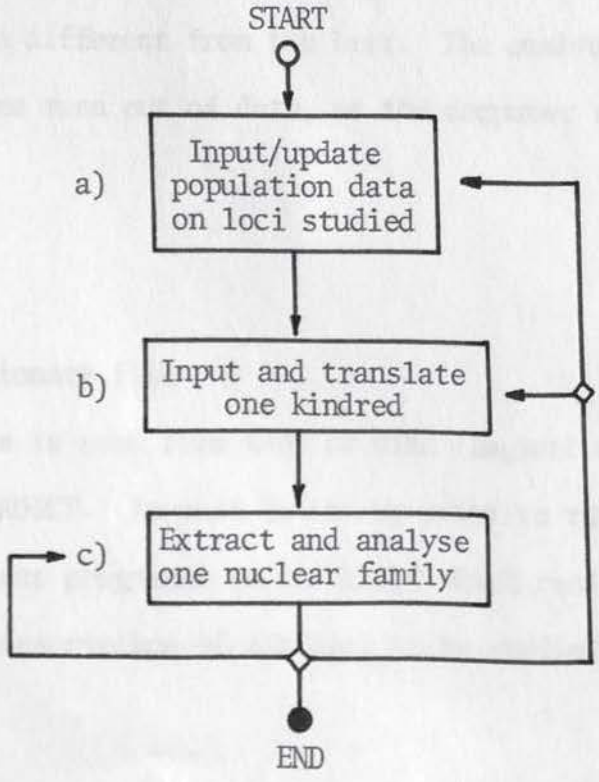
MNSs: Jk, Fy, Lu, Rh, P, Hp

Rh: Jk, Lu, Gm, Hp

P: Fy, Rh, Hp

Fy: Hp.

For the same practical reasons, large kindreds are broken into smaller units before being analysed. The unit is the nuclear family formed by a sibship with one or two parents and up to four grandparents. The nuclear families are analysed independently, one after the other, and although this obviously results in the loss of the information provided by the phenotypes of the other relatives, the loss is usually small and unlikely to be more than 10% on average. At any rate, the data currently available to most laboratories are made of single nuclear families analysed with complete efficiency by MARKs I, II and III. The analysis is carried out in the three stages outlined in the flow-diagram below:



a) A dictionary file is input into core. It contains the names of up to 80 loci found in the population, the alphanumeric codes used to describe their phenotypes on the data cards, the equivalent boolean codes of these phenotypes --i.e. the strings of '0' and '1' bits to be stored in W-arrays-- and the boolean codes of all haplotypes with their proportions in the population. These proportions can be changed before or after the analysis of any kindred.

b) The data on one kindred are read according to specifications to be described later. The dictionary is then consulted for the translation of phenotypes from alphanumeric to boolean codes.

c) A nuclear family is extracted from the kindred, legitimacy is checked, the mating type is determined for each possible pair of loci, and the lods are computed, printed and punched. This is repeated until the whole kindred has been analysed.

The allele proportions can be changed if the next kindred comes from a population different from the last. The analysis is carried out until the programme runs out of data, or the computer runs out of time.

3. INPUT

a) The dictionary file

This file is read from tape or disc (logical unit 10) by the subroutine RDICT. It must be set up prior to running MARK III, by an independent programme called LLDIC which reads the data cards containing the description of all loci to be studied. A card can carry the

information on either a haplotype or a phenotype, and for each locus, the phenotype cards are grouped together, followed by a blank card, a comment card, all the haplotype cards, and a final blank card.

The haplotype and phenotype cards give the following details:

i) the number of bits necessary to write the boolean representation of the haplotypes and phenotypes at that locus (in columns 1-2-3),

ii) the name of the locus (in columns 5-6-7),

iii) an 'H' or a 'P', depending on whether the card describes a Haplotype or a Phenotype (in column 10),

iv) on haplotype cards, the proportion of each haplotype in the population (i.e. the 'gene frequency' or allele proportion, in columns 14 to 18 incl.),

v) an alphanumeric code of up to three characters used to define the haplotype or phenotype (in columns 22-23-24),

vi) the boolean code, punched as a left-justified string of up to fifteen 1's and 0's seperated by two blanks (starting in column 28),

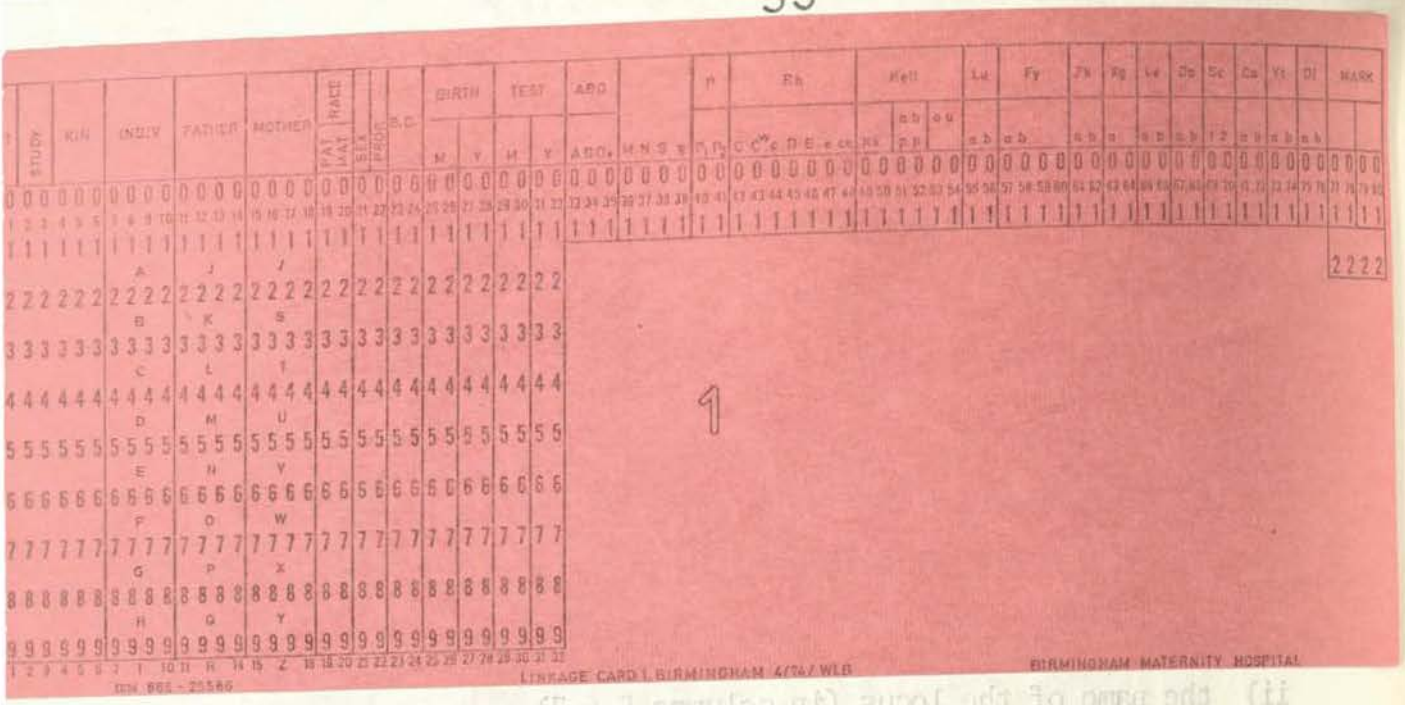
vii) an alternative alphanumeric code of up to four characters, if needed (in columns 77-78-79-80).

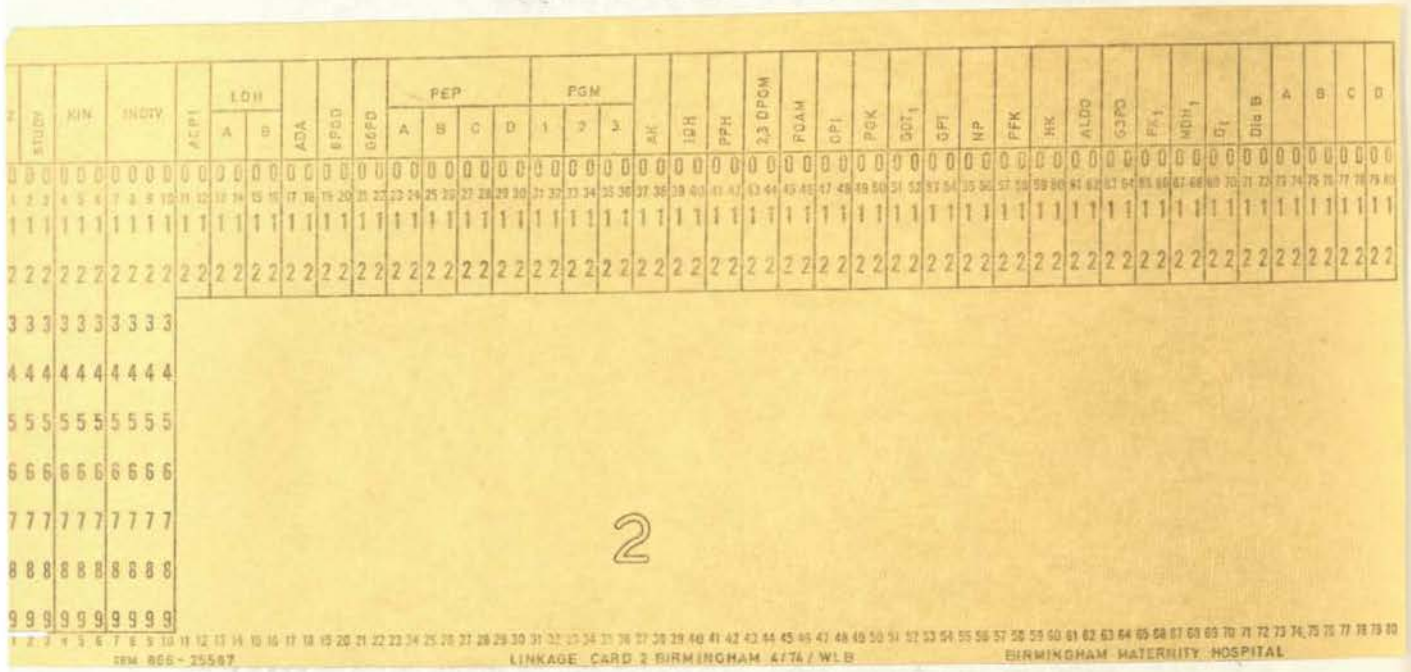
The FORTRAN format used to punch these seven items onto cards in that order is

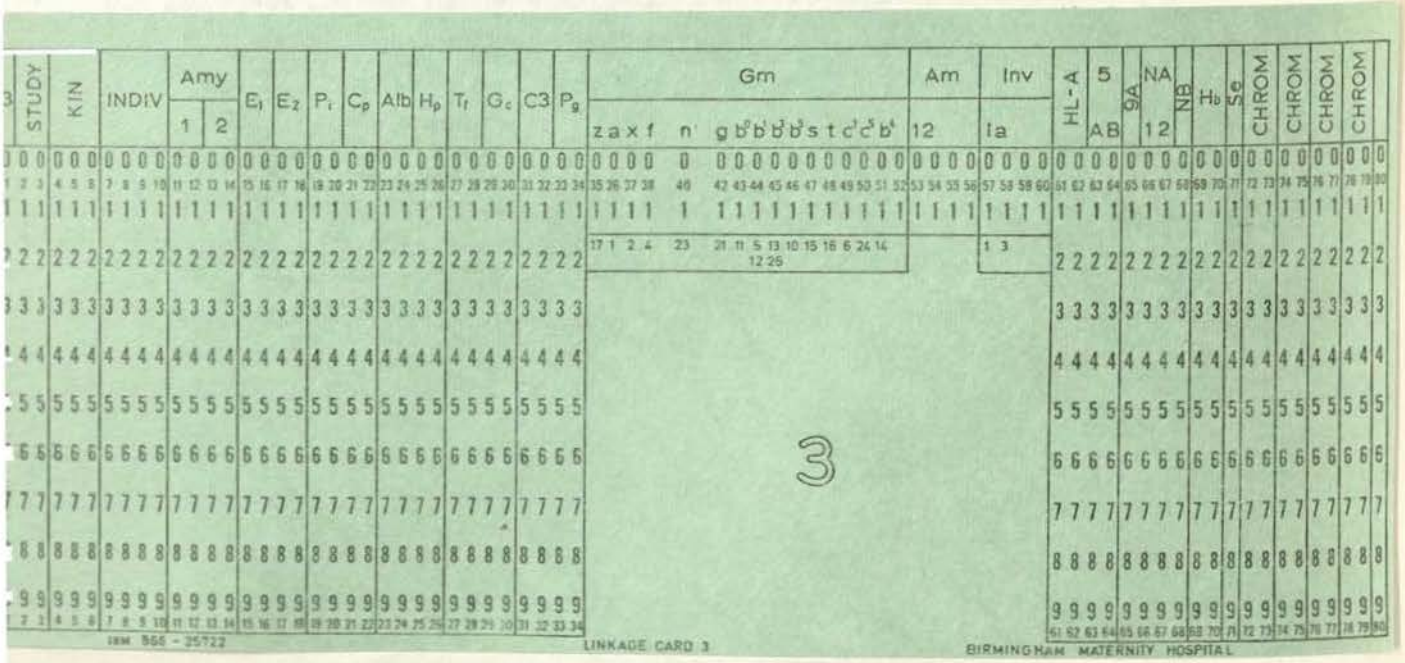
(I3, 1X, A3, 2X, A1, 3X, F5.4, 3X, A3, 1X, 15(2X, I1), 6X, A4).

When a haplotype is described by a string of 0's only, it must be punched on the last haplotype card. This is because dominant systems are recognized by the last haplotype having this null quality.

For each locus, all previous cards are preceded by a single locus card







Input cards used for pedigree data.

which must contain

i) the number of bits used for the boolean codes (columns 1-2-3),

ii) the code name for the locus (columns 5-6-7),

iii) an 'L' in column 10,

iv) a number in column 79.

This number will tell MARK III which alphanumeric code is to be used to make the translation into boolean. The number has the value 2 if the first code (in columns 22-23-24) is to be used,

3 if the alternative code (in columns 77-78-79-80) is,

4 if the translation has already been done (this applies to Rh and Gm phenotypes translated into boolean by the special routines RHBOO and GMBOO),

0 if the locus is not to be considered.

The format of the locus card is (I3, 1X, A3, 2X, A1, 68X, I1).

The dictionary used in Leiden is given in Appendix II.

b) Pedigree data

These are punched on the cards shown opposite. Three cards are available for each individual, the last two being optional. Within a data set, all individuals must have the same number of cards, even if there are no phenotypic data. Columns 1-10 contain an alphanumeric identification on the three cards and must always be filled if the card is used at all. The cards for each individual must be in numerical order but the individuals can be in any order within a kindred. This facility makes it necessary for any individual identified as a father or a mother on a child's card 1 (in columns 11-14 or 15-18) to have his or her own card 1.

The programme needs this to sort out the family relationships and will issue an error message if this is not done. When such input mistakes are detected, MARK III does not analyse the faulty kindred, and skips to the next one. The end of a kindred is specified by a blank card.

The alphanumeric codes used to describe the phenotypes of all loci are listed in Table 2. They must always be punched left-justified in the spaces provided on the cards. Table 2 also gives the code names and code numbers used on the data control cards that identify the input and determine how the data are to be analysed by MARK III.

c) Control cards

Each set of punched pedigree data read by MARK III is preceded by control cards instructing the programme on what to do as the analysis progresses. The full data deck must be ordered as follows:

A) Card A is the first card of the data deck and contains

i) up to five values of θ for the lods to be calculated. These proportions are multiplied by 100 and punched as integers.

ii) the code name (see Table 2) of the locus to be paired with all other loci. If all possible pairs of all loci are wanted, the code is 'ALL'.

The FORTRAN format is (5(1X, I2), 4X, A3).

Examples: (where ↑ means blank, starting in column 1)

↑05↑10↑20↑30↑40↑↑↑↑ALL

↑↑↑↑10↑↑↑↑30↑↑↑↑↑↑↑CR1

Table 2. Alphabetic list of all loci stored in the dictionary file, with code numbers, code names and alphanumeric codes for the phenotypes.

Locus	Code number	Code name*	Codes for phenotypes
Alpha-1-antitrypsin (Pi)	51	PI	F, S, M, FS, FM, MS.
ABO blood group	1	ABD	O, 1 for each test.
Acid phosphatase-1 (ACP1)	18	AP4	A, B, C, X, BA, CA, XA, CB, XB, XC.
Adenosine deaminase (ADA)	21	AD3	1, 2, 3, 21, 31, 32.
Adenylate kinase-1 (AK)	30	AK3	1, 2, 3, 21, 31, 32.
Albumin (Alb)	53	AL	X, N, XN.
Aldo	42	ALD	1, 2, 21.
Amylase-1 (Salivary, Amy-1)	47	AM1	A, B, C, X, BA, CA, XA, CB, XB, XC.
Amylase-2 (Pancreatic, Amy-2)	48	AM2	A, B, C, X, BA, CA, XA, CB, XB, XC.
Ceruloplasmin (Cp)	52	CP	A, B, AB, BA.
Cholinesterase-1 (E_1)	49	CH2	S, A, U, I.
"	49	CH3	F, A, U, AF, UF, I.
"	49	CH4	S, F, A, U, I, FS, AS, AF, US, UA,UF.
Cholinesterase-2 (C5) (E_2)	50	C5	O, 1
Chromosome marker A+ (in columns 72-73 of card 3)	66	CR1	1, 2, 3, 4, 21, 31, 41, 32, 42, 43.
Chromosome marker B+ (in columns 74-75 of card 3)	67	CR2	1, 2, 3, 4, 21, 31, 41, 32, 42, 43.
Chromosome marker C+ (in columns 76-77 of card 3)	68	CR3	1, 2, 3, 21, 31, 32.
Chromosome marker D+ (in columns 78-79 of card 3)	10	CR4	1, 2, 3, 21, 31, 32.
Codominant marker		COD	1, 2, 21.
C'3	57	C3	S, F, X, SF, XS, XF.
Diaphorase B (DiaB)	69	DA	1, 2, 3, 21, 31, 32.
Dominant marker		DM	1, O.

*When the last character of the code name is a number, it usually refers to the number of alleles in the system. The number of symbols was restricted to three because of using a computer restricted to three character words.

+The positions of the loci on the chromosomes are not known to the programme and summation of the lods must be handled informally.

cont'd/..

../cont'd.

