

Linkage Newsletter

Vol. 5 No. 2 May 1991

Published by Jurg Ott, Columbia University, New York¹.

Editorial Assistant: Katherine Montague

Tel. 212/960-2507 Fax: +1-212-568-2750 Bitnet: OTT@NYSPI

Postal address: Columbia University, Box 58
722 West 168th Street, New York, NY 10032

1. EDITORIAL

On January 1 of this year, I succeeded Prof. Lars Beckman as the editor of *Human Heredity*. This journal was established in 1950 as *Acta Genetica et Statistica Medica* and has a good tradition of publishing articles on methods and applications in various areas of human genetics. I would like to extend an invitation to the readers of this newsletter to submit manuscripts to *Human Heredity*, particularly in the areas of linkage analysis and gene mapping. A new category of articles will be created shortly ("Methodological Issues") in which new, particularly difficult, and/or important aspects of research methods may be discussed. A contribution to this section need not represent an entirely new investigation but it must be of general interest and of a high scientific standard.

2. LINKAGE COURSES

An advanced linkage course will be held in New York from Monday through Friday, October 14-18, 1991, the week following the International Congress of Human Genetics. Tuition for the 5-day course is \$100 (supported by a grant from the National Center for Human Genome Research). The maximum number of participants is 25. The course will take place in the microcomputer classroom of the Health Sciences Library of Columbia University.

Topics to be covered include: Introduction to the LINKAGE and MENDEL computer programs; handling of inbreeding loops, age-dependent penetrance, and sex-specific recombination fractions; problems of interference in multipoint mapping; introduction to models of disease heterogeneity; models for complex diseases; genetic heterogeneity; calculation of genetic risks, also under allelic association and allelic heterogeneity (as in CF); linkage analysis with pseudoautosomal loci.

¹ Support through grant HG00008 from the National Center for Human Genome Research is gratefully acknowledged.

Participants must be familiar with IBM PCs or compatible microcomputers. Extensive experience with a linkage program and/or an excellent background in statistical genetics and linkage analysis are additional criteria for admission.

The course will be advertised in scientific journals. To obtain additional material and an application form, please write to the address above. The application deadline is August 15, 1991.

An introductory course will be given from Wednesday through Saturday, August 28-31, 1991, in Cardiff, Wales (in the week following the Human Gene Mapping Workshop 11 in London). The course will be taught by myself, Lodewijk Sandkuijl, and Iain Fenton. The topics to be covered include: Introduction to population genetics, introduction to linkage analysis, practical aspects of data collection, two-point linkage, multipoint linkage, risk calculations, and linkage analysis for diseases with a complex mode of inheritance.

The total course fee is £650.- including full room and board accommodation. The number of participants is limited to 20. The computer classroom will be equipped with 20 PC-compatible computers with 80386 microprocessors. Requests for additional information and applications should be directed to:

Mrs. G. Gulliford, secretary to Prof. P.S. Harper
Institute of Medical Genetics, Heath Hospital
Cardiff, CF4 4XN, Wales
Fax +44-222-747603

or to

Dr. L.A. Sandkuijl, Voorstraat 27a
2611 JK Delft, The Netherlands
Fax +31-15-123638

3. SOFTWARE NOTES

3.1 Another bug in the SLINK program

The following contribution has been submitted by Dr. Weeks for inclusion in the newsletter:

Bug in SLINK

A bug in SLINK.PAS has recently been found. This bug will cause problems any time you try to simulate with only markers (i.e., no trait locus) and the markers aren't in the order 1, 2, 3,....

