Linkage Newsletter

Vol. 6 No. 1 January 1992

Published by Jurg Ott, Columbia University, New York[1].

Editorial Assistant: Katherine Montague
Fax: +1-212-568-2750 Tel. +1-212-960-2507 e-mail: OTT@NYSPI.BITNET

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

EDITORIAL

The Linkage Newsletter is now going into its sixth year of publication. Last year only two issues were mailed out, but we are again planning on three issues for this year. As in the past, the Newsletter will focus on all aspects of interest to researchers in human linkage analysis with particular emphasis on statistical problems and computer programs. Comments from readers are always welcome.

Several ad hoc versions of the LINKAGE programs in the C language exist. An "official" C version is also forthcoming — Mark Lathrop and Peter Cartwright have a trial C version in operation. The next LINKAGE version will be available both in C and Pascal.

ADDRESS CONFIRMATION

We currently have close to 900 individuals on our address list and keep receiving new requests for the Newsletter. To allow us to consolidate our address list and remove any entries for people who are no longer interested in receiving the Newsletter, please fill out the questionnaire attached and return it to us, preferably by fax.

All recipients are requested to fill out the questionnaire or their names will be removed from the mailing list!

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

LINKAGE COURSES

This spring, two introductory linkage courses will be held, one in New York and one in Zurich, Switzerland. The dates are as follows:

Zurich: April 13-16, 1992, in the Computer Center at the University of Zurich, Irchel campus. This course is organized jointly with Prof. Eric Kubli in Zurich.

New York: May 19-22, 1992, in the microcomputer classroom of the Health Sciences library, Columbia University, New York.

Registration is open for both courses until one month prior to the beginning of each course, but we always have more applicants than space available, so we encourage a rapid response. For information and application forms, please write (preferably by fax) to Katherine Montague, course coordinator, at the address given above.

There will again be an advanced linkage course at Columbia University this fall but a date has not yet been fixed.

SOFTWARE NOTES

Bug in SLINK program

Dr. Dan Weeks at the University of Pittsburgh sent in the following text:

October 1, 1991: Clipping Bug in SLINK

There is a bug in SLINK dealing with how SLINK clips up through the pedigrees. This bug mainly causes problems if the pedigree is numbered so that a father is numbered before his child, who is numbered before its mother. Please correct your version of SLINK as indicated below or order a new copy from us.

Find this line in SLINK.PAS in the procedure collapseup:

```
IF ((p^.foff<>NIL) AND (p^.geneloc=0)) OR ((p^.foff<>NIL) AND (NOT noloop))
```

and replace it by:

```
{10/1/91 Correction to clipping problem }
{ Step 1: Find the spouse }
   IF (p^.foff<>NIL) THEN
    IF p^.male THEN
     q:=p^.foff^.ma
    ELSE
     q:=p^.foff^.pa;
{ Step 2: Collapseup if there are children AND either spouse }
{       has not yet been assigned a genotype }
   IF (p^.foff<>NIL) THEN
   IF ((p^.geneloc=0) OR (q^.geneloc = 0) OR (NOT noloop))
{Old code did not test whether spouse was done or not }
{IF ((p^.foff<>NIL) AND (p^.geneloc=0)) OR ((p^.foff<>NIL) AND (NOT noloop))}
```

Users of SLINK are encouraged to order the current program version. In the last few weeks several changes have been implemented which should make the programs more useful. Details are given below.

The default buffer size (128 bytes) for the input or output files (whichever is usually larger) was increased to 4K bytes. This resulted in the addition of an extra line of code in nonstandard Pascal.

The critical levels for the maximum lod score (previously fixed at 1, 2, and 3) are now user defined. At start-up of the analysis programs, the user is prompted to enter 3 critical values. To run the analysis programs in batch mode, one may invoke them, for example, as MSIM < MCRIT.DAT where MCRIT.DAT is a file with one line containing 3 critical values.

The quadratic interpolation routine, QUADMAX, in the MSIM program has been rewritten, because it occasionally seemed to give nonsensical results. It is now based on formulas (8.8)-(8.14) in Ott (1991) and interpolates the maximum lod score given three pairs of values (Z, Θ). It does not carry out any extrapolation beyond the smallest or largest Θ value. If the two interval lengths differ by more than a factor of 4, no quadratic interpolation is carried out, because the quadratic approximation may not fit the lod score curve very well in those circumstances.