Locus	Code number	Code name	Codes for phenotypes
Duffy (Fy)	7	FY2	0, 1 for each test.
Galactose-1-phospho-uridyl transferase (Gt)	46	PUT	1, 2, 3, 21, 31, 32.
Gamma globulin marker (Gm)	60	GM	0, 1 for each test.
Gamma globulin marker (Am)	64	IGA	0, 1 for each test.
Glucose-3-phosphate-dehydrogenase (G3PD)	43	G3P	1, 2, 21.
Glutamic oxalacetic transaminase (GOT_1)	37	GOT	1, 2, 21.
Glutamic pyruvic transaminase (GPT)	38	GPT	1, 2, 3, 21, 31, 32.
Group-specific component (Alpha-2-globulin) (Gc)	56	GC3	1, 2, 3, 21, 31, 32.
Haemoglobin (Hb)	62	HM3	C, S, A, SC, CA, SA.
"	62	HM4	C, S, A, X, SC, CA, SA, XC, XS, XA.
Hexokinase (HK)	41	HK	1, 2, 21.
Haptoglobin (Hp)	54	HP	1, 2, 21.
Histocompatibility region (HL-A)	61	HLA	A, B, C, D, AB, AC, AD, BC, BD, CD.
5A/5B	11	5A	0, 1 for each test.
9A	12	9A	0, 1.
NB1	14	NB1	0, 1.
NA (1, 2)	13	NA	0, 1 for each test.
Inhibiting factor V (Inv 1)	59	INL	0, 1.
" (Inv a)	59	INA	0, 1.
" (Inv a and 1)	59	IN2	0, 1 for each test.
Isocitrate dehydrogenase (ICD)	31	ICD	1, 2, 21.
Kell (K, k)	5	KL2	0, 1 for each test.
" (K, k, Kp^a, Kp^b)	5	KL4	0, 1 for each test.
Kidd (Jk)	8	JK2	0, 1 for each test.
Lactate dehydrogenase A (LDH-A)	19	LDA	1, 2, 3, 21, 31, 32.
Lactate dehydrogenase B (LDH-B)	20	LDB	1, 2, 3, 21, 31, 32.

cont'd/..

../cont'd.

Locus	Code number	Code name	Codes for phenotypes
Lewis (Le)	9	LE	0, 1 for each test.
Lewis (Le) (informative about Secretor)	9	LES	0, 1 for each test.
Lutheran (Lu)	6	LU2	0, 1 for each test.
Malate dehydrogenase (MDH_1)	45	MDH	1, 2, 3, 21, 31, 32.
MNSs blood group	2	MNS	0, 1 for each test.
Nucleoside phosphorylase	39	NP	1, 2, 21.
P_1 blood group	3	PE1	0, 1.
Pepsinogen (Pg)	58	PSG	A, B, C, D, AB, BA, AC, CA, AD, DA, BC, CB, BD, DB, CD, DC.
Peptidase A	23	PA3	1, 2, 3, 21, 31, 32.
Peptidase B	24	PB3	1, 2, 3, 21, 31, 32.
Peptidase C	25	PC3	1, 2, 3, 21, 31, 32.
Peptidase D	26	PD3	1, 2, 3, 21, 31, 32.
Phospho-fructo-kinase (PFK)	40	PFK	1, 2, 21.
Phosphoglyceric acid mutase (PGAM)	34	PGA	1, 2, 21.
Phosphoglucomutase-1 (PGM_1)	27	PG1	1, 2, 3, 21, 31, 32.
Phosphoglucomutase-2 (PGM_2)	28	PG2	1, 2, 21.
Phosphoglucomutase-3 (PGM_3)	29	PG3	1, 2, 21.
Phosphoglycerate kinase (PGK)	36	PGK	1, 2, 3, 4, 21, 31, 41, 32, 42, 43.
Phospho-hexose-isomerase (PHI)	35	PHI	1, 2, 3, 21, 31, 32.
Pyruvate kinase (PK_1)	44	PK	1, 2, 21.
Phosphopyruvate hydratase (PPH)	32	PPH	1, 2, 21.
Rhesus (Rh) (C, C^W, c, D, E, e)*	4	RH6	0, 1 for each test.
Secretor ABH (Se)	63	SE	0, 1.
Sex	17	SEX	M, F.

*The C^W response is interpreted as negative when omitted.

cont'd/..

../cont'd.

Locus	Code number	Code name	Codes for phenotypes
Transferrin (Tf)	55	TF	B, C, X, CB, XB, XC.
2,3 DPGM	33	23D	1, 2, 21.
6-phospho-gluconate-dehydrogenase (6PGD)	22	6P4	A, B, C, X, XC, XB, XA, CB, CA, BA.
Non-specific marker* (in columns 77-78 of card 1)	15		
Non-specific marker* (in columns 79-80 of card 1)	16		
Non-specific marker* (in columns 73-74 of card 2)	70		

*To be used with the code names and phenotypic codes of any of the other loci.

Table 2. Alphabetic list of all loci stored in the dictionary file, with code numbers, code names and alphanumeric codes for the phenotypes.

B) Card B is the second card of the data deck and contains

i) 'XX8' in columns 4-5-6 if some of the haplotypic proportions listed in the dictionary file are to be changed before the first kindred is input,

ii) the record number of the first pair of loci to be analysed. 1 is added to the record number of each following pair.

The format is (3X, A3, I6)

Examples: ↑↑↑XX8↑↑↑↑↑1

↑↑↑↑↑↑↑↑372

C) This card is optional, but one such card is required for each locus whose haplotypic proportions are to be updated in the dictionary file. It must only be used if 'XX8' has been punched in card B, or after a card K (see below). Each card C contains:

i) the name of the locus to be updated,

ii) up to ten real numbers representing all haplotypes at that locus. The integer parts specify the boolean codes, and the fractional parts specify the proportions in the population.

The format is (A3, 1X, 10(F5.3)).

For example, the boolean haplotypic codes and proportions of the ABO system (A_1: 101, 0.209; A_2: 100, 0.070; B: 010, 0.061; O: 000, 0.660) would be punched as follows:

ABD↑5209↑4070↑2061↑0660

or alternatively,

ABD5.2094.0702.0610.660

D) This card is required after any number of C cards.

Punch 'ZZZ' in columns 1-2-3.

E) Card E is used in conjunction with cards F and G to initialize a new set of data. It tells the programme how many data cards are to be read for each individual. The first two columns must contain
00 if only card 1 is used,
20 if card 2 is used with card 1,
03 if card 3 is used with card 1,
23 if both cards 2 and 3 are used with card 1.

F) This card is needed with cards E and G to initialize a new data set. It contains a '1' in all columns corresponding to the code numbers (see Table 2) of the loci to be analysed. The format is (70I1).

G) This card is used with cards E and F to initialize a new data set. It gives the code names (see Table 2) of the loci to be analysed. The names must be given in the same order as they appear on card F, i.e. in the order specified by their code numbers.

H) These cards contain the pedigree data as instructed in the previous subsection (III-B-3-b). Each card must have its own number (i.e. 1, 2 or 3) punched in its first column.

I) Blank card I must follow each kindred. This card can be itself followed by another kindred, or one of cards J and K, or it can be the last card of the data deck.

J) This card is optional and can be used after card I to announce a new data set. If used, it must be followed by cards E, F, G, H and I. Punch '1$_{\uparrow\uparrow}$XX7' in columns 1 to 6 inclusive.

K) Card K is optional and can be used after card I to announce a change in the dictionary file, prior to initializing a new data set. It must be followed by cards C, D, E, F, G, H and I.
Punch '1$_{\uparrow\uparrow}$XX8' in columns 1 to 6 inclusive.

In summary, the data cards are punched for the following purposes:

A+B: job initialization

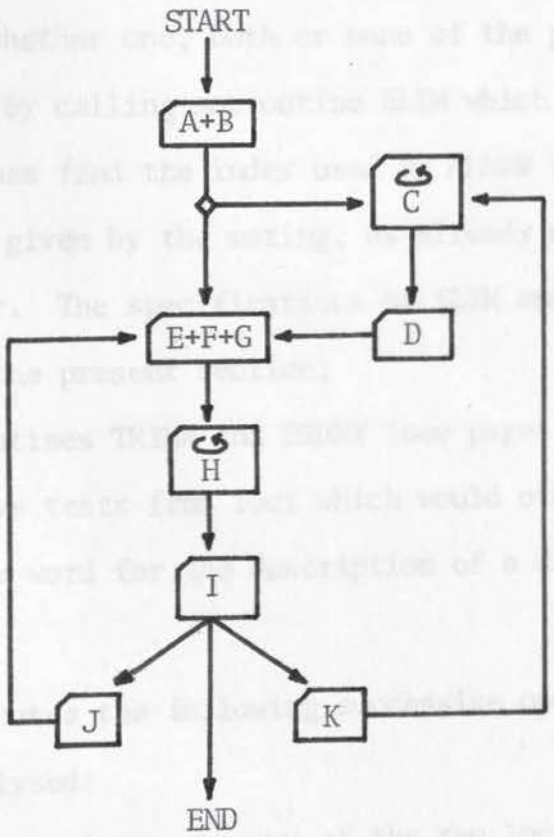
C+D: updating of haplotypic proportions

E+F+G: initialization of new data set

H: pedigree data

I+J+K: branching instructions.

The data deck must therefore be ordered in accordance with the following flow-diagram:



4. ANALYSIS AND OUTPUT

The COMMON block of information used by MARK III is shared by the main programme MLINK and four major subroutines, all listed in Appendix I: RDICT, INWIN, TRANS and BLINK. RDICT inputs the dictionary file into the COMMON block, forms all haplotype W-arrays, and changes the haplotypic proportions if necessary; INWIN reads the pedigree data, calls routine TRANS to translate the phenotypes into their boolean codes, and extracts nuclear families from bigger kindreds.

The linkage analysis per se is then carried out by BLINK which first looks successively at each locus in the following manner:

a) it finds the W-array of its haplotypes by calling FNDWS (see page 26),
b) checks on the children's legitimacy by calling ILLEG (see page 41),
c) determines whether one, both or none of the parents are informative for linkage by calling subroutine ELIM which in turn calls JHOM. These routines find the index used by ALNKW to designate the type of information given by the mating, as already mentioned in section A of this chapter. The specifications of ELIM and JHOM will be found at the end of the present section;
d) it calls routines TRIMW and PRUNW (see pages 44 and 43) to eliminate uninformative tests from loci which would otherwise require more than one computer word for the description of a diplotype.

BLINK then executes the following successive operations on all pairs of loci to be analysed:

e) it finds the haplotype arrays of the two loci in the population by calling FNDWS for each locus (see page 26),

f) sets up the W-array of observed phenotypes in the sibship under consideration. This is done by a subroutine called FCPGW which extracts the parental and grandparental phenotypes stored in the COMMON block of information. Specifications for this routine will also be found at the end of this section;

g) it generates the set of haplotypes at these two loci in the population by calling ABCW (see page 31),

h) determines the parental possible diplotypes by calling PPRDW for each parent (see page 37),

i) calls routine ALNKW (see page 47) to calculate the lods,

j) prints and punches the results.

Each punched card shows the results for one pair of loci with the following information:

- the study name
- the kindred name
- the nuclear family number
- the code names of the two loci
- the lods for five values of θ
- the type of information: P for paternal
  M for maternal
  I for both parents (intercrosses)
- the two boolean representations of the parental phenotypes at both loci
- the number of children in the sibship
- a record number for the card.

The format used for the punched cards is

(A2, A3, I2, 2A3, 5F8.3, A1, 2I5, I2, I6).

ELIM (NLOC, MAR, IRO, ICOL, HABW, IHAB, NFAT)

Input-output: MAR

Parameters:

MAR: Matrix of dimenstions (IRO, ICOL) containing all markers of a nuclear family in integers representing boolean strings.

NLOC: No. of the locus investigated. It is also the no. of the column of MAR where the phenotypes at that locus are stored.

NFAT: No. given to the row of MAR where all the father's phenotypes are stored.

HABW: Population haplotype array of IHAB words.

Operation:

2^{21} and 2^{22} are respectively added to the father's phenotype at locus no. NLOC if the father's and mother's phenotypes are potentially informative for linkage because they are or may be heterozygous at that locus. An unknown parent is declared potentially informative.

Requirements:

JHOM, ZBIT.

Usage:

To find out if a mating is informative for linkage at a defined pair of loci. The routine is called independently for each locus and is given the parents' phenotypes and all haplotypes present at that locus in the population.

JHOM (HAW, IHA, NPHEN)

This is a FUNCTION.

Parameters:

HAW: Haplotype array of IHA words for locus A.

NPHEN: Integer word containing a phenotype in boolean form.

Operation:

NPHEN is declared uninformative and JHOM = 0 if heterozygosity is impossible at that locus. This happens when

1) NPHEN = 0 (i.e. a recessive phenotype)
2) NPHEN is equal to one of the haplotypes in HAW and the last one differs from zero (i.e. the phenotype is homozygous at a locus whose alleles show codominance).

NPHEN is declared informative and JHOM = 1 when heterozygosity is possible. This happens when

1) NPHEN is negative, i.e. unknown,
2) the last haplotype in HAW equals zero and NPHEN does not (i.e. the phenotype is dominant)
3) NPHEN differs from all haplotypes in HAW and the last one differs from zero (i.e. the phenotype is heterozygous for two alleles showing codominance)

No requirements

Usage:

To find out if phenotype NPHEN is informative about linkage or not, given all haplotypes at that locus.

FCPGW (MAR, IRO, ICOL, NLOCB, NLOCC, NIND, LPA, PCW, IPC)

Output: LPA, PCW.

Parameters:

MAR: Matrix of dimensions (IRO, ICOL) containing all phenotypes at all loci of a whole nuclear family.

NLOCB and NLOCC: Nos. of two columns where the phenotypes of loci nos. NLOCB and NLOCC are stored in matrix MAR.

NIND: Number of individuals in the nuclear family stored in MAR. NIND= number of children + 2 parents + 4 grandparents. NIND $\leq$ IRO.

LPA: Array of six words containing the parental and grandparental phenotypes at two loci.

LPA(1): Father
LPA(2): Paternal grandfather
LPA(3): Paternal grandmother
LPA(4): Mother
LPA(5): Maternal grandfather
LPA(6): Maternal grandmother

PCW: Array of IPC words and (2(NIND-6)+4) dyads containing the phenotypes of a sibship at two loci.

Operation:

Setting up of a W-array PCW containing the phenotypes of a sibship at two loci, and of a six-word integer array LPA containing the parental and grandparental phenotypes at these two loci, given the matrix MAR of all phenotypes for this nuclear family.

Requirements:

ADDEW, LMARK, RNKBW.

C. HANDLING OF LINKAGE RESULTS

1. STORAGE AND GRAPHICAL DISPLAY

The masses of data collected around the world, and the impressive number of combinations of loci informative for linkage (up to $(n^2-n)/2$ combinations for n loci) would make publication and storage of results very cumbersome and expensive if a summary of these could not be produced. The lod curve for all values of θ tends to be part of a parabola as sample size increases, and its shape can therefore be imagined or approximately reconstructed when as few as two of its values are given. The lods at $\theta = 0.1$ and $\theta = 0.3$ are useful points to choose because of their positions on the curve and because their values for several types of matings are given in Morton's papers and other published tables (Maynard-Smith et al., 1961; Race and Sanger, 1968, 1974 in press). Linkages can thus be detected by the calculation of these two lods for each combination, a procedure which greatly saves on time and money, and allows the presentation of lods on several combinations of loci in the form of two triangular matrices (J.H. Edwards, 1972; Bjarnason et al., 1973; Chautard-Freire-Maia, 1974).

In large samples, the ratio of these two lods is expected to be constant and to depend solely on the true recombination fraction $\bar{\theta}$ between the two loci. This can readily be seen, for instance, in double backcrosses where the phase is known:

The lod is then $\log_{10} \frac{\theta^R(1-\theta)^D}{.5^N}$

where $N = R+D$, and R and D are the observed numbers of recombinant and non-recombinant children respectively. R and D are expected to be equal

to $N\bar{\theta}$ and $N(1-\bar{\theta})$ respectively, so that the ratio of the two lods is

$$\frac{\text{lod at } \theta = 0.1}{\text{lod at } \theta = 0.3} = \log_{10} \frac{.1^{N\bar{\theta}} \, .9^{N(1-\bar{\theta})}}{.5^{N}} \div \log_{10} \frac{.3^{N\bar{\theta}} \, .7^{N(1-\bar{\theta})}}{.5^{N}}$$

$$= (N \log_{10}(.1^{\bar{\theta}} \, .9^{(1-\bar{\theta})}/.5)) \; / \; (N \log_{10}(.3^{\bar{\theta}} \, .7^{(1-\bar{\theta})}/.5))$$

$$= \log_{10}(.1^{\bar{\theta}} \, .9^{(1-\bar{\theta})}/.5) \; / \; \log_{10}(.3^{\bar{\theta}} \, .7^{(1-\bar{\theta})}/.5)$$

which is constant for any true recombination fraction $\bar{\theta}$. Figure 1 gives this ratio plotted against the value of θ between 0.0 and 0.5. There is a vertical asymptote because the lod for $\theta = 0.3$ equals zero when $\theta = 0.397$.

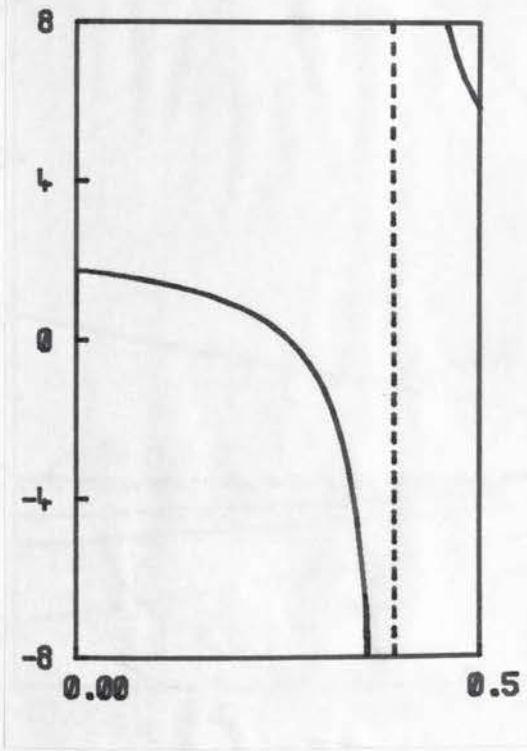


Fig. 1. Expected ratio of the lod at $\theta = 0.1$ and that at $\theta = 0.3$ plotted against the true recombination fraction $\bar{\theta}$ between two loci, for double backcrosses where the phase is known.