The fix to this bug is simple: Just change the line in SLINK.PAS containing
j:=order[1];

to

j:=1;

as indicated below:

Original code:

```
FOR i:=1 TO 35 DO write(output, '-');
writeln(output);
{Setup unlinked trait locus}
IF trait<>0 THEN
j:=order[trait]
ELSE
j:=order[1];
FOR i:=1 TO nlocus DO
IF i<>trait THEN
IF order[i]<j THEN
order2[i]:=order[i]+1
ELSE
order2[i]:=order[i];
```

Corrected code:

```
FOR i:=1 TO 35 DO write(output, '-');
writeln(output);
{Setup unlinked trait locus}
IF trait<>0 THEN
j:=order[trait]
ELSE
j:=1;
FOR i:=1 TO nlocus DO
IF i<>trait THEN
IF order[i]<j THEN
order2[i]:=order[i]+1
ELSE
order2[i]:=order[i];
```

If you have any questions about this fix, please contact:

Daniel E. Weeks

University of Pittsburgh

Department of Human Genetics

130 DeSoto Street, A300 Crabtree Hall

Pittsburgh, PA 15261

Tel. (412) 624-3066

FAX: (412) 624-3020

WEEKS@PITTVMS.BITNET or WEEKS@VMS.CIS.PITT.EDU

3.2 FTREE pedigree drawing program

Dr. Rodney C.P. Go at the University of Alabama submitted a copy of his FTREE-Family Tree Drawing Program. It is written in Fortran and comes in versions for Vax and IBM PC-compatible computers. The manual is supplied in a disk file, with one version in ASCII format and one in WordPerfect format.

The FTREE program may be obtained from:

Rodney C.P. Go, Ph.D.
 UAB - University Station
 Birmingham, AL 35294-0008. Tel. 205/934-6107

3.3 Usage notes to LINKAGE version 5.10.

In previously mailed versions of the PREPLINK program (Turbo Pascal only), the M compiler switch (first program line, third number on that line) was used to control the maximum amount of memory available to the program. Such a limitation was necessary for proper functioning of one of the program features, that is, to see a directory listing of files. However, the limitation imposed by the M switch may not leave enough memory if you want to specify a large number of loci. Therefore, the current Turbo Pascal version of PREPLINK no longer contains the M switch and also does not allow one to obtain a directory listing before specifying a file name for input.

When you try running MLINK with the *dostream* program constant set to false, an error occurs. The error may be fixed by adding *and dostream* to a line towards the end of the *iterpeds* procedure, 11 lines before the start of the *initmlink* procedure (in the file MLK3.PAS of the DOS version). The corrected line reads:

```
if score and (not risk) and dostream
  then writeln(stream,tlike-scorevalue);
```

The UNKNOWN program (DOS and other versions) does not properly read the datafile when more than one quantitative factor is specified. The error occurs in the *getquan* procedure; the corrected code reads as follows:

```
procedure getquan(VAR locus:locuspoint);
VAR
  i:INTEGER;
begin {getquan}
  WITH locus^ DO
  begin
    READLN(datafile,ntrait);
    IF ntrait>maxtrait THEN inputerror(31,system,ntrait);
    IF ntrait<=0 THEN inputerror(32,system,nclass);
```

```

    FOR i:=1 TO ntrait+2 DO READLN(datafile);
  END;
END; {getquan}

```

3.4 Benchmark tests for LINKAGE programs

The benchmark used previously to compare running times of linkage programs on different machines (see Linkage Newsletter vol. 3, December 1989) is no longer suitable for today's microcomputers as it runs too quickly and, thus, partially measures speed of video output. Our new benchmark consists of a 12-member family with an inbreeding loop in which a recessive disease and three markers are segregating. It was run on various machines in the times shown below. I would like to thank Drs. Catherine Falk, New York, and John Rice, St. Louis, for running the benchmark on their Sun machines.

Generally, version 5.10 of the LINKAGE programs runs approximately 10% faster than version 5.04. The benchmark results listed below were obtained with version 5.10 of the LINKAGE programs (Turbo Pascal for DOS machines). All machines were equipped with numeric coprocessors as indicated. For the DOS machines the clock speed of the microprocessor is also given. The run time is the time in seconds taken by the MLINK program of the LINKAGE package to calculate two likelihoods for the 12-member family described above (elapsed time except where noted). The UNKNOWN program was executed and a speedfile produced prior to the test run. Program constants were the same in each of the runs listed below. The benchmark data set is available on disk 5c (see appendix).