Prospero Pascal and OS/2

All our Pascal programs have been updated to conform to standard Pascal as much as possible. Executable versions are compiled with Prospero Pascal because this compiler can produce programs running under OS/2. Most programs (eg. SLINK, TLINKAGE) are now available in compiled forms running under DOS and OS/2. The major advantages of running under OS/2 are that 1) programs can run in the background (in an OS/2 window) while one carries out other work in the foreground, and 2) larger problems can be analyzed under OS/2 than under DOS. For example, I recently used ILINK to analyze 1,000 pedigrees comprising a total of 22,000 individuals (2 loci, 2 and 4 alleles) with OS/2.

For a while we had some problems with the OS/2 versions. Every so often the system would halt a program and announce that it had led to a protection violation. After removing two compiler switches previously used (checking for stack space, initializing heap space to zero) these errors no longer occur.

Program bugs and problems

Users of the LINKAGE programs have reported the bugs and problems outlined below. I am grateful for any error reports — other people working in this field will surely appreciate being made aware of potential problems.

In the PREPLINK program, for *quantitative trait* locus types, a multiplier for the variance in heterozygotes versus that in homozygotes can be chosen provided that only two alleles and a single trait phenotype are defined. With more than two alleles or one trait, this multiplier has no effect in the program but still occurs in the datafile, which may be misleading (Joseph Terwilliger).

There is presently no check in the LINKAGE programs whether a loop in the data is made known to the programs or whether the number of loops declared exceeds the constant MAXLOOP. In either case, the program may terminate normally but provide an incorrect result. The latter problem is relatively easy to fix, but it will take some programming to create a loop detecting procedure. We are currently working on such a procedure. Some (but not all) undeclared loops lead to a program error in that the stack space is exhausted (Marcy Speer).

In the TLINKAGE programs, if two null loci (two-locus disease models) are specified, they must be listed in the locus order in such a way that the null locus with a higher locus number is listed after the other null locus. For example, if 1 and 2 are the null loci, locus orders 3 1 4 2 and 1 2 4 3 are all right, but the locus order 2 1 3 4 leads to a cryptic error message of *Array bound exceeded* (Chantal Mérette).

In the CFACTOR program, which is invoked before the CILINK program, when all family data are uninformative for linkage, the TEMPPED.DAT output file is empty. In that case, CILINK crashes when it tries to read TEMPPED.DAT. A check has been implemented in the Prospero Pascal version of CILINK to test for an empty TEMPPED.DAT file (Chantal Mérette).

In the LODSCORE program, in procedure getpen (file inl.pas), line 26 should be commented out or deleted. The line to be deleted is

```
      FOR i:=1 TO nallele DO
read(datafile,pen[0,i,2,1]);
```

(Marie-Claude Babron).

Tom Burroughs, Department of Psychiatry, Washington University, St. Louis, reports in their Linkage Bulletins two problems with the **LINKAGE programs on a Sun 4.1.1** computer:

Problem 1: PREPLINK won't read and write data for numbered alleles properly. Solution: A line of code must be added after the 101st line of procedure writedata. Below, the line to be inserted and the three lines preceding it are shown:

```
   write(outdata,contrait:7:3);
   writeln(outdata,' << MULTIPLIER FOR VARIANCE IN HETEROZYGOTES');
  END;
  numbers: begin end;  { <= insert this line}
```

Problem 2: LCP won't run properly in OpenWindows in a command tool. After typing lcp, about 12 lines are written to the screen at which point the screen locks up. Solution: Invoke LCP in a shell tool with the scroll mode off, rather than a command tool. While the highlighting and boxes which surround the menu do not appear as neatly as in the PC version, the program will run this way.

ELECTRONIC BULLETIN BOARD

As outlined in a previous issue of this Newsletter, the BIOSCI electronic newsgroup network comprises various newsgroups, some of which are of particular interest to people working in linkage and genome analysis. The network allows its users to contact people around the world without having to learn a variety of computer addressing tricks. Any user can simply post a message to his/her regional BIOSCI node and copies of the message will be distributed automatically to all other subscribers on all of the participating networks.

The following newsgroups may be of particular interest to readers of this newsletter:

NEWSGROUP NAME	TOPIC
CHROMOSOME-22	Mapping/Sequencing of Human Chromosome 22
GENETIC-LINKAGE	Newsgroup for genetic linkage analysis
HUMAN-GENOME-PROGRAM	Human genome issues, NIH sponsored

Information about BIOSCI may be obtained from biosci@genbank.bio.net; to subscribe to one of the newsgroups, send a message in plain English to one of the following addresses, depending on your geographic location:

Your location	Address
Continental Europe	biosci@bmc.uu.se
United Kingdom	biosci@uk.ac.daresbury
North/South America	biosci@genbank.bio.net

The GENETIC-LINKAGE newsgroup is an ideal forum for exchange of ideas and discussion of current topics.