A pair of lods can also be plotted as a point on a grid (Edwards et al., 1972), with the lod at $\theta = 0.3$ against that at $\theta = 0.1$, in which case the line joining this point to the origin is expected to be at a constant angle α with the x-axis for any true $\bar{\theta}$. This is because the cotangeant of the angle α is equal to the constant ratio of the two lods:

$\cot \alpha = x/y =$ (lod at $\theta = 0.1$) / (lod at $\theta = 0.3$) = constant.

For a given $\bar{\theta}$, the lines of expectation always point in the same general direction although their angles vary slightly with the mating types considered. Figure 2 shows these lines for five values of θ in double backcrosses with phase known and phase unknown. The figure was first drawn by J.H. Edwards. Actual linkage results can be plotted on it for comparison with the expectations (Edwards et al., 1972; Verma, 1974).

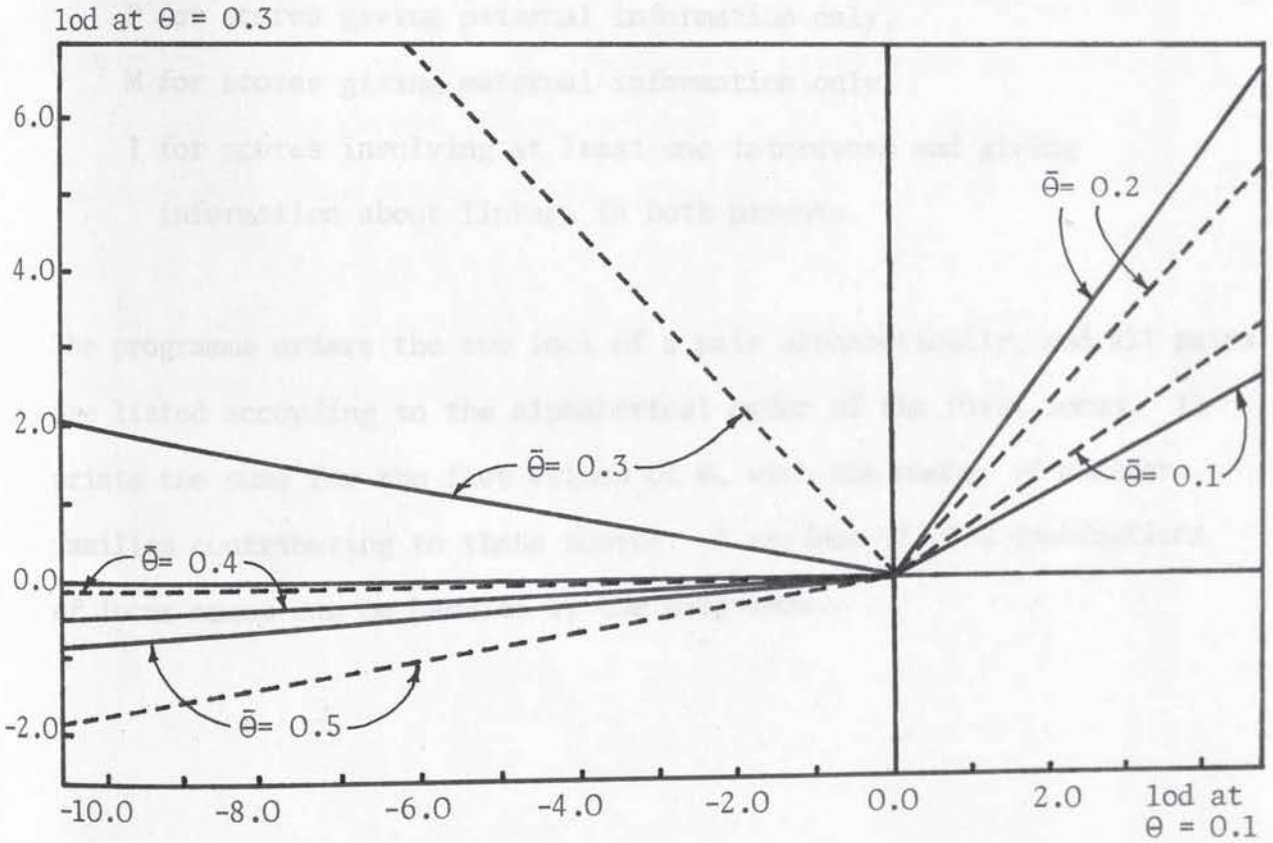


Fig. 2. Lod at $\theta = 0.3$ plotted against the lod at $\theta = 0.1$. The rays are the expected trajectories for the given true recombination fractions $\bar{\theta}$ when phase is known (- - -) and phase is unknown (———). Obviously, the more information there is, the further away actual points are from the origin.

2. ACCUMULATION

The lods punched in one or more runs of MARK III can be read as data and added up for each pair of loci by the programme LODSUM. The punched lods must be preceded by a control card telling the programme which type of lods are to be added. A single letter is punched in column 1 of the control card:

T if the sum of all scores is wanted,

P for scores giving paternal information only,

M for scores giving maternal information only,

I for scores involving at least one intercross and giving information about linkage in both parents.

The programme orders the two loci of a pair alphabetically, and all pairs are listed according to the alphabetical order of the first locus. It prints the sums for the five values of θ, with the number of nuclear families contributing to these scores. A maximum of 2852 combinations of locus names can be handled by the programme.

D. APPLICATION

The MARK III and LODSUM programmes have now been used for about a year in both Leiden and Winnipeg to calculate the lods of almost 400 pairs of loci. Most of the results have been presented at the second international workshop on human gene mapping, and include those of Marion Lewis on the PPH-Rh linkage, with strong suggestions that the loci for the enzyme uridine monophosphate kinase (UMPK) and the blood group Scianna (Sc) are also linked to the Rh blood group locus on chromosome no. 1, and the confirmation by A. Arnason of the linkage between the cholinesterase-2 locus (E_2) and the transferrin locus (Tf).

My contribution to these analyses has mainly consisted in adapting and installing the programmes, and in giving instructions on how to use them.

IV LINKAGE IN AUTOSOMAL TRISOMIES

Trisomies provide an additional way of observing the segregation of chromosomes and of the loci they carry. Apart from causing a specific syndrome, the extra chromosome should predictably affect the segregation pattern of the phenotypes defined at the loci it carries. Analysis of its effects and the detection of linkage depend on an understanding of the cytological mechanisms involved in trisomy. These are not fully covered by the literature which will be more easily reviewed after non-disjunction is studied in detail. In the model presented below, asynapsis, the non-pairing of two homologous chromosomes at meiosis, is implied, and considered together with cases of non-disjunction at the first meiotic division.

A. MODEL FOR NON-DISJUNCTION

Let us consider two homologous chromosomes associated in a tetrad of four chromatids at the diplotene stage of meiosis (Fig. 3).

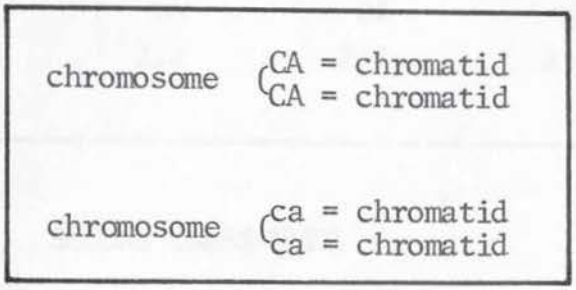


Fig. 3. Literal representation of two homologous chromosomes pairing at meiosis, one with upper case letters, the other with lower case. C and c define the centromere; A and a are alleles of a locus. For clarity the centromere is assumed to be a duplex structure, although the evidence for this is ambiguous and the terminology obscure.

If a cross-over occurs between a locus and its centromere, two of the four chromatids are recombinants, and two are non-recombinants (Fig. 4).

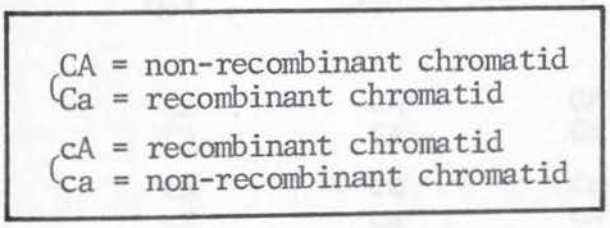


Fig. 4. Chromosomes of Fig. 3 after one cross-over between the locus studied and the centromere. Two of the four chromatids are now recombinants.

Providing the chromatids are involved at random, the average result will be the same for any number of cross-overs and there will be an equal number of apparent recombinant and non-recombinant chromatids (Fig. 5).

FIRST CROSS-OVER

Ca CA cA ca 2:2	Ca CA ca cA 2:2	CA Ca cA ca 2:2	CA Ca ca cA 2:2

SECOND CROSS-OVER

1	CA CA ca ca 0:4	Ca CA ca cA 2:2	CA Ca cA ca 2:2	Ca Ca cA cA 4:0
2	Ca CA cA ca 2:2	CA CA ca ca 0:4	Ca Ca cA cA 4:0	CA Ca ca cA 2:2
3	Ca CA cA ca 2:2	Ca Ca cA cA 4:0	CA CA ca ca 0:4	CA Ca ca cA 2:2
4	Ca Ca cA cA 4:0	Ca CA ca cA 2:2	CA Ca cA ca 2:2	CA CA ca ca 0:4

Fig. 5. Above the line are the four possible outcomes of a cross-over involving two non-sister chromatids. Under each of these are the four possible combinations resulting from a second cross-over. The numbers under each combination indicate the ratio 'recombinant : non-recombinant' chromatids. A simple count shows that recombinants and non-recombinants occur with equal frequencies.

The word chromatid will be used in the sense of the structure defined by a locus and its centromere to avoid repeated use of 'apparent'. Normally, a pair of homologous chromosomes contributes only one of its four chromatids to a gamete and, on average, there is an equal number of recombinant and non-recombinant gametes after one or more cross-overs.

In trisomy, however, a gamete receives two of these four chromatids. Such disomic gametes can thus display two recombinant chromosomes, two non-recombinant ones, or one of each kind (Fig. 6).

Chromosome type in the gamete	Gamete type			
	Unigrandparental		Bigrandparental	
Two recombinant chromosomes	Ca Ca	cA cA	Ca cA	
Two non-recombinant chromosomes	CA CA	ca ca	CA ca	
One of each	CA cA	ca Ca	CA Ca	ca cA

Fig. 6. Literal representation of all possible chromosome combinations in disomic gametes resulting from non-disjunctional events in the parental cells shown in Fig. 5.

A crucial distinction must be made between two types of gametes for which terms will have to be introduced. The first type carries two copies of a single locus originating from only one grandparent, and will be called unigrandparental. The second type carries both parental loci originating from two grandparents and will thus be called bigrandparental (Fig. 7).

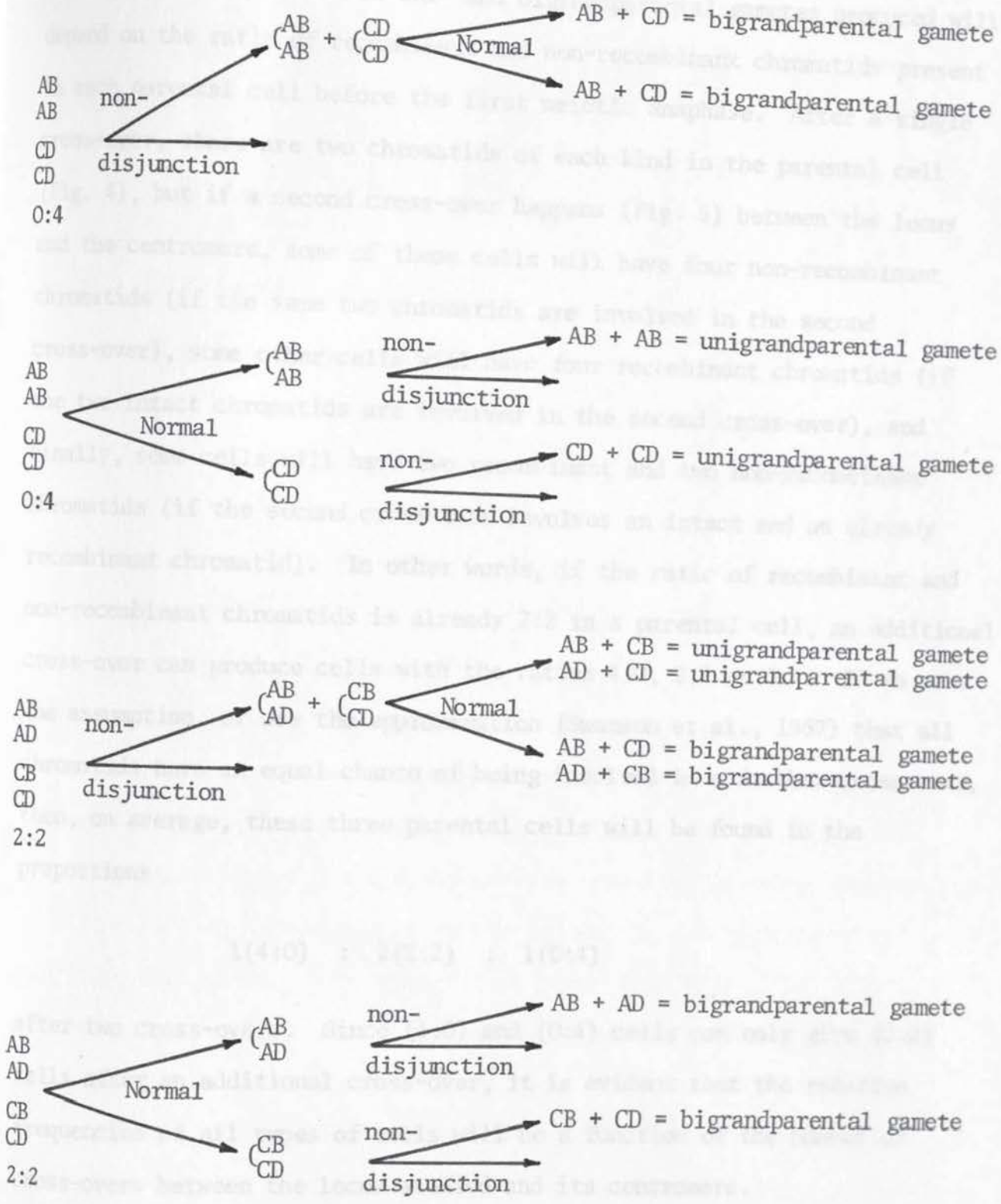


Fig. 7. Segregation of alleles B and D in disomic gametes at meiosis. A and C represent the centromere. Bigrandparental gametes are produced by half the (2:2) cells with non-disjunction at meiosis I, all (2:2) cells with non-disjunction at meiosis II, and all (0:4) and (4:0) cells with non-disjunction at meiosis I. Bigrandparental gametes are homozygous when the B and D alleles are identical, and heterozygous when the B and D alleles are different.

The relative frequencies of uni- and bigrandparental gametes procuced will depend on the ratio of recombinant and non-recombinant chromatids present in each parental cell before the first meiotic anaphase. After a single cross-over, there are two chromatids of each kind in the parental cell (Fig. 4), but if a second cross-over happens (Fig. 5) between the locus and the centromere, some of these cells will have four non-recombinant chromatids (if the same two chromatids are involved in the second cross-over), some other cells will have four recombinant chromatids (if the two intact chromatids are involved in the second cross-over), and finally, some cells will have two recombinant and two non-recombinant chromatids (if the second cross-over involves an intact and an already recombinant chromatid). In other words, if the ratio of recombinant and non-recombinant chromatids is already 2:2 in a parental cell, an additional cross-over can produce cells with the ratios 4:0, 2:2 or 0:4. If we make the assumption, or use the approximation (Swanson et al., 1967) that all chromatids have an equal chance of being involved in a further cross-over, then, on average, these three parental cells will be found in the proportions

$$1(4:0) \ : \ 2(2:2) \ : \ 1(0:4)$$

after two cross-overs. Since (4:0) and (0:4) cells can only give (2:2) cells after an additional cross-over, it is evident that the relative frequencies of all types of cells will be a function of the number of cross-overs between the locus studied and its centromere.

From what has been said, it can be seen that the proportion --defined as $P_n(2:2)$-- of (2:2) cells after exactly $\underline{n}$ cross-overs is equal to half the proportion of (2:2) cells after (n-1) cross overs, plus the proportion of (0:4) and (4:0) cells after (n-1) cross-overs. In terms of probability,

this becomes

$$P_n\ (2{:}2) = P_{n-1}\ (4{:}0 \text{ and } 0{:}4) + \frac{P_{n-1}\ (2{:}2)}{2}$$

$$= 1 - \frac{P_{n-1}\ (2{:}2)}{2}$$

where $P_o\ (2{:}2)$ is known to be nought.

In terms of series, this is equal to

$$P_n\ (2{:}2) = \sum_{i=1}^{n} (-\tfrac{1}{2})^{i-1}$$

$$= 2(1-(-2)^{-n})/3 = f_n.$$

The gametes therefore originate from a proportion f_n of (2:2) cells, and a proportion $(1-f_n)$ of (4:0) and (0:4) cells. In each case, the ratio of uni- and bigrandparental gametes depends on when non-disjunction happens. If we define as u and v the proportions of non-disjunctions occurring at divisions I and II of meiosis respectively (where u+v is equal to 1 if the locus and centromere are on the same chromosome; otherwise both u and v are zero), it is clear from Fig. 7 that (0:4) and (4:0) cells give u bigrandparental gametes while (2:2) cells give $(\frac{1}{2}u + v)$.

It follows that, in a large sample of disomic gametes having experienced exactly n cross-overs, the proportion y_n of bigrandparental gametes is expected to be

$$y_n = f_n\ (\tfrac{1}{2}u + v) + (1-f_n)u$$

$$= f_n\ (1-\tfrac{1}{2}u) + (1-f_n)u.$$

The maximum number of cross-overs can theoretically reach infinity but is necessarily much smaller in practice. The proportion of bigrandparental gametes expected after all possible numbers of cross-over --from zero to $\underline{n}$-- is

$$y = (1-2\theta)y_o + \sum_{i=1}^{n} a_i y_i$$

$$= (1-2\theta)u + \sum_{i=1}^{n} a_i y_i \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(1)$$

where 2θ is the proportion of parental cells experiencing at least one cross-over, a_i is the proportion of those experiencing $\underline{i}$ cross-overs, and $\sum_{i=1}^{n} a_i = 2\theta$.

The expected values for $\underline{y}$, calculated from this equation under various conditions, are given in Table 3.