Machine	Run time
Toshiba 3100-e (80286-12/80287)	2099
BUS Laptop 386SX (80386SX-16/80387)	765
IBM PS/2 model 70 (80386-16/80387)	710
Toshiba 5200 (80386-20/80387-20)	449
Dell System 310 (80386-20/80387-20)	446
BUS 386 (80386-25/80387-25)	426
Vaxstation 3100 model 30 (Vax Pascal, CPU time)	237
Vaxstation 3100 model 38 (Vax Pascal, CPU time)	167
Everex Step (80486-33)	98
BUS 486 (80486-33)	98
Sun SLC (rated at 12.5 MIPS)	70
Sun 4/370 (rated at 16 MIPS)	49
Sun Sparcstation 1+ (rated at approx. 16 MIPS)	45

In the comparisons given above, the purchase price of the machines should also be taken into account. The faster Sun machines listed run approximately twice as fast as the fastest DOS machines but their cost (with educational discount) is less than twice that of the DOS machines.

3.5 LINKAGE programs for MS-OS/2

The LINKAGE programs are now also available for running under OS/2 (IBM-compatible microcomputers; see list of programs attached). Details of the OS/2 implementation will be given in the next issue of this newsletter, and the benchmark problem will be run under this version when we receive version 1.3 of OS/2.

One important aspect of the OS/2 implementation is that the Prospero compiler in principle allows for arrays larger than 64KB. Records containing such arrays, however, cannot exceed 64KB. Whereas this restriction is not too serious, it also turns out that memory for large arrays cannot be allocated by *new(p)*, where *p* is a pointer pointing to an array larger than 64KB. This limits the total number of loci that can be analyzed jointly, but the OS/2 version of the LINKAGE programs can still accommodate problems larger than can be run under MS-DOS.

As mentioned in the previous issue of the newsletter, to run under OS/2, the LINKAGE programs were adapted to Prospero Pascal. Unfortunately, the U.S. address and telephone number of the Prospero Software company are no longer correct. Readers interested in purchasing Prospero Pascal should contact Prospero Software, 190 Castelnau, London SW13 9DH, England; fax +44-81-748-9344, tel. +44-81-741-8531. I have no connection with that company whatsoever and am providing this address in response to readers who tried in vain to contact the company at their U.S. office.

4. APPENDIX: List of programs available

These programs are designed for IBM type microcomputers, unless indicated otherwise. Program versions for the Macintosh SE/30 are being developed by Dr. Daniel Weeks, who is now at Pittsburgh University. Below, sets of files are arranged as numbered disks. Most of these 'disks' hold up to 360 KB characters, but some (identified with 'DD') contain up to 720 KB.

** For ordering instructions, please write to us at the address above **

Abbreviations:

TPS= Turbo Pascal (TP) source code (compiler needed).

TPC= Turbo Pascal, compiled for IBM PC

FSC= Fortran (Microsoft v.4.01), source and compiled.

EXE= executable code

Item Contents

LINKAGE programs version 5.10, disks 4a-4c (Prospero Pascal, DOS or OS/2) and disks 5a-5e (Turbo Pascal version 5, DOS). The programs are in archived form and will have to be uncrunched using a program supplied on disks 4a and 5a. Printed documentation (version 5.10, May 1991) will be provided. Our benchmark pedigree is on disk 5c.

LINKAGE programs in Prospero Pascal (general and CEPH pedigrees; DOS or OS/2, compiled for OS/2 only): Users with a coprocessor installed in their machines (eg. all 80486 machines) order disks 4a and 4b, those without a coprocessor order disks 4a and

4c.

- 4aSource code, utility programs, and documentation
- 4bExecutable code using coprocessor
- 4cExecutable code not making use of coprocessor

LINKAGE programs in Turbo Pascal (general and CEPH pedigrees; DOS only):

For general pedigrees order 5a-d, for 3-generation pedigrees order 5a-c and 5e.

