# Linkage Newsletter

Editorial Assistant:  Katherine Montague
Fax: 212-568-2750
Tel. 212-960-2507
e-mail:  jurg.ott@columbia.edu

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

## E-mail address

Our old bitnet address (ott@nyspi) is no longer valid and will soon reject mail sent to it.  If you want to send us e-mail, please use only the address given above.  Also, for those who use e-mail but receive this newsletter by postal mail, please let us know -- we would prefer sending the newsletter by e-mail only.

## LINKAGE COURSES

The next two linkage courses are Advanced Courses.  They will take place as follows:

October 3-7, 1994, at the University of Zurich, Irchel Campus Computer Center (Switzerland;  maximum number of participants is 14), and

January 9-13, 1995, at Columbia University, New York (maximum of 20 participants).

To obtain information on these courses, please write to Katherine Montague, course coordinator, by e-mail or fax.  These courses are for researchers with experience in using the LINKAGE programs, or who have an otherwise excellent understanding of genetic linkage analysis.  Also, course participants must be familiar with PCs.  *Topics:*  Working with age at disease onset;  models for genetic heterogeneity;  genetic linkage and allelic association;  complex traits;  etc.  As usual, the main focus will be on practical exercises and linkage analyses carried out by participants on IBM PC's using the LINKAGE and other programs.  Each session will begin with a theoretical introduction on the material to be worked on.  We will use our new book instead of mimeographed handouts (Terwilliger and Ott, "Handbook for Human Genetic Linkage," Johns Hopkins University Press, 1994).  Next October, the book will be handed out at the course but for later courses (eg. January 1995), participants are expected to buy the book and bring it to the course.

The next introductory courses will be held in the spring or early summer of 1995, one each at Columbia University New York and at the University of Zurich (dates not yet set).

# SOFTWARE NEWS

## PC version 5.2 of LINKAGE

Because of as yet unresolved problems with version 5.2 of some LINKAGE programs (PC, Turbo Pascal), for general pedigrees we are now distributing version 5.1 while for the three-generation pedigrees version 5.2 is used.

## FastSLINK

As mentioned in the last newsletter, Drs. Daniel Weeks and Alejandro Schaffer have improved SLINK so that it runs considerably faster than the original version. This C-version is now on Dr. Week's ftp server, watson.hgen.pitt.edu. A brief description of the files available on that server follows:

```
----------------------------------------------------------------------
readme            :        Readme file for this directory
newapm.readme     :        Readme about the new version of the APM programs
newdosapm.tar.Z   :        DOS version of the new APM programs
newdosapm.zip     :   DOS version of the new APM programs
newdosapm.readme  :        Readme about the DOS version of the new APM programs
newdosapm.tar.Z   :        DOS version of the new APM programs
slink.tar.Z       :        SLINK simulation program (original version) Weeks DE,
                           Ott J, Lathrop GM (1990) Am J Hum Genet 47:A204
fastslink.tar.Z   :        Fast version of SLINK.  SLINK as modified by Alejandro
                           Schaffer and Daniel Weeks to use the algorithms in
                           FASTLINK v. 1;  cf. Cottingham Jr RW, Idury RM, Schaf-
                           fer AA, Am J Hum Gen 53(1993), 252-263
cintmax.tar.Z     :        Linkage analysis under mapping functions
simapm.tar.Z      :        SLINK-based simulation program for APM method (Unsta-
                           ble Unix-hack -- use at own risk)
----------------------------------------------------------------------
```

## Borland Pascal version 7.0

We are gradually converting all DOS Pascal programs to Borland Pascal 7. The main advantage is that programs may be compiled to run in protected mode so that they can make use of extended memory. Thus, for many users a heap space overflow will no longer be a problem. Programs may still be compiled to run in real mode (as in earlier versions of Turbo Pascal) in which case execution speed will be about 10% higher than when they run in protected mode. In either case, however, the usual restrictions of Turbo/Borland apply, in particular, the data segment is still limited to 64KB. Program version without these limitations are available for OS/2 (NDP Pascal).

## Bugs in CFACTOR program

In the three-generation programs (CLINKAGE), pedigree input files are preprocessed by the CFACTOR program such that larger numbers (20 or more on the PC) loci may be analyzed jointly. Dr. Weeks has reported two bugs that will result in somewhat errone-ous likelihoods. The bugs and their suggested corrections are given below.

2

## == Bug #1 ==

In each of the two procedures, getpaphase and getmaphase, the following two lines occur:

```
gpahet:=(pa^.gene1[i]<>pa^.gene2[i]) and gpaknown;
gmahet:=(pa^.gene1[i]<>pa^.gene2[i]) and gpaknown; <=BUG
```

The second line must be corrected in that all occurrences of 'pa' in the buggy lines should be changed to 'ma', since everything should refer to the grandmother, rather than to the grandfather. So, the second line, in each of the two procedures, should be

```
gmahet:=(ma^.gene1[i]<>ma^.gene2[i]) and gmaknown; <=CORR
```