Condition	y = proportion of bigrand-parental gametes
Absolute linkage	u
Obligatory single cross-over	1 - (u/2)
Always two cross-overs	(1/2) + (u/4)
Always lots of cross-overs (i.e. $\underline{n} \to \infty$)	2/3

Table 3. Proportion of bigrandparental gametes expected for a locus on a trisomic chromosome.

A glance at this table reveals that $\underline{y}$ could only be zero in two circumstances: the first happens when the locus studied is on the trisomic chromosome, is absolutely linked to the centromere (θ=0) and all non-

disjunctions occur at division II of meiosis (u=0). This combination of events is so unlikely that it only leaves the second possibility for $\underline{y}$ to equal zero: i.e. when the locus studied is not on the trisomic chromosome. This means that if $\underline{y}$ can be evaluated, its best estimate for any locus on a trisomic chromosome is, on average, bound to be different from zero, and the locus can therefore be assigned to that chromosome, no matter how far it lies away from its centromere, given sufficient data.

Since unigrandparental and bigrandparental gametes can only be distinguished when produced by heterozygous persons, the problem is then reduced to that of distinguishing between homozygous and heterozygous gametes.

In humans, however, searching for heterozygous disomic gametes is not realistic and the only source of information on their presence is that provided by the heterozygous trisomic individuals they produce. Heterozygous gametes will necessarily cause an increase in the proportion of heterozygous individuals, since every gamete will lead to a heterozygous phenotype.

B. LITERATURE REVIEW

Phenotypic shifts have always been assumed to happen in trisomics, and have been observed and documented in their progeny in organisms such as the Jimson weed (Datura) and the fruit fly (Drosophila). They were first mentioned for man in 1960 (Bateman, 1960) soon after the discovery of the extra chromosome in patients with Down's syndrome (Lejeune et al., 1959a, b, c), and were clearly defined algebraically by Crow (in Shaw and Gershowitz, 1962).

Following Lejeune's discovery, several authors started comparing the distribution of phenotypes in populations of trisomic patients with that in normal populations. This had already been done in 1953 by Lang-Brown et al. but the hope of mapping chromosome 21 was now added to that of detecting an association with the condition (Shaw and Gershowitz, 1962, 1963; Chown et al., 1962, 1965; Chown and Lewis, 1963; Kaplan et al., 1964; Hutton and Smith, 1964; Yokoyama, 1966; Evans et al., 1966, 1968; Goodman and Thomas, 1966, 1968; Hsia et al., 1969).

However, most studies did not make clear the distinction between association and linkage, as they mostly consist in trying to find out a deviation from the phenotypic proportions observed in disomic controls, without considering that a lack of heterozygotes would probably be a sign of association, but that an excess could be due to either association or linkage, the distinction being impossible without pedigree data. In fact, the distribution of parental mating types has rarely been taken into account, all parental information usually being ignored, even when available.

Misunderstanding and misinterpretation of significance tests have also added to the confusion. Most reports use samples that do not even approach the prohibitive sizes required by the X^2 tests performed (see Rundle, 1973), 'control' populations are often chosen indiscriminately, and one-way conclusions are drawn even when there is no evidence in favour of or against any locus on chromosome 21. As yet, no linkage has been advanced without being subsequently contradicted.

The lack of positive results has encouraged authors to publish new cytological models, most of which are incomplete and some not quite right. Patau et al. (1963) and de Grouchy (1965) went as far as assuming that most anomalies seen in trisomics were due to single genes located on the extra chromosome and produced chromosome maps of mendelian characters such as 'failure to thrive' and 'developmental retardation'. Penrose (1963) presented a model where disomic gametes received two loci at random out of the four present in the original tetrad of the parental cell. Saldanha and Beçak (1963) thought the recessive phenotype was perhaps only produced by a double or even a triple dose of the recessive allele. Goodman's model (1965) was more complete, covering non-disjunction in mitosis and in both meiotic divisions, crossing-over, dominance and codominance, dosage effects and overdose effects. However, he had more parameters than degrees of freedom and was forced to admit with Thomas in 1966 that

> 'In the absence of rigorous cytological evidence regarding frequency of chiasma formation and the stage or stages at which non-disjunction occurs, there is little basis for choosing one model over another.' (Goodman and Thomas, 1966)

In 1968, they finally concluded:

> 'The results of this study should discourage the application of gene frequency analysis to allelic systems exhibiting complete dominance in an effort to establish linkage of trisomic chromosomes' (Goodman and Thomas, 1968).

Meanwhile, quantitative biochemical studies were going on (Trubowitz et al., 1962; Brandt, 1962; O'Brien and Groshek, 1962; Alter et al., 1962; Cox, 1965; Baikie et al., 1965; Rosner et al., 1965; Yokoyama, 1966; Berman et al., 1967; Nadler et al., 1967; Pantelakis et al., 1970) and were best summed up by Lawler (1962):

> 'In mongols almost any physiological measurement gives anomalous results and it cannot be automatically assumed that the direct effects of gene action are being measured'.

The various effects of different trisomies on the same proteins in barley (McDaniel and Ramage, 1970) show remarkably well that association does not necessarily imply linkage.

More recent frequency analyses on trisomy 21 have reached higher degrees of statistical complexity (Ball et al., 1972; Rittner and Schwinger, 1973a, 1973b; Rundle, 1973; Rundle and Sudell, 1973) but contradictions still occur, and at present the only two loci known with certainty to be on chromosome 21 --superoxide dismutase-1 and anti-viral protein-- have been found by using somatic cell hybridization while the possible linkage of the Ag lipoprotein to a marker chromosome 21 has been suggested after the study of one pedigree showing a ratio of either 8:0 or 5:3 of non-recombinants to recombinants (New Haven Conference, 1973).

Further investigations on the phenotypic proportions of trisomic populations are unlikely to yield clearer results, and the development of cytological

techniques is a more obvious way of trying and improving chromosome maps. However, studies aimed at detecting an excess of heterozygous trisomics born to heterozygous parents is still indicated, not only because they make possible the detection of unknown linkages, some of which cannot be tested in vitro, but also because they can extract information on what goes wrong in meiosis in the case of loci already known to be on a trisomic chromosome.

This approach was first presented for autosomes at the fourth international congress of human genetics in Paris (Lockie and Edwards, 1971), and the method of analysis with W-arrays was later outlined by J.H. Edwards (1972).

C. PROBABILITY MODELS

Although the analysis could be rapidly programmed by using W-arrays as proposed by J.H. Edwards (1972), local computer difficulties prevented me from doing so and included intractable system defects on the UNIVAC 418-III of the Queen Elizabeth Hospital, and the change over of the University computer from an ICL KDF9 to a 1906A. Since no IBM 370 was freely available in this country and that data on trisomies were immediately available, I had to undertake a less efficient analysis restricted to two alleles per locus, and do the programming for a smaller IBM 1130 computer that I could operate myself. The efficiency attained is unlikely to be less than the sum of the proportions of the two commonest alleles. Provided no propositus is ascertained through both the centromeric phenotype (trisomy) and the phenotype at the locus, any estimate of their relationship will be unbiassed.

1. ASSUMPTIONS

Let us make the basic assumption of codominance when there are three normally non-silent alleles, Although inactivation of the extra chromosome is a formal possibility, the type of inactivation associated with an X chromosome in the female is known not to occur with chromosome 13, 18 and 21 from the absence of any conspicuously late labelling and, in trisomy 13 and 18, from the absence of related chromatin body (J.H. Edwards, personal communication). (In trisomy 21 an inactive chromosome might be difficult to identify due to its small size.) Under this assumption, the first step in estimating the proportion of heterozygous trisomics produced by heterozygous gametes is to exclude all known mosaics from the analysis. Trisomics born to homozygous

parents should also be excluded, as they can only contribute false information, if any. The effects of random deviations from the expected proportions of parental phenotypes can be eliminated by stratifying the anlaysis according to parental mating class.

In a first attempt, aimed at mapping chromosome 21, translocation cases can be included as they would experience the same phenotypic shifts as trisomics. At any rate, they constitute a minority of a few percent (Wright et al., 1967; Mikkelsen, 1970; Uchida, 1970) and will not drastically affect the results, especially if significance tests are not used on small samples. Considering the strong maternal age effect, and the fact that a paternal origin of the extra chromosome has been shown in only two out of sixteen informative cases (Sasaki and Hara, 1973; Uchida, 1973 and personal communication), one can use the rough approximation that all extra chromosomes are maternal, and restrict the analysis to matings with heterozygous mothers. This greatly simplifies the algebra and does not seriously affect conclusions on evidence of linkage if one adopts the likelihood approach.

The conclusions are thus based on the relative phenotypic proportions of affected children in various mating types, and therefore, on the double assumption that trisomy does not alter the quality of any of these phenotypes, and that no allele of any locus studied has any substantial effect on the manifestation of the syndrome.

Obviously, the phenotypes observed in children are assumed to be produced by the children's own alleles and not to be due to transfused blood or maternal plasma proteins having passed through the placental barrier.

2. CODOMINANT SYSTEMS

Heterozygous mothers can produce - at least theoretically - both uni- and bigrandparental gametes. Let us call b the proportion of their trisomic offspring coming from bigrandparental gametes. These affected individuals carry alleles from three of their grandparents; their expected proportions and observed frequencies are given in Table 4 for the three informative mating classes. All three genotypes can be dist-

Mating class		expected phenotypes			observed phenotypes		
Mother	Father	gg	gh	hh	gg	gh	hh
gh	gg	c/2	b+c/2	-	E	F	-
gh	gh	c/4	b+c/2	c/4	G	H	I
gh	hh	-	b+c/2	c/2	-	J	K

Table 4. Phenotypes of trisomic children in three mating classes for a system of two codomimant alleles g and h. The parameter b (b= 1.0-c) is the proportion of trisomics who have received three alleles from three grandparents, two of whom are maternal.

inguished for codominant systems, and the likelihood of observing any particular set of data is then

$$L = \text{constant} \times (c/2)^{E+K}(b+c/2)^{F+H+J}(c/4)^{G+I}$$

and its natural logarithm

$$\ln L = \text{constant} + (E+K)\ln(c/2) + (F+H+J)\ln(b+c/2) + (G+I)\ln(c/4)$$

presents no difficulty.

If the assumptions hold, the evaluate of b should be zero when the locus studied is not on the trisomic chromosome, and it should have some positive value, which it is impossible to predict in our present state of knowledge, if the locus is on that chromosome.

A drawing of the support curve for all values of b will clearly show the state of information on any locus, and assist judgement on whether the null hypothesis of no linkage remains unchallenged because it is true or because the sample is too small. As explained in Chapter II, the curve can also be summarized by its evaluate and the two values of b between which the support is less than two units below the maximum.

3. DOMINANT-RECESSIVE SYSTEMS

Dominant-recessive pairs of alleles are more difficult to analyse since the propositus' genotype can never be known with certainty if dominant, while maternal heterozygosity can only be identified by the presence of a recessive child. Restricting the analysis to families with recessive children would introduce difficult problems of bias. Because the presence of dominant-recessive alleles on a trisomic chromosome will cause a decrease in the frequency of recessive trisomics, a practical solution is to consider each family with a dominant mother separately, and calculate the probability r of her having a recessive trisomic child, taking into account the information given on her genotype by the allelic proportions of the population and the phenotypes of her husband and her normal children which, by nature, are assumed to have normal segregation. Extending the analysis to mating classes where one or both parents are unknown will also add some information.

Table 5 shows the expected proportions of recessive trisomics for all combinations of parental genotypes, in the case of a dominant-recessive pair of alleles T and t. The proportions are defined in terms of c.

Parental genotypes		r
Mother	Father	
TT	TT	0
Tt	TT	0
tt	TT	0
TT	Tt	0
Tt	Tt	c/4
tt	Tt	$\frac{1}{2}$
TT	tt	0
Tt	tt	c/2
tt	tt	1

Table 5. Proportion r of recessive trisomics expected in each combination of parental genotypes. The parameter c (= 1.0 - b) is the proportion of trisomics who do not receive three alleles from three grandparents.

Only two combinations of parental genotypes are informative: H_1 (:Tt mother x Tt father) and H_2 (:Tt mother x tt father). But the genotypes of dominant individuals must be inferred, and one must find the probability that the parents' phenotypes represent one of these two combinations in order to calculate the probability r_1 or r_2 of their having a recessive trisomic child.

When both parents are dominants, the probability that they are both heterozygotes will depend on their normal children's phenotypes. If one of the children is recessive, both parents must be heterozygotes. If no child is recessive, the more numerous they are, the more unlikely it is that the parents are heterozygotes. What we want to know is $P(H_1|S)$, the probability that both are heterozygotes, given S, their normal children's phenotypes. This is

$$P(H_1|S) = \left(P(H_1) \times P(S|H_1)\right) / \left(P(H_1) \times P(S|H_1) + P(N_1) \times P(S|N_1)\right)$$

where N_1 represents any combination of genotypes where at least one parent is not heterozygous.

We know that $P(H_1) = (2pq)^2/(1-q^2)^2 = (2q/1+q)^2$ where $\underline{p}$ and $\underline{q}$ are the allele proportions of T and t.

$P(S|H_1) = (\frac{3}{4})^n$ where $\underline{n}$ is the number of children in the sibship S of normal children.

$P(N_1) = 1-P(H_1) = 1-(2q/1+q)^2$.

$P(S|N_1) = 1$ if there is no recessive child.

$= 0$ if there is at least one recessive child.

Therefore, $P(H_1|S) = 1$ if there is at least one recessive child.

When all children are dominants,

$$P(H_1|S) = (2q/1+q)^2(\tfrac{3}{4})^n / (2q/1+q)^2(\tfrac{3}{4})^n + 1-(2q/1+q)^2$$

$$= 3^n q^2 / 3^n q^2 + 4^{n-1}(1+2q-3q^2) .$$

When one parent is recessive, one can also find the probability $P(H_2|S)$ that the other dominant parent is heterozygous, given their children's phenotypes.

In this case, $P(H_2) = (2pq/(1-q^2)) = 2q/(1+q)$

and $P(S|H_2) = (n!/D!R!)(\frac{1}{2})^D(\frac{1}{2})^R$ where D and R are the numbers of dominant and recessive disomic children in the sibship S.

$P(N_2) = p^2/(1-q^2) = (1-q)/(1+q)$.

$P(S|N_2) = \underline{x} = 1$ if $R=0$

$= 0$ if $R \geq 1$.

Therefore,

$$P(H_2|S) = (2qn!/(1+q)D!R!)(\tfrac{1}{2})^{D+R} / (qn!/(1+q)D!R!2^{n-1} + \underline{x}(1-q)/(1+q))$$

$$= qn!/(qn! + D!R!2^{n-1}\underline{x}(1-q))$$

$= 1$ when $R \geq 1$

$= q/(q+2^{n-1}(1-q))$ when $R=0$.

When one or both parents are unknown the probability r_i that they produce recessive trisomic children can most easily be defined in terms of r_1 and r_2 as shown in Table 6. The likelihood of the

Phenotypic mating class		r= probability of having a recessive trisomic child.
Mother	Father	
T	T	$r_1 = c \cdot P(H_1 \mid S)/4$
T	t	$r_2 = c \cdot P(H_2 \mid S)/2$
T	?	$r_3 = (1-q^2)r_1 + q^2 r_2$
?	T	$r_4 = (1-q^2)r_1 + q^2 \cdot P(H_2 \mid S)/2$
?	t	$r_5 = (1-q^2)r_2 + q^2$
?	?	$r_6 = (1-q^2)r_4 + q^2 r_5$

Table 6. Expected proportion of recessive trisomic children in all informative mating classes, for a dominant-recessive pair of alleles, assuming that all extra chromosomes have a maternal origin.

observed phenotypes in trisomics belonging to the same family is $L = \left((A+B)!/A!B!\right) r^A (1-r)^B$ where A and B are the numbers of recessive and dominant trisomics in the family. Each family has to be computed separately. The logarithms of their likelihoods can then be added on a computer to produce a graphical output as in the case of codominant systems.

D. COMPUTERIZATION

Computer programmes were written in FORTRAN for an IBM 1130 to do the analysis described in the previous section. The programmes are listed in Appendix I.

Codominant systems are analysed by the programme GTRIC:
The card input consists of

a) 21 cards containing the log likelihoods that were given by a previous run to the 101 values of b, from 1.00 to 0.00, and that are to be added to the current run. The FORTRAN format is (5E13.6), and the cards are left blank if there are no previous logs.
b) A 22nd card containing the phenotype frequencies E, F, G, H, I, J and K of table 4, followed by the name of the locus studied. The format is (7I3, 51X, A3).

The output consists of the total log likelihoods for all values of b. These are

a) punched on 21 cards in the format used for input,
b) printed with the input phenotype frequencies, the evaluate of b, and the name of the locus,
c) standardized with the maximum support at 0.00, and plotted against b if they are in the range (0.00, -5.00).

Dominant-recessive systems are analysed by the programme GTRID which uses the subroutines GREAD and GPHEN. The card input consists of

a) A first card containing the name of the locus studied, two codes representing the dominant phenotype, one code for the recessive phenotype, one code for an unknown phenotype, and the proportion of the dominant allele in the normal disomic population. The FORTRAN format is (A3, 1X, 4A2, F3.3).

b) 21 cards with the log likelihood of each value of b, as for codominant systems.

c) The pedigree data coded as instructed in a) above.
There is one card per individual, the cards are grouped in families and in this order for each family: father, mother, trisomic propositus, sibs, and finally a blank card. The format is variable and is given in routine GREAD which must be updated each time a new locus is investigated.

The punched and plotted outputs are identical to those of codominant systems. The printed output gives a list of the items on the first input card, the total log likelihoods for all values of b, the evaluate of b, and the frequencies of dominant and recessive trisomics observed in each mating class.

Various sets of log likelihoods can be added together by the programme GSUPP which plots the overall results and prints the totals in tabular form with the name of the locus, the evaluate and the 2-unit support limits of b. GSUPP accepts n sets of 21 cards, preceded by a data card containing the value of n and that of an index which takes the value 1 when the total log likelihoods are desired on 21 punched cards. The format used for this first card is (2I2).

E. DATA ON TRISOMY-21

The method has been tested and used on 821 non-mosaic, non-negro patients with Down's syndrome, their parents and sibs. One hundred and forty-eight were taken from the paper by Lang-Brown et al. (1953), and 164 from that by Evans et al. (1966). The rest was kindly made available to Prof. J.H. Edwards, and consisted of 337 Manitoban cases collected by Dr. Irene A. Uchida and provided by Dr. B. Chown of Winnipeg, Canada, 70 cases sent by Dr. E.W. Lovrien of Portland in Oregon, U.S.A., 56 cases received from Dr. A.D. Merritt and Dr. P.M. Conneally of Indianapolis in Indiana, U.S.A., and 46 cases from Prof. M.A. Ferguson-Smith of Glasgow, Scotland.