- 5aLCP and other management programs
- 5bVarious utility and batch programs, documentation files
- 5cSource code; benchmark data (in unarchived form)
- 5d Executable code, general pedigrees
- 5eExecutable code, 3-generation pedigrees

OTHER PROGRAMS (Turbo Pascal version 5 except where noted)

- 6Source code to disk 8 (TPS).
- 7NOCOM program for analysis of mixture of distributions. Includes the COMPMIX and HIST program. FSC
- 8Linkage Utility programs (see list below). TPC
- 9PC-LIPED (two-point linkage analysis, up to 5 alleles per locus), version Oct. 1988. Includes SEXLODS program for approx. separation of male and female recombination fractions. FSC
- 9aLIPED, same as disk 9 except that up to 6 alleles per locus are allowed and program requires more memory to run.
- 10abcPC-WRITE version 3.02 text editor for data entry (Quicksoft, Inc.). [3 disks]
- 15HOMOG programs and MTEST program to carry out homogeneity tests TPS
- 16HOMOG programs and MTEST program to carry out homogeneity tests TPC
- 17Kermit V2.30 program for electronic communication.
- 18LIPEDMAX program version Nov. 1987 for iterative estimation of age of onset parameters (lognormal distribution of age at onset). FSC
- 20aSLINK simulation program, DOS and OS/2 version, Prospero Pascal source code and documentation file [1 DD disk]
- 20bSLINK for DOS and OS/2, compiled for machines with coprocessor [1 DD disk]
- 20cSLINK for DOS and OS/2, compiled (ISIM not included), not requiring coprocessor [1 DD disk]
- 21 SLINK simulation program, VAX VMS version [1 DD disk]
- 22a 2-locus LINKAGE programs (TMLINK, TLINKM, TILINK) for DOS and OS/2 [1 DD disk]
- 22b 2-locus LINKAGE programs for VAX VMS [1 DD disk]

 We keep a list of people who ordered programs from us and/or who have taken our linkage courses. These individuals regularly receive the LINKAGE NEWSLETTER which is so far being mailed free of charge a few times a year.

LITERATURE (for information purposes only)

J. Ott: "Analysis of Human Genetic Linkage", Johns Hopkins University Press, Baltimore and London, 1985 (\$35). A list of corrections is available from me on request. Japanese translation by Soft Science, Tokyo, 1987. New completely revised edition will be available in September, 1991.

E.A. Thompson: "Pedigree Analysis in Human Genetics", Johns Hopkins University Press, Baltimore and London, 1986 (\$35).

K.E. Davies (editor): "Human Genetic Diseases - A Practical Approach". IRL Press, Oxford England and Washington, D.C., 1986 (\$25, softbound; \$40, hardbound).

PROGRAMS CONTAINED On DISKS 6/8 (version no. in parentheses)

2BY2 (1.0) carries out Fisher's exact test in 2x2 tables ($n < 8000$).

ASSOCIATE (2.3), for two loci with codominant alleles, partitions the chi-square for phenotypic association into two components, (1) due to allelic association, (2) other phenotypic association.

BINOM (1.63) calculates binomial probabilities ($n < 8000$).

CELLIP (2.2) calculates points on a confidence ellipsis for two jointly estimated variables.

CHIPROB (2.2) computes the upper tail probability of the chi-square distribution.

CONTING (2.4) calculates chi-square for contingency table data.

EQUIV (2.6) calculates equivalent fully informative observations.

HIST (2.3) produces a histogram.

LSURF/LSMAX (3.3/1.3) calculate the lod score surface over the x_1, x_2 -plane in 3-point linkage analysis (all 3 orders), where x_1 and x_2 are the map distances from locus 1 to 2 and from locus 2 to 3, respectively. Input is offspring counts from phase known data.

MAPFUN (2.31) converts recombination fractions into map distances (6 mapping functions) and vice versa.

NORINV (1.31) accurately computes the normal deviate from a given tail probability.

NORPROB (3.2) accurately computes the tail probability associated with a normal deviate, x .

PERMUTE (2.3) produces a list of all $n!/2$ orders of n gene loci.

PIC (1.3) computes for given alleles at one locus the PIC value and heterozygosity.

RERI (2.2) calculates and combines relative risks from a set of 2×2 tables and carries out homogeneity tests among the tables.

VARCO3 (2.41) approximates mean and variance of an MLE of a variable x from three values of x and their log likelihoods or lod scores. The likelihood is approximated by a normal density, i.e., the log likelihood is quadratic.

VARCO6 (2.21) approximates means, variances and correlation for two jointly estimated variables, x and y , from six points (x, y) and associated likelihood values.