BENCHMARK

In the May 1991 (Vol. 5 No. 2) issue of this Newsletter, run times (in seconds) of our new benchmark problem (available on disk 5b, directory *bench*) were shown. Dr. Dan Weeks provided the following additional information, where for comparison the first four lines are copied from the previously published table:

Machine	Run time
Toshiba 5200 (80386/7-20)	449
Vaxstation 3100 model 38	167
AT compatible 80486-33	98
Sun Sparcstation 1+	45
SUN SparcStation 2, 33MHz SUN Pascal 2.0	31
Macintosh IIfx (68030-40MHz MPW Pascal) LINKAGE 5.04	257

The SUN Pascal compiler seems very sensitive to the optimization level. At the default optimization, the compile time is much faster but the program then takes 82.3 seconds to run! The 30.2 "user" seconds is with the -O3 optimization level.

We often receive requests for information on **Macintosh** versions for the LINKAGE programs. Dr. Weeks started developing such a version while he was in New York and now has a test version working. However, only the analysis programs have been converted but not LCP or LSP. There is thus no very useful Macintosh version available. Please address all requests for information to Dr. Weeks at University of Pittsburgh, e-mail:

dweeks@watson.hgen.pitt.edu

QUESTIONS AND ANSWERS

Q_1: We are analyzing our family data with the help of LINKAGE, HOMOG, and LINKAGE UTILITY programs. For one highly polymorphic DNA marker, no evidence for linkage was found when the whole sample of 14 families was analyzed, although some individual families showed complete linkage. The HOMOG program provided considerable support for heterogeneity: p=0.0058 (H_2 vs. H_1), p=0.0316 (H_1 vs. H_0), and p=0.0037 (H_2 vs. H_0). After the exclusion of 7 families showing a conditional probability of linkage of less than 0.05, the maximum lod score for the remaining families reaches 5.09 at zero recombination fraction. Does this provide significant evidence for linkage?

A_1: The classical criterion for linkage is a maximum lod score of at least Z=3, which corresponds to a likelihood ratio for linkage of at least 1000:1 or a p-value of at most 10^{-Z}=0.001. In the presence of an admixture of linked and unlinked families, the overall maximum lod score may be small or even zero but a significant test for heterogeneity would clearly provide evidence for linkage (in some of the families). However, to retain the stringency of the criterion for linkage, the test for heterogeneity should only be declared significant when the associated likelihood ratio is at least 1000:1, which is not the case for your data. (For statistical reasons, the current HOMOG programs no longer provide p-values.) Since your results do point in the direction of heterogeneity, adding more families may well lead to significance.

Q_2: What is the interpretation of the generalized lod score calculated by ILINK? It reaches 8.07 in our family sample for a three-point analysis between the disease and two highly polymorphic marker loci.

A_2: For three loci, the generalized lod score calculated by ILINK is the logarithm to base 10 of the likelihood ratio, $L(\Theta_1, \Theta_2) / L(½, ½)$. It measures the overall evidence that the three loci are linked as opposed to being all unlinked. When one locus is a disease locus and the other two are markers, one usually knows that the markers are linked with each other. The question of interest is then generally whether the disease is linked with the two markers, given that the markers are linked. The generalized lod score overstates the evidence for linkage between disease and the markers. A better measure for linkage is obtained, for example, from the LINKMAP program, which assumes a fixed map distance between the markers.

CORRECTIONS IN NEW BOOK

A revised version of my *Analysis of Human Genetic Linkage* (1991 is now available for $47.50 from the publisher (Johns Hopkins University Press, Baltimore) or through bookstores. Readers (notably Drs. Deborah Meyers and Marcy Speer) have alerted me to the following missprints:

Page 38, Problem 2.2: Replace *200 cM* by *100 cM.*

Page 117, lines 21-23: The last sentence in this paragraph should read: The second child has genotype 121/222 or 122/221, *each of which requires at least one recombination in the father or the mother.*

Page 137, first line, should read:

...between the loci C and *D*.

Page 148, line 11: Replace f_{dd} by f_{DD}.

Page 149, table 7.1, line d1/d1: replace ½ by *½r* for P(g;r) (as on line above it).

QUESTIONNAIRE on Linkage Newsletter

Your name and address: (please print)

Would you like to stay on our mailing list? (yes/no)

Would you like to receive the Newsletter (mark none, one or both possibilities)

- by postal mail?

- by e-mail?
  (please provide your e-mail address)

Please return to:
Katherine Montague
Columbia University, Box 58
722 West 168 Street Fax: +1 (212) 568 2750
New York, NY 10032 Tel: +1 (212) 960 2507