Table 7 shows the loci investigated at each of these six sources. Rh, Fy, AK, Amy-2, PGM_1 and 6PGD are already known to be on chromosome 1, ACP1 on chromosome 2, Hp on 16, ADA on 20 (see Chapter I), and MN possibly on chromosome 2 (New Haven Conference, 1973), but they were included in this analysis as controls.

The systems were divided into codominant and dominant-recessive ones, as shown in Table 8.

The familial phenotypic data are summarized in Tables 9 and 10. Details on the data can be found in Appendix III.

Locus	Lang-Brown et al. (1953)	Evans et al. (1966)	Manitoba	Oregon	Indiana	Scotland (Glasgow)
ABO blood group	✓	✓	✓	✓	✓	✓
Secretor status (Se)		✓		✓		
MNSs blood group	✓	✓	✓	✓	✓	✓
P blood group	✓	✓	✓	✓	✓	✓
Rh blood group	✓	✓	✓	✓	✓	✓
Kell blood group	✓	✓	✓	✓	✓	✓
Duffy blood group (Fy)			✓	✓	✓	✓
Kidd blood group (Jk)			✓	✓	✓	
Lutheran blood group (Lu)		✓	✓	✓		
Group specific component (Gc)			✓	✓		✓
Haptoglobin (Hp)			✓	✓	✓	✓
Lipoprotein (Lp)				✓		
Phosphoglucomutase-1 (PGM_1)					✓	✓
Red cell acid phosphatase (ACP1)					✓	✓
Adenylate kinase (AK)				✓		✓
Adenosine deaminase (ADA)				✓		✓
Ability to taste phenyl-thiocarbamide (PTC)				✓		
Pancreatic amylase (Amy-2)				✓		
Gm gammablobulin marker			✓	✓		
Inv gammaglobulin marker			✓			
6-phosphogluconate dehydrogenase (6PGD)				✓		✓
Pseudocholinesterase E_1						✓
Transferrin (Tf)						✓

Table 7. Origin of the trisomic data.

	Codominant systems	Dominant-recessive systems	p
ABO		not O (i.e. A, B and AB) and O	0.34
Se		Se and se	0.52
MN	M and N		
P		P_1+ and P_1-	0.54
Rh	C and c		
Kell	K and k	K+ and K-	0.04
Fy	a and b	a+ and a-	0.42
Jk	a and b	a+ and a-	0.51
Lu	a and b	a+ and a-	0.04
Gc	1 and 2		
Hp	1 and 2		
Lp		a+ and a-	0.39
PGM_1	1 and 2		
ACP1	a and b		
AK	1 and 2		
ADA	1 and 2		
PTC		Taster and non-taster	0.45
Amy-2	a and b		
Gm	1 and 5		
Inv		1+ and 1-	0.20
6PGD	a and b		
E_1	U and A		
Tf	C and B		

Table 8. Names of the two alleles considered for all the systems studied, and proportion p of the dominant alleles (for dominant-recessive systems) in the normal disomic population.

Mother	ab							aa				bb				Total
Father	aa		ab			bb		aa	ab		bb	aa	ab		bb	
Trisomic	aa	ab	aa	ab	bb	ab	bb	aa	aa	ab	ab	ab	ab	bb	bb	
Locus (a,b)																
MN (M,N)	45	43	42	85	44	38	38	64	49	54	43	50	22	31	17	665
Rh (C,c)	28	35	39	87	49	53	48	16	29	23	44	36	49	41	91	668
Kell (K,k)	0	0	0	1	0	10	10	0	0	0	0	0	13	20	389	443
Fy (a,b)	16	19	27	38	27	29	37	11	19	12	15	16	20	33	49	368
Jk (a,b)	23	18	12	40	33	20	28	31	22	30	20	25	21	23	12	358
Lu (a,b)	0	0	1	2	0	9	12	0	0	0	0	0	14	8	282	328
Gc (1,2)	20	4	3	4	5	0	3	31	11	4	5	3	1	1	1	96
Hp (1,2)	6	5	17	17	8	18	12	2	4	3	7	5	14	15	25	158
PGM_1 (1,2)	9	7	2	3	1	2	0	20	7	6	2	2	1	0	1	63
ACP1 (a,b)	1	0	3	9	1	7	5	0	1	2	3	5	5	9	7	58
AK (1,2)	6	2	1	0	0	0	0	86	6	3	0	0	0	0	0	104
ADA (1,2)	2	6	0	0	0	0	0	39	0	1	0	0	0	0	0	48
Amy2 (a,b)	4	1	1	0	0	0	0	1	3	1	0	0	0	0	0	11
Gm (1,5)	4	9	6	23	7	20	15	0	6	3	7	8	17	9	42	176
E_1 (U,A)	1	1	0	1	0	0	0	35	2	1	0	0	0	0	0	41
Tf (C,B)	1	0	0	0	0	0	0	40	1	0	0	0	0	0	0	42
6PGD (a,b)	0	1	0	0	0	0	0	107	0	2	0	0	0	0	0	110

Table 9. Phenotypic frequencies of propositi with Down's syndrome, in all parental mating classes for the codominant systems.

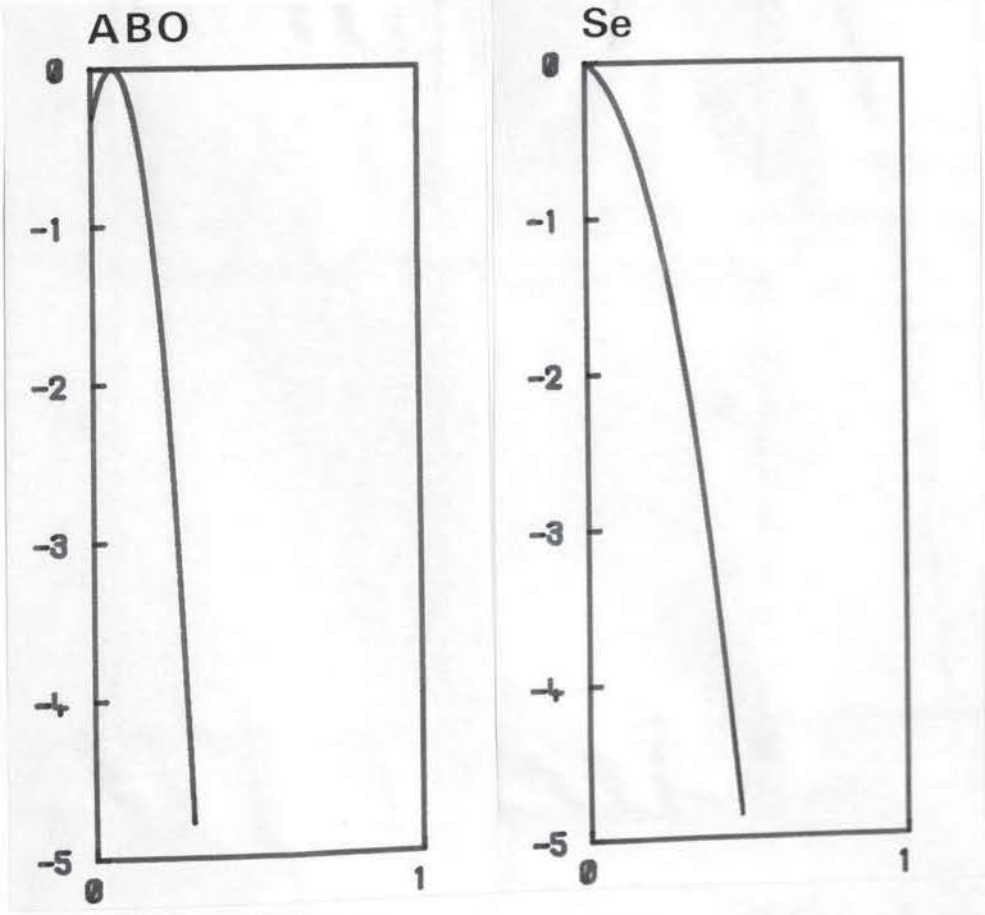
Mating class	Mother	Father	ABO		Se		P		K		Fy(a)		Jk(a)		Lu(a)		Lp(a)		PTC		Inv	
			T	t	T	t	T	t	T	t	T	t	T	t	T	t	T	t	T	t	T	t
1	T	T	181	29	51	3	327	42	0	1	36	3	10	2	0	0	9	5	0	0	1	0
2	T	t	96	53	17	7	74	45	3	5	11	10	3	0	0	0	1	5	1	2	1	2
3	T	?	11	2	7	4	29	2	2	0	0	0	0	0	0	0	2	0	1	0	0	0
4	?	T	3	1	4	1	9	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
5	?	t	3	3	0	0	1	1	0	9	0	0	0	0	0	0	0	0	0	0	0	0
6	?	?	46	48	63	20	71	20	16	77	0	0	0	0	8	74	2	1	0	0	0	0
7	t	T	91	66	15	6	68	49	10	6	13	5	1	2	0	0	4	1	1	1	0	0
8	t	t	-	154	-	5	-	30	-	174	-	19	-	1	-	0	-	4	-	0	-	0
9	t	?	3	24	4	1	3	4	0	27	0	1	0	0	0	0	0	1	0	1	0	0
Totals:			434	380	161	47	582	193	31	299	60	39	14	5	8	74	18	17	3	4	2	2
			814		208		775		330		99		19		82		35		7		4	

Table 10. Frequencies of phenotypes of propositi with Down's syndrome, in all parental mating classes for the dominant-recessive systems studied. T and t designate the dominant and recessive alleles respectively. The four propositi tested for Inv were more than one year old.

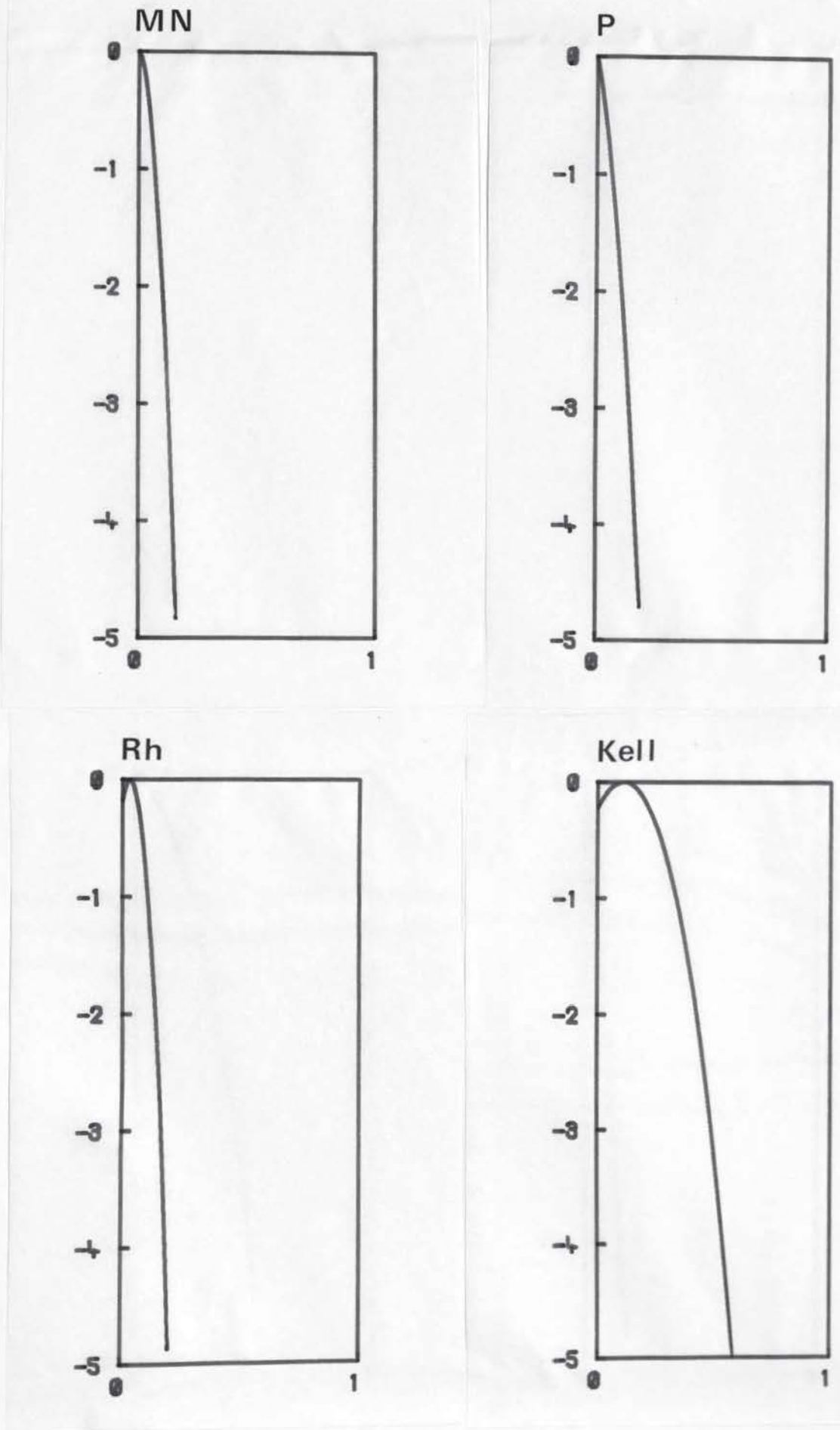
F. RESULTS AND DISCUSSION

First, it is obvious from Table 9 that nothing can be said about linkage with the loci for cholinesterase E_1, transferrin and 6-PGD since only three, one and one trisomics respectively are informative, although up to 110 of them were tested. All the other loci were analysed although again the sample sizes were often too small for decisive conclusions to be envisaged. The results, details of which can be found in Appendix IV, are shown in Figure 8 where the support for each locus is plotted against the probability $\underline{b}$ that individuals with Down's syndrome carry alleles from both maternal grandparents. All these support curves are summarized in Table 11.

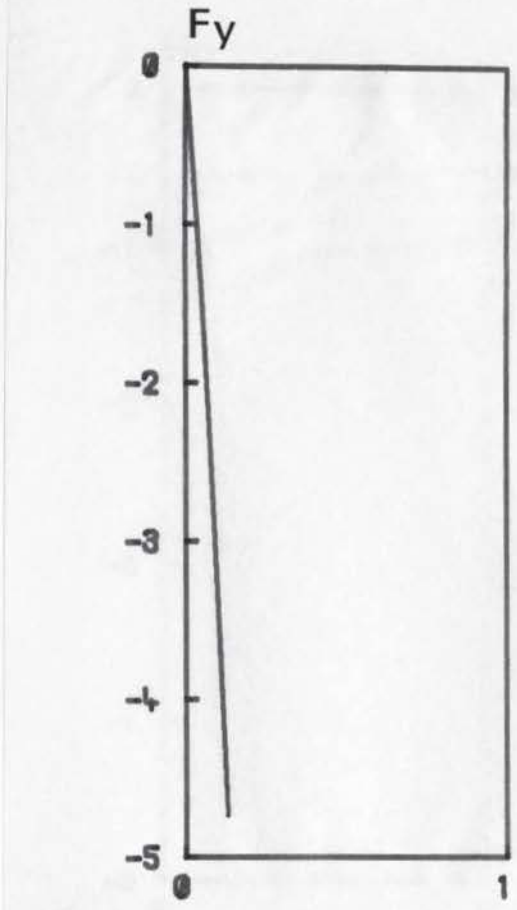
Fig. 8. Support plotted against the probability $\underline{b}$ that individuals with Down's syndrome carry alleles of a given locus from both maternal grandparents.

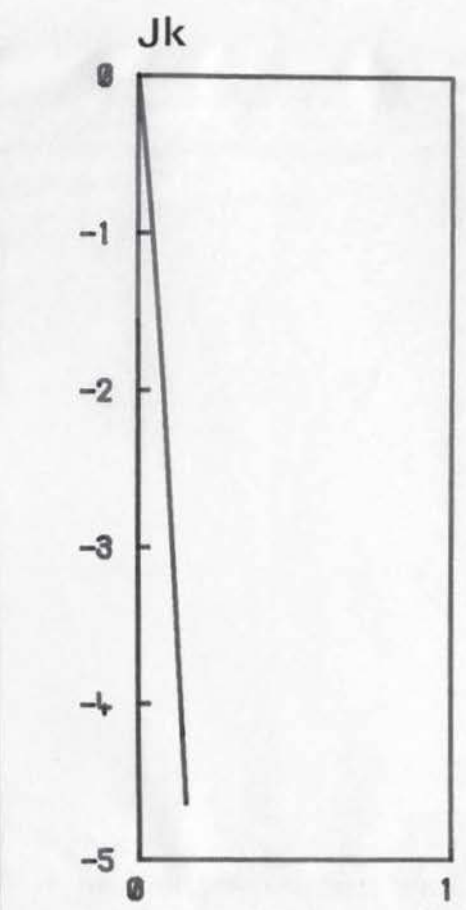


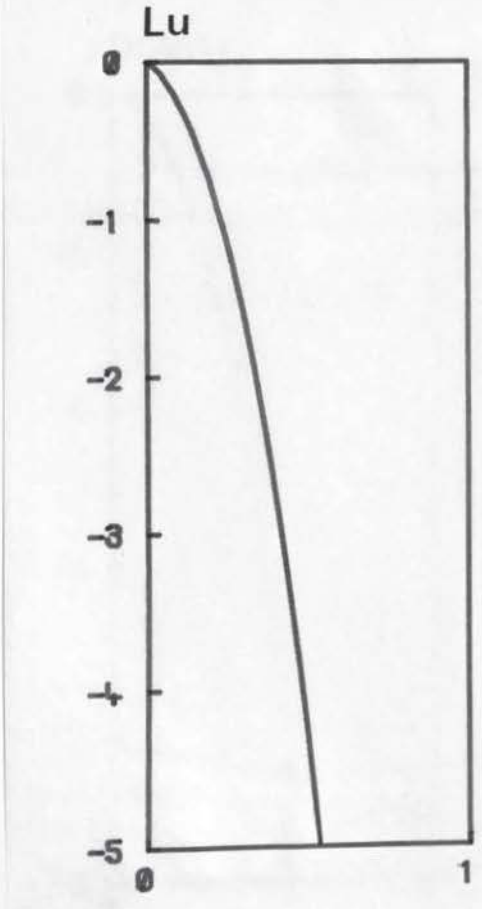
cont'd/..

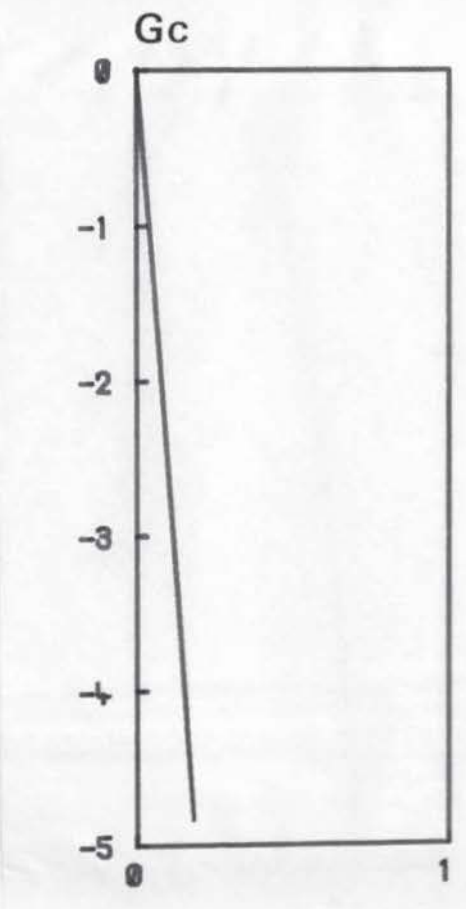


cont'd/..

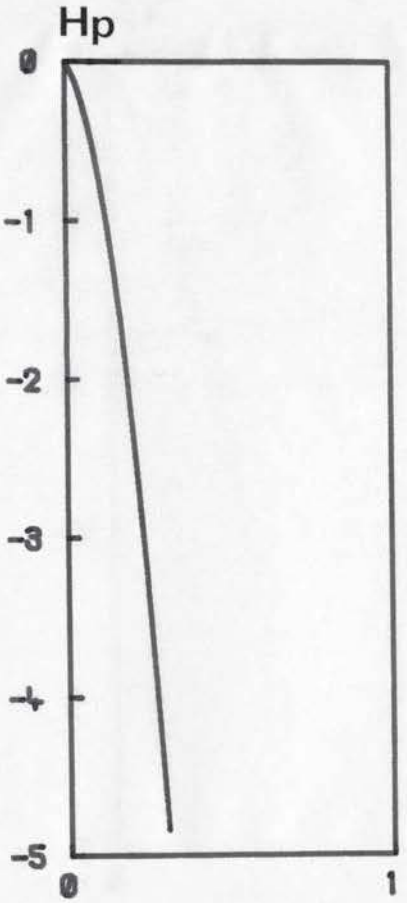


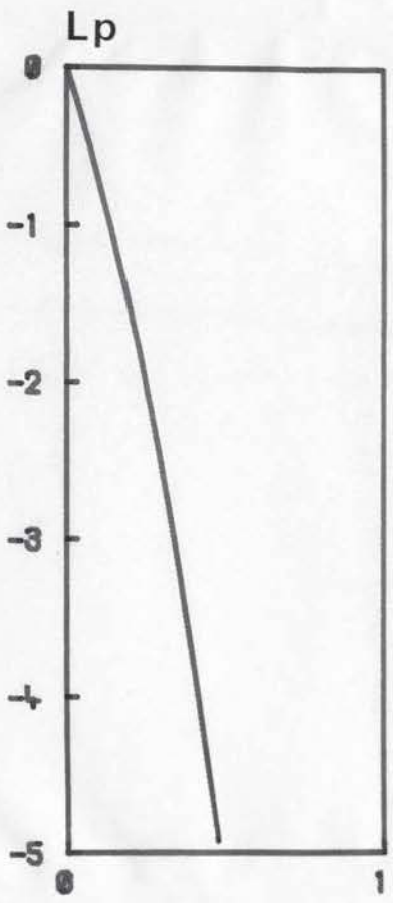


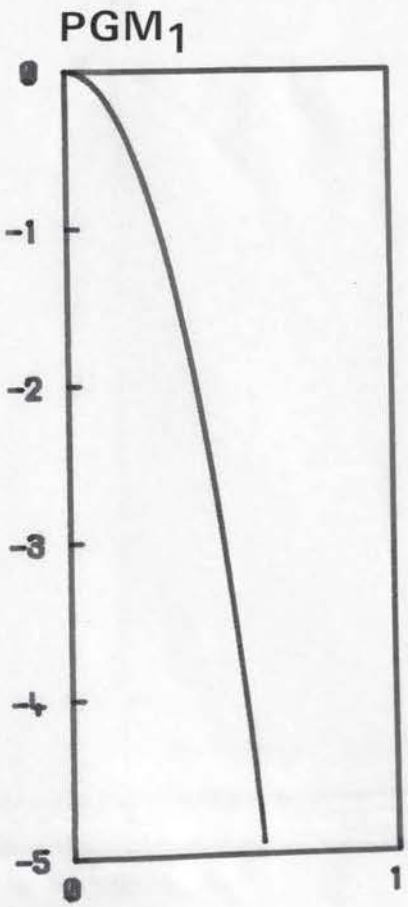


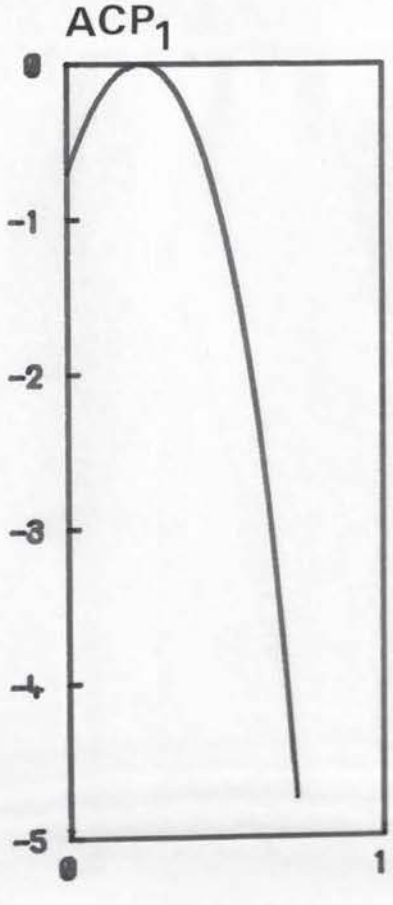


cont'd/..

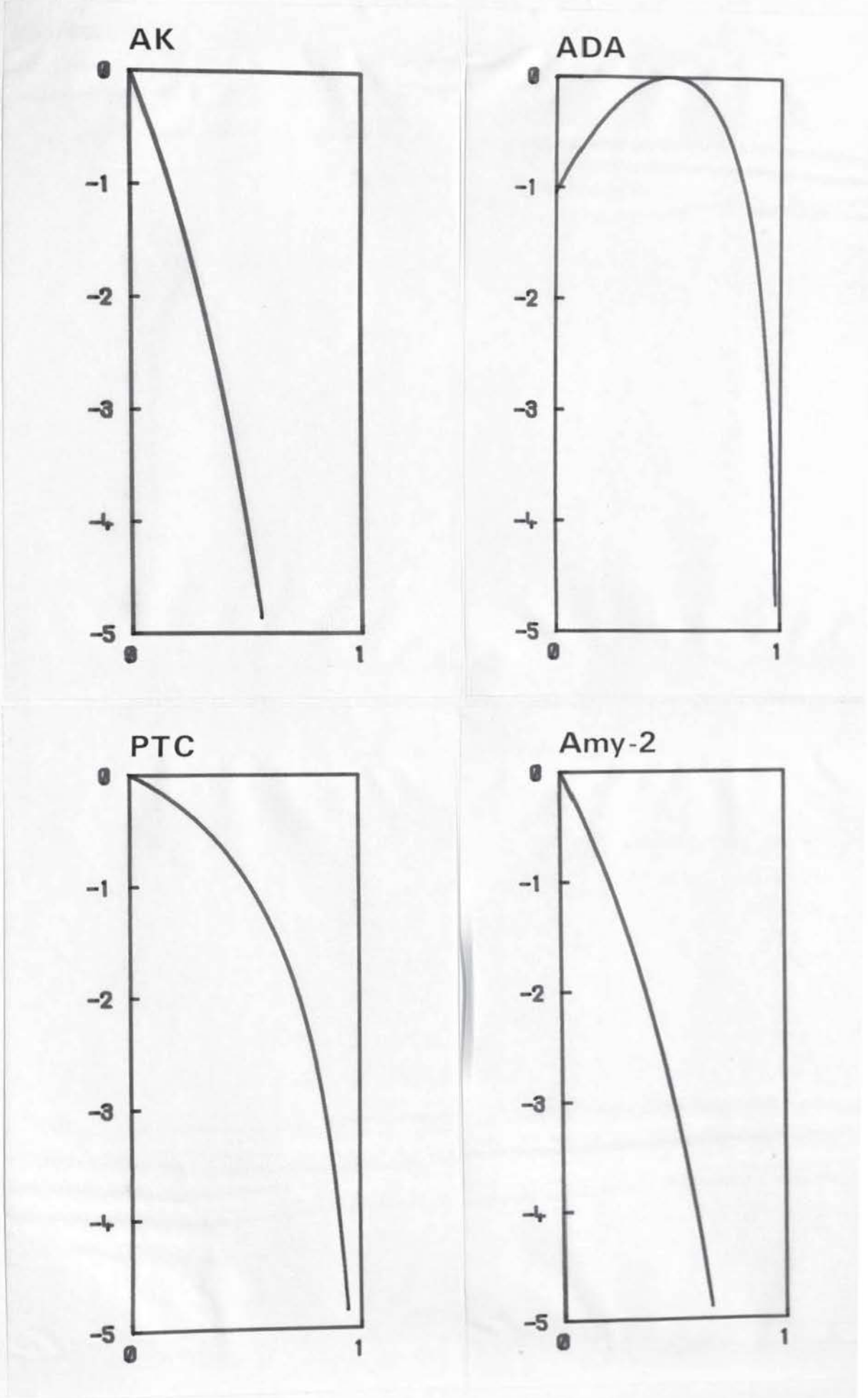






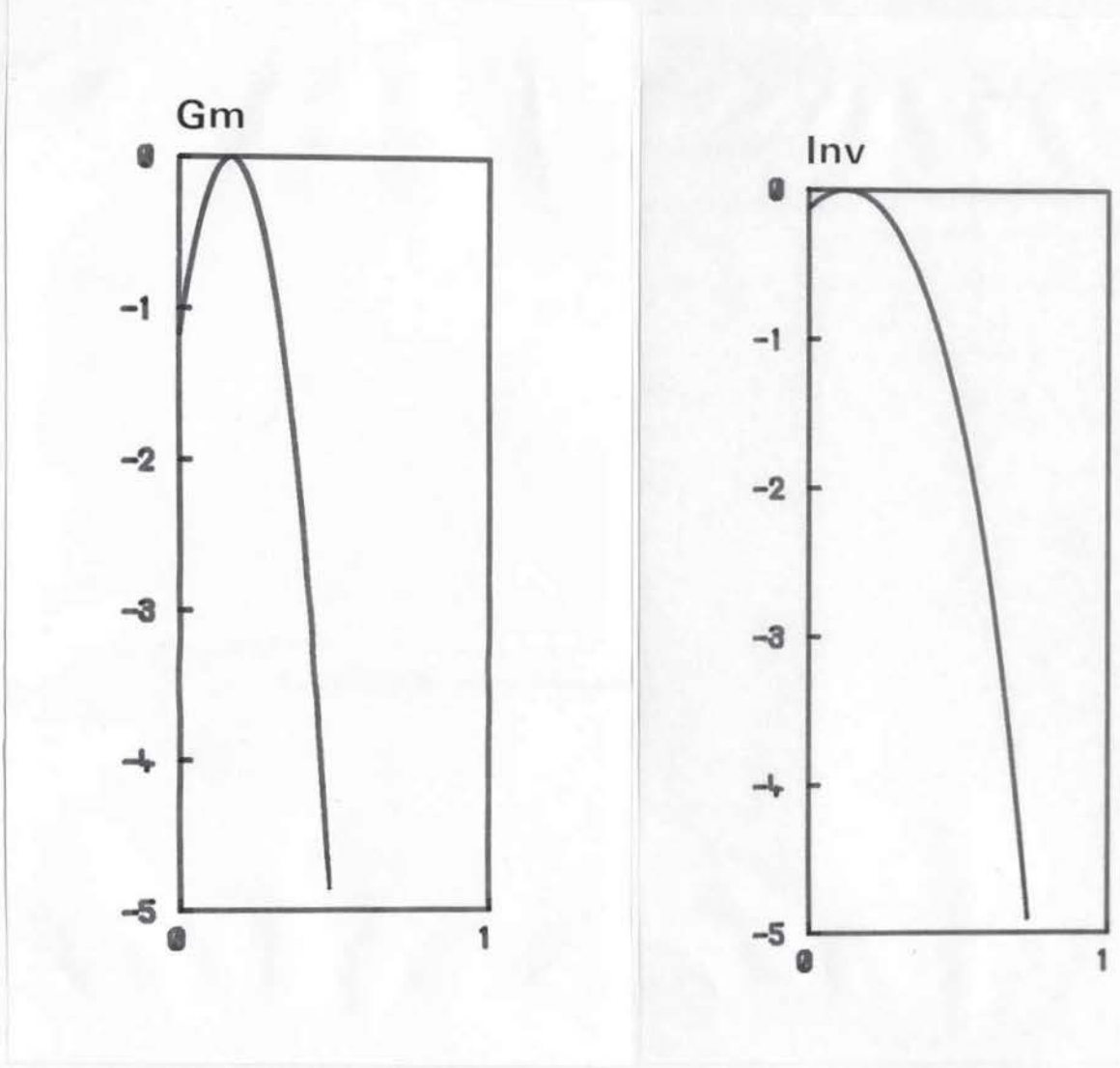


cont'd/..



cont'd/.

../cont'd



Locus	Evaluate of $\underline{b}$	2-unit support limits
ABO	0.06	(0.00, 0.21)
Se	0.00	(0.00, 0.31)
MN	0.00	(0.00, 0.10)
P	0.00	(0.00, 0.10)
Rh	0.03	(0.00, 0.15)
Kell	0.10	(0.00, 0.43)
Fy	0.00	(0.00, 0.07)
Jk	0.00	(0.00, 0.09)
Lu	0.00	(0.00, 0.34)
Gc	0.00	(0.00, 0.09)
Hp	0.00	(0.00, 0.18)
Lp	0.00	(0.00, 0.25)
PGM_1	0.00	(0.00, 0.40)
ACP_1	0.23	(0.00, 0.58)
AK	0.00	(0.00, 0.31)
ADA	0.50	(0.00, 0.92)
PTC	0.00	(0.00, 0.74)
Amy 2	0.00	(0.00, 0.37)
Gm	0.17	(0.00, 0.37)
Inv	0.12	(0.00, 0.56)

Table 11. Summary of the support curves of Figure 8. For each locus the evaluate of $\underline{b}$ is given with the two values of $\underline{b}$ between which the support is less than two units below the maximum.

Since one does not know what the evaluate would be if a locus were on chromosome 21, one has to be rather conservative in dividing the systems analysed into three categories:

1) Those which can safely be said not to be anywhere on chromosome 21: ABO, MN, P, Rh, Fy, Jk, Gc, Hp and Lp.
2) Those that are probably not on chromosome 21, although the evidence is not absolutely conclusive: Se, Lu and AK.

3) Those where the data are not sufficient to firmly support any final conclusion in favour of or against linkage:
Kell, PGM_1, ACP1, ADA, PTC, Amy-2, Gm, Inv, 6PGD, E_1 and Tf.

This is in perfect agreement with the known chromosome assignments of ten of these loci (see Chapter I and previous section), as well as the a priori expectation that none of the thirteen other loci would be on a chromosome that represents only 1.63% of the haploid autosomal set (Paris Conference, 1971). Since there was enough information on only five of these loci (ABO, P, Jk, Gc and Lp), the probability of finding at least one of them on chromosome 21 was only $(1.0 - (1.0 - 0.0163)^5) = 0.08$. The findings of this method thus corroborate the negative results obtained by the earlier trisomy analyses already mentioned, and have the additional advantage of clearly showing which results could be due to lack of information.

The reliability of the method can also be tested by investigating gamma-globulin markers in infants. The curve for Gm in Figure 8 takes into account only propositi who were over six months old, since gammaglobulins are known to cross the placental barrier during pregnancy and could possibly be detected in the child up to the age of six months. When the method is applied to Gm in infants, the assumption that one is testing the child's own chromosome products does not hold and the evaluate shows the probability of identifying the maternal serum phenotype in the child. This probability is clearly 1.0 at birth, as can be seen in Figure 9a, and it gradually decreases with age towards zero (Fig. 9b, c, d, e and Tables 12 and 13). The same tendency is true for Inv (Fig. 10 and Table 13). More data would be needed to show whether the apparent persistence of this effect for Gm up to the age of two is fortuitous or not.

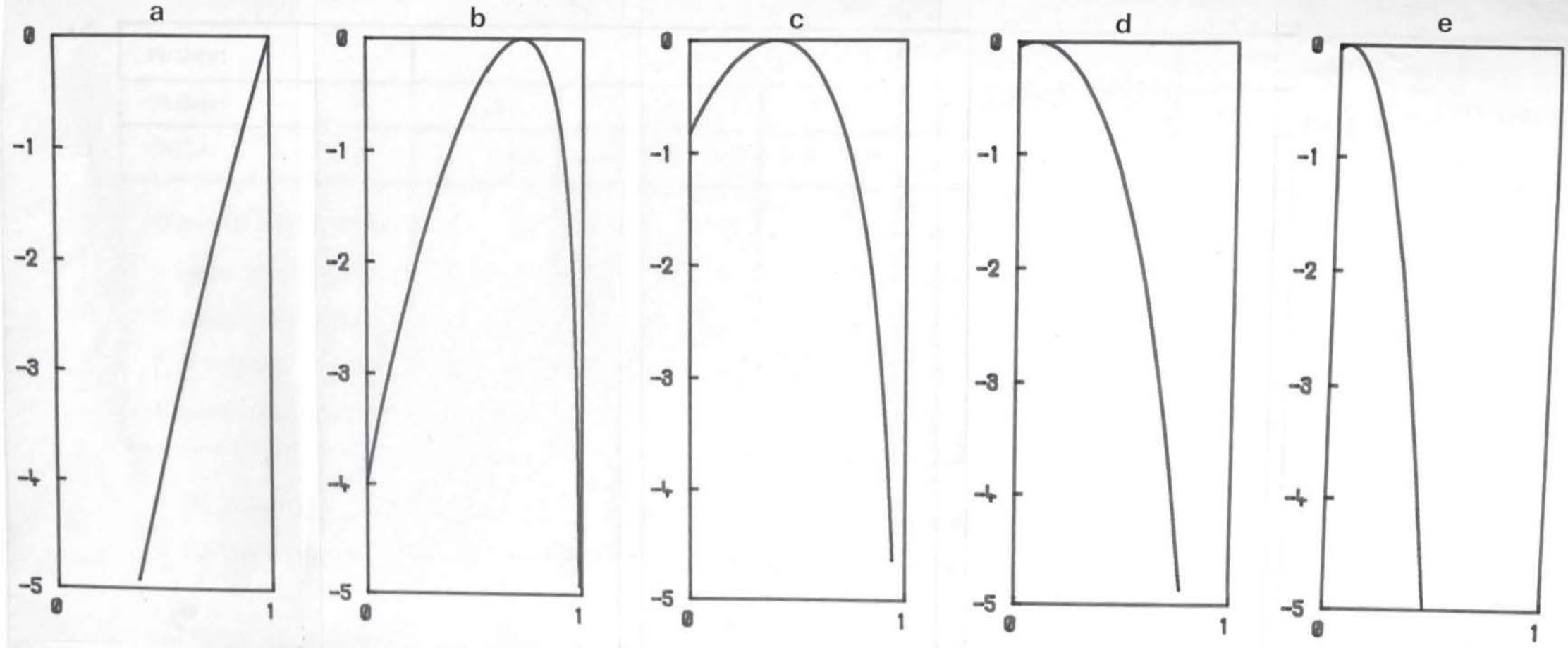


Fig. 9. Support plotted against probability b for Gm data on propositi of various age groups:
a) from birth to one month incl.
b) 5 weeks to 6 months
c) 7 months to 2 years
d) 3 to 5 years
e) 6 years and older.

Mother:	1,5							1,1				5,5				Total
Father:	1,1		1,5			5,5		1,1	1,5		5,5	1,1	1,5		5,5	
Child:	1,1	1,5	1,1	1,5	5,5	1,5	5,5	1,1	1,1	1,5	1,5	1,5	1,5	5,5	5,5	
From birth to 1 month	0	0	0	6	0	7	0	0	1	0	1	0	1	0	4	20
5 weeks to 6 months	0	2	0	5	0	5	2	2	0	0	2	0	5	7	2	32
7 months to 2 years	0	1	1	3	1	3	1	0	2	1	1	2	4	1	10	31
3 to 5 years	1	1*	2	6	1	1	1	0	2	0	4	3	1	0	2	25
6 years and older	3	6+	3	10	4	12	12	0	2	2	1	3	9	8	26	101

* The other Gm tests show that this individual cannot carry both maternal alleles.

\+ The other Gm tests show that two of these six individuals cannot carry both maternal alleles.

Table 12. Frequencies of Gm phenotypes for various age groups of Manitoban propositi in all parental mating classes.

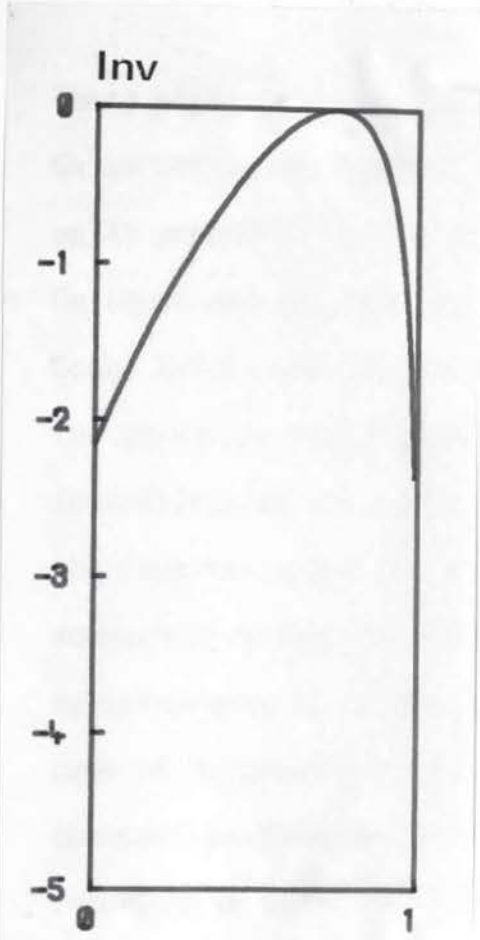


Fig. 10. Support plotted against probability $\underline{b}$ for Inv data on propositi aged from one month to one year inclusive.

Gm	From birth to 1 month	1.00 (0.71, 1.00)
	5 weeks to 6 months	0.71 (0.23, 0.96)
	7 months to 2 years	0.40 (0.00, 0.84)
	3 to 5 years	0.08 (0.00, 0.58)
	6 years and older	0.04 (0.00, 0.32)
Inv	1 month to 1 year	0.74 (0.03, 0.99)
	2 years and older	0.12 (0.00, 0.56)

Table 13. Evaluates of $\underline{b}$ for Gm and Inv, and the two values of $\underline{b}$ between which the support is less than two units below the maximum, for various age groups of Manitoban propositi.

These might be of some interest since there is no evidence to say that Gm is not on the tip of the long arm of chromosome 21. Other locations on 21 are ruled out by the firm absence of close linkage between the Gm locus and the break point in 21-translocations (Hustinx, 1966; Cook, 1974), and the presence of two Gm alleles in a patient lacking the short arm and the proximal part of the long arm of a chromosome 21 (Dutrillaux et al., 1973). Further, the absence of close linkage between the centromere and the Pi locus (Cook, 1974) which is about 27 centimorgans away from Gm (Cook, 1974) relegates the possibility of Gm being on chromosome 21 to the very tip of the long arm. The 6/21 translocation case of Borgaonkar et al. (1973) carries both the normal and translocated chromosomes from her Gm(1+5+) mother, she is Gm(1-5+) and therefore a recombinant if Gm is on the translocated chromosome. This is also compatible with Gm being on the tip of the long arm of 21 since a crossover would have had to happen between the centromere and the distant locus.

Finally, a word should be said about the possible association of certain phenotypes and Down's syndrome.

First, it is worth noting that the striking similarity detected by Penrose (1957) between trisomics and their mothers at several blood groups does not seem to be substantiated on these more extensive data, which include his original data (Lang-Brown et al., 1953), since heterozygous mothers do not have excesses of heterozygous children with Down's syndrome at any of these loci.

Second, it is obvious that the present method will not detect the association of a homozygous phenotype with the syndrome since this would result in a lack of heterozygotes and only show that the locus is not on chromosome 21. A casual inspection of Table 9 shows that this is precisely the case for the Gc locus. A break-down of the sample is given in Table 14 which shows that if the lack of heterozygotes is due to technical errors, then the three centres err in the same direction. The effect is more likely to be due to an action of trisomy, probably resulting in a low production by the Gc alleles, although this does not seem to have been detected in the 141 cases of Down's syndrome with 54 observed heterozygotes (54.3 expected) reported by Rittner and Schwinger (1973a).

Mother:	1,2							1,1				2,2				
Father:	1,1		1,2			2,2		1,1	1,2		2,2	1,1	1,2		2,2	Total
Child:	1,1	1,2	1,1	1,2	2,2	1,2	2,2	1,1	1,1	1,2	1,2	1,2	1,2	2,2	2,2	
Oregon	11	4	0	4	2	0	2	7	5	2	2	1	1	0	1	42
Scotland	7	0	1	0	1	0	1	14	5	1	1	1	0	0	0	32
Manitoba	2	0	2	0	2	0	0	10	1	1	2	1	0	1	0	22
Total	20	4	3	4	5	0	3	31	11	4	5	3	1	1	1	96

Table 14. Gc segregation in Down's syndrome, as reported by three centres.

V SEX CHROMOSOME ABNORMALITIES

When loci are assigned to specific trisomic chromosomes, their distributions in trisomic individuals and their families give some information on the origin of the defect since these are partially determined by the relative proportions of non-disjunctions at meiosis I and II. Although such data do not seem to have been gathered in sufficient numbers to permit a thorough investigation of the origin of autosomal trisomies, they are available for families of people with abnormalities of the sex chromosomes. These include the data on the X-linked Xg blood group of 506 propositi with the 47,XXY Klinefelter's syndrome and their families, kindly provided to J.H. Edwards by Ruth Sanger of the MRC Blood Group Unit in London, and B. Chown of the Rh Laboratory in Winnipeg.

A. XXY KLINEFELTER'S SYNDROME

The data are summarized in Table 15.

Father	Mother	-	+
+	+	7	92
+	-	8	6
-	+	7	42*
-	-	6	0
?	+	3	44
?	-	4	3
+	?	3	3
-	?	2	2
?	?	39	235
Totals		79	427

*One mother was not tested but the propositus has an Xg(a+) sister.

Table 15. Xg information on 506 XXY Klinefelter males and some of their parents.

As in the previous chapter, we can define b and c as the respective proportions of XXY propositi who received alleles of an X-linked locus from both or only one of their maternal grandparents, and add a new category a for those who received one X-borne allele from each parent. (Normal males can only transmit the allele they received from their mother.) The expected proportions r_i of Xg(a-) propositi in all parental mating classes i have been published by Edwards (1971). They are given in Table 16.

Mating type		r = expected proportion of recessive individuals
Father	Mother	
+	+	$r_1 = cq/(1+q)$
+	-	$r_2 = 1-a$
-	+	$r_3 = q(a+c)/(1+q)$
-	-	$r_4 = 1.0$
?	+	$r_5 = q(aq+c)/(1+q)$
?	-	$r_6 = 1-ap$
+	?	$r_7 = cpq + q^2(1-a)$
-	?	$r_8 = (q + (a+c)p)/(1+p)$
?	?	$r_9 = q(q+cp)$

Table 16. Expected proportions of Xg(a-) children, given for all mating types in terms of the parameters a and c defining the parental origin of the children's sex chromosomes. p and q are the proportions of the dominant and recessive alleles respectively.

If we define R_i and D_i as the observed numbers of recessive and dominant propositi in each class, the likelihood function is

$$L = \text{constant} \times \prod_{i=1}^{9} (r_i)^{R_i} (1-r_i)^{D_i} \quad \ldots\ldots\ldots\ldots (2)$$

and can easily be iterated on a computer to find the evaluates. When this is done with the data of Table 15, the evaluates and their two-unit support limits are as follows:

a = 0.41 (0.34, 0.49)

b = 0.38 (0.30, 0.46)

c = 0.21 (0.18, 0.26)

Figure 11 shows the position of these evaluates on a Streng diagram, together with the curve limiting their two-unit support region. Points outside this region are unlikely to represent the true values of these

parameters. This means that about 40% of the XXY Klinefelter propositi receive an extra X chromosome from their fathers, and 60% from their mothers.

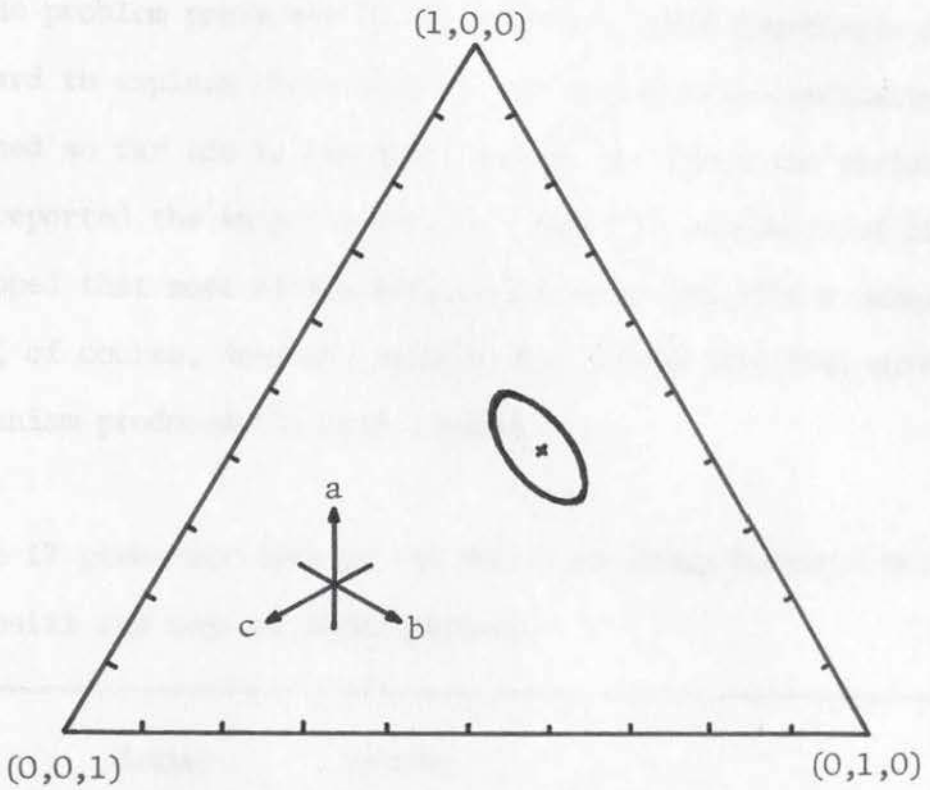


Fig. 11. Evaluates of the meiotic parameters a, b and c, and their two-unit support region for 47,XXY Klinefelter males.

Although non-disjunctions must happen at meiosis I in fathers, inferences about the type of maternal non-disjunction is not so simple because of the possibility of crossing-over between the two X chromosomes in the female. In fact, if we assume mitotic or somatic non-disjunctions to be negligibly rare, we can use equation (1) of chapter IV to describe the relationships between the recombination fraction, the proportions of non-disjunctions at division I and II of meiosis, and the maximum number of cross-overs between the locus and the X centromere in the female. Solving this equation for all unknown parameters is obviously impossible in the present state of knowledge and will only be possible when one knows more precisely where the Xg locus is on the X chromosome.

B. THE ORIGIN OF 46,XX MALES

The method for investigating the origin of 47,XXY males can be extended to the problem presented by 46,XX males. Many hypotheses have been put forward to explain their origin, but no definite conclusion has been reached so far (de la Chapelle, 1972). Although the various workers who reported the known cases have different standards of diagnosis, it is hoped that most of the accumulated data describe a unique syndrome. This, of course, does not exclude the possibility that more than one mechanism produces the same condition.

Table 17 gives the data of the MRC Blood Group Research Unit on 39 propositi and some of their parents.

Father	Mother	-	+
+	+	1	14
+	-	3*	1
-	+	0	4
?	+	1	6
?	-	1	0
?	?	1	7
Totals		7	32

*One father was not tested but the propositus has Xg(a+) sisters.

Table 17. Xg information on 39 46,XX men and some of their parents.

Parental phenotypes and mating proportions do not significantly differ from those in the normal population. A closer look at the data reveals

that the Xg^a distribution simply cannot follow either of the XY male and XX female distributions since impossible observations are recorded in both cases: if we denote the phenotypes in parentheses as (father, mother, propositus), we see that one (++-) and two (+--) did not inherit their fathers' genes as XX females do, while one (+-+) did, contrary to the expectation in XY males. Another (?--) has Xg(a+) sisters, and is presumably (+--). In addition, the total frequencies are significantly different from the expectations in XY males, even when the (+-+) is arbitrarily removed, and an XX female distribution - expected if a gene similar to the Sxr gene in mice were causing the syndrome - is almost ruled out by the virtual absence of recurrence in siblings and of consanguinity among the parents (de la Chapelle, 1972). This leaves at least two possible mechanisms remaining and they will be studied more closely.

The first is an XXY origin, in which case 46,XX men would arise from 47,XXY zygotes that subsequently lose their Y, at least in those tissues where chromosomal analysis is done. This is plausible since several XX/XXY mosaics are known to have a very low proportion of XXY cells. These cells could go unnoticed or even completely disappear in some individuals whose Xg^a distribution would evidently be the same as that in XXY men.

When equation (2) is used with the 46,XX men data to estimate the parameters a, b and c, maximum support is given to the following combination of parametric values

$$a = 0.20$$

$$b = 0.56$$

$$c = 0.25$$

but the two-unit support region would cover the almost entirety of the Streng diagram (Fig. 12) except for the very corners corresponding to the Xg distributions of normal males and females. Support for the

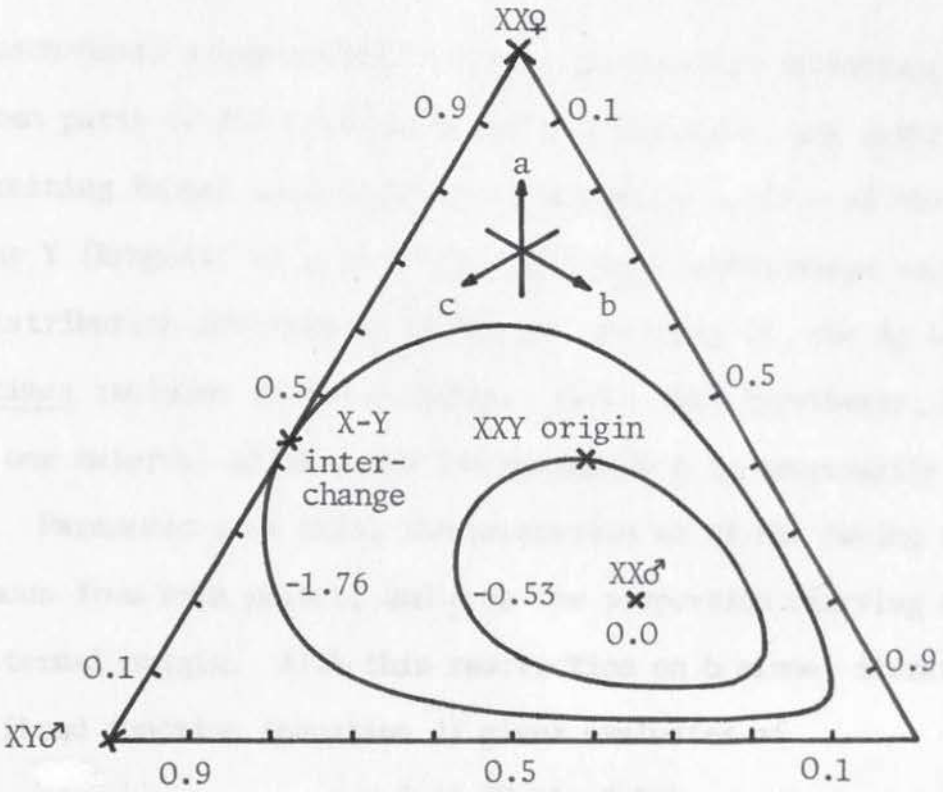


Fig. 12. Evaluates of the meiotic parameters a, b and c, for 46,XX men, and positions of the points corresponding to the Xg distributions in normal males and females, and those representing the two theories of the XXY origin and the X-Y interchange. The points on each of the two contours receive equal support.

hypothesis of the XXY origin (i.e. to the parametric values of $\underline{a}$ = 0.41, $\underline{b}$ = 0.38 and $\underline{c}$ = 0.21) is only 0.53 units less than that given to the evaluates. Therefore, the hypothesis cannot be rejected, but it clearly could not be with so few data, whether it was true of false.

However, these few data still carry some information, although small, and the method of support can be used to find out if the XXY theory is better supported than the other mechanism proposed by Ferguson-Smith (1966). The answer will not be conclusive but will accurately describe the actual

state of knowledge about the two hypotheses. It will also have the advantage of being additive with any future results reported in terms of support.

Ferguson-Smith's hypothesis is that a chromosomal interchange takes place between parts of the paternal X and Y chromosomes, and involves the male determining factor located on the centromeric portion of the short arm of the Y (Krmpotic et al., 1972). The small interchange would give the Xg distribution observed in XX men if, and only if, the Xg locus were sometimes included in the exchange. Under this hypothesis, XX men receive only one maternal allele, and the parameter b is necessarily equal to zero. Parameter a is still the proportion of XX men having inherited one Xg locus from each parent, and c is the proportion carrying a single locus of maternal origin. With this restriction on b alone, iteration of the likelihood function (equation 2) gives evaluates of

$$a = 0.43 \ (0.05, 0.79)$$

$$c = 0.57 \ (0.21, 0.95)$$

with their two-unit support limits. The support for this combination of parametric values is 1.23 units less than that for the XXY hypothesis. This is a minimum difference since the XXY hypothesis was defined by the evaluates obtained with the XXY data, while the interchange hypothesis was defined by using the XX data.

The hypothesis of the XXY origin thus receives slightly more support than the X-Y interchange hypothesis although the apparent absence of maternal age effect argues against the XXY origin. A definite conclusion will only be possible when further data become available. A mixture of both mechanisms might then be shown to be the solution.

VI CONCLUSION

MARK III has already proved to be a useful tool for routine linkage analyses of diploids in both Winnipeg and Leiden. It is hoped that the likelihood approach used throughout this thesis will also simplify some of the other problems presented by centromeric linkage and chromosomal variants. The analysis of pooled data on both trisomies 13 and 18 is already under way in this department and should soon produce clear results. The controversy on the origin of most 46,XX men will also be easily solved with the proposed method when more data become available.

BIBLIOGRAPHY

Alter, A.A., Lee, S.L., Pourfar, M. & Dobkin, G. (1962). Leukocyte alkaline phosphatase in mongolism; a possible chromosome marker. J. Clin. Invest., 41: 1341.

Baikie, A.G., Loder, P.B., DeGruchy, G.C. & Pitt, D.B. (1965). Phosphohexokinase activity of erythrocytes in mongolism: Another possible marker for chromosome 21. Lancet, i: 412-414.

Ball, J.R.B., Brackenridge, C.J., McKay, H. & Pitt, D.B. (1972). Hp distributions in Down's syndrome. Clin. Genet., 3(5):334-340.

Barnard, G.A. (1949). Statistical inference. J. Roy. Stat. Soc., B11: 115-139.

Bateman, A.J. (1960). Blood-group distribution to be expected in persons trisomic for the ABO gene. Lancet, i: 1293-1294.

Berman, J., Hulten, M. & Lindsten, J. (1967). Blood-serotonin in Down's syndrome. Lancet, i: 730.

Bjarnason, O., Bjarnason, V., Edwards, J.H., Fridriksson,S., Magnusson, M., Mourant, A.E. & Tills, D. (1973). The blood groups of Icelanders. Ann. hum. genet., 36: 425-458.

Boone, C.M., Chen, T.R. & Ruddle, F.H. (1972). Assignment of three human genes to chromosomes (LDH-A to 11, TK to 17, and IDH to 20) and evidence for translocation between human and mouse chromosomes in somatic cell hybrids. Proc. Nat. Acad. Sc. U.S., 69: 510-514.

Borgaonkar, D.S., Bias, W.B., Chase, G.A., Sadasivan, G., Herr, H.M., Golomb, H.M., Bahr, G.F. & Kunkel, L.M. (1973). Identification of a C6/G21 translocation chromosome by the Q-M and Giemsa banding techniques in a patient with Down's syndrome, with a possible assignment of Gm locus. Clin. Genet., 4: 53-57.

Brandt, N.J. (1962). Genes on the Mongol chromosome? Lancet, 2: 837.

Chautard-Freire-Maia, E.A. (1974). Linkage relationships between 22 autosomal markers. Ann. hum. genet., 38(2): 191-198.

Chown, B. & Lewis, M. (1963). ABO blood groups on mongols. Can. Med. Ass. J., 89: 906.

Chown, B., Lewis, M. & Kaita, H. (1962). Atypical Duffy inheritance in three Caucasian families: a possible relationship to mongolism. Am. J. hum. Genet., 14: 301-308.

Chown, B., Lewis, M. & Kaita, H. (1965). Atypical Fy^b in some families with a mongol child; a revocation. Am. J. hum. Genet., 17: 188.

Cook, P.J.L. (1974). The genetics of α_1-antitrypsin in man., Ph.D. thesis, University of London.

Cox, R.P. (1965). Regulation of alkaline phosphatase in skin fibroblast cultures from patients with mongolism. Exptl. Cell. Res., 37: 690-692.

de Grouchy, J. (1965). Chromosome 18: A topologic approach. J. Pediat., 66: 414-431.

de Grouchy, J., Lamy, M., Thieffry, S., Arthuis, M. & Salmon, C. (1963). Dysmorphie complexe avec oligophrénie: délétion des bras courts d'un chromosome 17-18. C.R. Acad. Sci., 256: 1028.

de la Chapelle, A. (1972). Analytic review: nature and origin of males with XX sex chromosomes. Am. J. hum. Genet., 24: 71-105.

Donahue, R.P., Bias, W.B., Renwick, J.H. & McKusick, V.A. (1968). Probable assignment of the Duffy blood group locus to chromosome 1 in man. Proc. Nat. Acad. Sc. U.S., 61(3): 949-955.

Dutrillaux, B., Jonasson, J., Kerstin, L., Lejeune, J., Lindsten, J., Petersen, G.B. & Saldaña-Garcia, P. (1973). An unbalanced 4q/21q translocation identified by the R but not the G and Q chromosome banding techniques. Ann. Génét., 16(1): 11-16.

Edwards, A.W.F. (1972). Likelihood. Cambridge University Press.

Edwards, J.H. (1971). On the distribution of phenotypes in XXY males and their parents. J. med. Genet., 8(4): 434-438.

Edwards, J.H. (1972). A marker algebra. Clin. Genet., 3: 371-380.

Edwards, J.H. (1972b). Linkage studies. In S.W. Wright, B.F. Crandall & L. Boyer, eds., Perspectives in cytogenetics, C.C. Thomas, Springfield, Illinois.

Edwards, J.H., Allen, F.H., Glenn, K.P., Lamm, L.U. & Robson, E.B. (1972). The linkage relationships of HL-A. In Histocompatibility Testing 1972, Munksgaard, Copenhagen, Denmark, 745-751.

Edwards, J.H., Fraccaro, M., Davies, P. & Young, R.B. (1962). Structural heterozygosis in man: analysis of two families. Ann. hum. Genet., 26: 163-168.

Elsevier, S.M., Kucherlapati, R.S., Nichols, E.A., Creagan, R.P., Giles, R.E., Ruddle, F.H., Willecke, K. & McDougall, J.K. (1974). Assignment of the gene for galactokinase to human chromosome 17 and its regional localisation to band q21-22. Nature, 251: 633-666.

Evans, D.A.P., Donohue, W.T., Bannerman, R.M., Mohn, J.F. & Lambert, R.M. (1966). Blood-group gene localization through a study of mongolism. Ann. hum. Genet., 30: 49-67.

Evans, D.A.P., Wren, P.J.J., Donohoe, W.T.A., Bullen, M.F., Lewis, M., Kaita, H., Chown, B. & Uchida, I. (1968). Further observations on Kell blood groups in families ascertained via a mongol propositus. J. med. Genet., 5(4): 310-313.

Falk, C.T. & Edwards, J.H. (1970). A computer approach to the analysis of genetic data for the detection of linkage. Genetics, 64: 518.

Ferguson-Smith, M.A. (1966). X-Y chromosomal interchange in the aetiology of true hermaphroditism and of XX Klinefelter's syndrome. Lancet, ii: 475-476.

Ferguson-Smith, M.A., Newman, B.F., Ellis, P.M., Thomson, D.M.G. & Riley, I.D. (1973). Assignment by deletion of human red cell acid phosphatase gene locus to the short arm of chromosome 2. Nature New Biol., 243: 271.

Goodman, H.O. (1965). Phenotypic shifts in trisomy. Am. J. hum. Genet., 17: 111-124.

Goodman, H.O. & Thomas, J.J. (1966). ABO frequencies in mongolism. Ann. hum. Genet., 30: 43-48.

Goodman, H.O. & Thomas, J.J. (1968). Kell types in Down's syndrome. Ann. hum. Genet., 31: 369-372.

Haldane, J.B.S. & Smith, C.A.B. (1947). A new estimate of the linkage between the genes for colour-blindness and haemophilia in man. Ann. Eug., 14: 10-31.

Hsia, D.Y.-Y., Shih, L.-Y., Easterberg, S., Farquhar, J., Kim, C.B., Yeh, S. & Young, A. (1969). The distribution of genetic polymorphisms among patients with Down's syndrome, phenylketonuria and cystic fibrosis of the pancreas. Am. J. hum. Genet., 21: 285-289.

Hustinx, T.W.J. (1966). Cytogenetisch onderzoek bij enige families. Drukkerij-Uitgeverij Brakkenstein.

Hutton, A.C. & Smith, G.F. (1964). Haptoglobins and transferrins in patients with Down's syndrome. Ann. hum. Genet., 27: 413-415.

Kaplan, S., Li, C.C., Wald, N. & Borges, W. (1964). ABO frequencies in mongols. Ann. hum. Genet., 27: 405-412.

Krmpotic, E., Szego, K., Modestas, R. & Molabola, G.B. (1972). Localization of male determining factor on short arm of Y chromosome. Clin. Genet., 3: 381-387.

Lang-Brown, H., Lawler, S.D. & Penrose, L.S. (1953). The blood typing of cases of mongolism, their parents and sibs. Ann. Eug., 17: 307-336.

Lawler, S.D. (1962). Genes on the mongol chromosome? Lancet, ii: 837.

Lawler, S.D. (1964). Localisation of autosomal genes in man. Human Biology, 36: 146-156.

Lejeune, J., Gautier, M. & Turpin, R. (1959a). Les chromosomes humains en culture de tissus. C.R. Acad. Sc., 248: 602-603.

Lejeune, J., Gautier, M. & Turpin, R. (1959b). Etude des chromosomes somatiques de neuf enfants mongoliens. C.R. Acad. Sc., 248: 1721-1722.

Lejeune, J., Turpin, R. & Gautier, M. (1959c). Le mongolisme, premier exemple d'aberration autosomique humaine. Ann. Génét. 1(2): 41-49.

Lockie, S.J. & Edwards, J.H. (1971). The interpretation of genetic markers in mongolism. Excerpta Medica, Internat. Congress Series, 233: 114.

Maynard-Smith, S., Penrose, L.S. & Smith, C.A.B. (1961). Mathematical tables for research workers in human genetics. London, Churchill.

McDaniel, R.G. & Ramage, R.T. (1970). Genetics of a primary trisomic series in barley: identification by protein electrophoresis. Can. J. Genet. Cytol., 12: 490-495.

McDougall, J.K., Kucherlapati, R.S. & Ruddle, F.H. (1973). Localization and induction of the human thymidine kinase gene by adenovirus 12. Nature New Biol., 245: 172-175.

Migeon, B.R. & Miller, C.S. (1968). Human-mouse somatic cell hybrids with single human chromosome (group E): link with thymidine kinase activity. Science, 162: 1005-1006.

Mikkelsen, M. (1970). A Danish survey of patients with Down's syndrome born to young mothers. Ann. N.Y. Acad. Sc., 171: 370-378.

Miller, O.J., Allerdice, P.W., Miller, D.A., Breg, W.R. & Migeon, B.R. (1971). Human thymidine kinase gene locus: assignment to chromosome 17 in a hybrid of man and mouse cells. Science, 173: 244-245.

Morton, N.E. (1955). Sequential tests for the detection of linkage. Am. J. hum. Genet., 7: 277-318.

Morton, N.E. (1956). The detection and estimation of linkage between the genes for elliptocytosis and the Rh blood type. Am. J. hum. Genet., 8: 80-96.

Morton, N.E. (1957). Further scoring types in sequential linkage tests, with a critical review of autosomal and partial sex linkage in man. Am. J. hum. Genet., 9(1): 55-75.

Nadler, H.L., Inouye, T. & Hsia, D. Y.-Y. (1967). Enzymes in cultivated human fibroblasts derived from patients with autosomal trisomy syndromes. Am. J. hum. Genet., 19: 94-99.

New Haven Conference (1973). First international workshop on human gene mapping. Birth defects: original article series, X: 3, 1974. The National Foundation, N.Y.

O'Brien, D. & Groshek, A. (1962). The abnormality of tryptophane metabolism in children with mongolism. Arch. Dis. Childhood, 37: 17-20.

Pantelakis, S.N., Karaklis, A.G., Alexiou,D., Vardas, E. & Valaes, T. (1970). Red cell enzymes in trisomy-21. Am. J. hum. Genet., 22: 184-193.

Paris Conference (1971). Standardization in human genetics. Birth defects: original article series, VIII: 7, 1972. The National Foundation., N.Y.

Patau, K., Therman, E. & Smith, D.W. (1963). Partial-trisomy syndromes and the mapping of trisomy loci. Proc. 2nd Int. Cong. hum. Genet. (Rome, Sept. 1961), 2: 1203-1206.

Penrose, L.S. (1957). Similarity of blood antigens in mother and mongol child. J. ment. Defic. Res., 1: 107-110.

Penrose, L.S. (1963). Measurements of likeness in relatives of trisomics. Ann. hum. Genet., 27: 183-187.

Race, R.R. & Sanger, R. (1968). Blood groups in man. Blackwell Scientific Publications, Oxford & Edinburgh, 5th ed.

Race, R.R. & Sanger, R. (1974). Blood groups in man., 6th ed. In press.

Rittner, C. & Schwinger, E. (1973a). Studies in Down's syndrome. II. Association studies with blood, serum and enzyme groups, and with the Au/SH antigen. Clin. Genet., 4: 398-406.

Rittner, C. & Schwinger, E. (1973b). Studies in Down's syndrome. IV. Investigations on the 'Xh' factor. Clin. Genet., 4: 407-409.

Rosner, F., Ong, B.H., Paine, R.S. & Mahanand, D. (1965). Biochemical differentiation of trisomic Down's syndrome (mongolism) from that due to translocation. New Engl. J. Med., 273: 1356-1361.

Ruddle, F.H. & Chen, T.R. (1972). In Perspectives in Cytogenetics, Wright, S.W., Crandall, B.F. & Boyer, L., editors, C.C. Thomas, Illinois.

Rundle, A.T. (1973). Theoretical implications of the application of polymorphic systems to gene location on a trisomic chromosome, and its application to haptoglobin phenotype frequencies in Down's syndrome. Clin. Genet., 4: 520-535.

Rundle, A.T. & Sudell, B. (1973). Red cell polymorphisms in Down's syndrome. Gene frequencies and phenotype associations. Clin. Genet., 4: 536-547.

Saldanha, P.H. & Beçak, W. (1963). Genotypic distribution among human autosomal aneuploids: a theoretical consideration. Acta genet. stat. med., 13: 67-76.

Sasaki, M. & Hara, Y. (1973). Paternal origin of the extra chromosome in Down's syndrome. Lancet, ii: 1257-1258.

Shaw, M.W. & Gershowitz, H. (1962). A search for autosomal linkage in a trisomic population: blood group frequencies in mongols. Am. J. hum. Genet., 14: 317-334.

Shaw, M.W. & Gershowitz, H. (1963). Blood group frequencies in mongols. Am. J. hum. Genet., 15: 495-496.

Smith, C.A.B. (1953). The detection of linkage in human genetics. J. Roy. Stat. Soc., B., 15: 153-192.

Steinberg, A.G. & Morton, N.E. (1956). Sequential test for linkage between cystic fibrosis of the pancreas and the MNS locus. Am. J. hum. Genet., 8: 177-189.

Swanson, C.P., Merz, T. & Young, W.J. (1967). Cytogenetics. Prentice-Hall, New Jersey.

Thieffry, S., Arthuis, M., de Grouchy, J., Lamy, M. & Salmon, G. (1963). Délétion des bras courts d'un chromosome 17-18: dysmorphies complexes avec oligophrénie. Arch. Franç. Pédiat., 20: 740.

Trubowitz, S., Kirman, D. & Masek, B. (1962). The leucocyte alkaline phosphatase in mongolism. Lancet, ii: 486-487.

Uchida, I.A. (1970). Epidemiology of mongolism: the Manitoba study. Ann. N.Y. Acad. Sc., 171: 361-369.

Uchida, I.A. (1973). Paternal origin of the extra chromosome in Down's syndrome. Lancet, ii: 1258.

Verma, C. (1974). A study of variant haemoglobins among immigrants with special reference to the position of the beta locus. Ph.D. thesis, University of Birmingham.

Wald, A. (1947). Sequential analysis, New York, Wiley.

Weiss, M.C. & Green, H. (1967). Human-mouse hybrid cell lines containing partial complements of human chromosomes and functioning human genes. Proc. Nat. Acad. Sc. U.S., 58: 1104-1111.

Wright, S.W., Day, R.W., Muller, H. & Weinhouse, R. (1967). The frequency of trisomy and translocation in Down's syndrome. J. Pediat., 70: 420-424.

Yokoyama, M. Immunogenetic studies on mongolism. Hawaii med. J., 26: 121-124.