## == Bug #2 ==

The procedure getdoublef factorizes (Lathrop, Lalouel, and White 1986) pedigrees in specific situations. Evidently, it does this also in cases where factorization should not be applied. The complete discussion of this bug is quite elaborate and is not reproduced here. Dr. Weeks recommends as a solution that the two lines indicated below should be deleted, that is, the curly brackets {} should be inserted in the lines identified by DELETE:

```
procedure getdoublef;
{Doublef contains information on factorization
 of completely informative loci}
var
  i,j : integer;
  paside1,paside2,maside1,maside2 : boolean;
begin
  FOR i:=1 TO nlocus DO doublef[i]:=false;
  FOR i:=2 TO nlocus-1 DO
    IF pahet[i] and mahet[i] and not intercross[i] THEN
      IF paphase[i] and maphase[i] THEN doublef[i]:=true
      ELSE
        {Factorize if phase unknown for surrounding loci
         on at least one side for each parent}
        begin
          paside1:=false;
          paside2:=false;
          maside1:=false;
          maside2:=false;
    {     IF not paphase[i] THEN   <= DELETE }
          begin
            FOR j:=1 TO i-1 DO
              IF paphase[j] THEN paside1:=true;
            FOR j:=i+1 TO nlocus  DO
              IF paphase[j] THEN paside2:=true;
          END;
    {     IF not maphase[i] THEN   <= DELETE }
          begin
            FOR j:=1 TO i-1 DO
              IF maphase[j] THEN maside1:=true;
            FOR j:=i+1 TO nlocus  DO
              IF maphase[j] THEN maside2:=true;
          END;
          IF not ((paside1 and paside2)
              or (maside1 and maside2)) THEN doublef[i]:=true;
        END;
  END;
```

## Bug in LINKMAP program

The report presented below has been submitted by Dr. Alejandro Schaffer. Over a year ago, an analogous report was sent to me by Dr. Weeks but at the time I considered it to be more a nuisance than a bug; at least it is an inelegant aspect of the LINKMAP program. Dr. Schaffer's report is as follows:

There is a bug in LINKMAP in LINKAGE 5.1 and LINKAGE 5.2 in which components of the theta vector which are supposed to be 0.0 are slightly positive. This occurs when that component is decreased from a larger positive number (e.g., 0.5) down to 0.0. What is particularly nasty about this bug is that whether it occurs or not depends on the number of steps of decrease as shown in the output transcript below. Note the non-infinite answer in the last run for the theta vector 0.112 0.210 0.000 in contrast to the infinite answer for what appears to be the SAME theta vector in the first two runs. The only difference between the runs is how many steps of decrease are used to reduce from 0.210 to 0.000. As my example shows, a bad consequence of this bug is that when the pedigree data implies that there must be a recombination and having the theta component be 0.0 should give a -infinity log likelihood, you instead can get a semi-plausible negative number. Furthermore, having the actual value be nonzero (when it should be zero) drastically slows down the computation for that theta. I have repaired the bug in LINKMAP and MLINK for what will be FASTLINK 2.2. I am not ready to release FASTLINK 2.2 because I want to take care of some other things that users requested.

NB: The transcript below was prepared with FASTLINK 2.1 (which still has the bug). I checked that LINKAGE 5.1 and 5.2 give the same buggy result on a Sparcstation 10.

Caution: The occurrence of this bug may be architecture-dependent as testing for 0.0 is well known to be a hard problem.

The program output referred to above is given in the Appendix.

## Bug in LOOPS program

Dr. Weeks reported a bug in the LOOPS program (OS/2 version only). Close to line 500, the following two lines occur:

```
readln(loopfile);
writeln(loopfile);
```

Correction:       The second line should be changed to
                  writeln(outfile).

Note (J.O.): This error evidently occurred in the adaptation from Turbo Pascal to NDP Pascal as it is not present in the DOS version. I am grateful to Dr. Weeks for pointing it out to me.

## No bug in Vax version of LINKAGE - and a word of caution

Over a year ago a 'bug' was reported here describing an apparent difference between the Vax and DOS versions of LINKAGE. The Vax version seemed to incorrectly produce a lod score of zero for a single doubly homozygous offspring of a phase-unknown double intercross mating in the presence of allelic association. The reason for the discrepancy between the two versions has now been traced to the program constant FITMODEL - it was set to false in the Vax version and to true in the DOS version. In the program description, this constant is said to have an effect in the ILINK program when more than just recombination fractions are to be estimated. It evidently has other effects as well. So, it is prudent to set FITMODEL equal to true at all times except in special circumstances.

# APPENDIX

```
*******************************************************************
Run   1 - LINKMAP : p12 p14 p1 p17
*******************************************************************
Program UNKNOWN version 5.1 (1-Feb-1991)
The following maximum values are in effect:
      10 loci
      55 single locus genotypes
      10 alleles at a single locus
     500 individuals in one pedigree
       3 marriage(s) for one male
       3 quantitative factor(s) at a single locus
      20 liability classes
      10 binary codes at a single locus
YOU ARE USING LINKAGE (V5.1 (1-Feb-1991)) WITH 4-POINT AUTOSOMAL DATA
Program LINKMAP version  5.10 (1-Feb-1991)
FASTLINK version  2.10 (21-Mar-1994)
The program constants are set to the following maxima:
       8 loci in mapping problem (maxlocus)
      12 alleles at a single locus (maxall)
     157 recombination probabilities (maxneed)
   50000 maximum of censoring array (maxcensor)
     180 haplotypes = n1 x n2 x ... where ni = current # alleles at locus i
   16290 joint genotypes for a female
   16290 joint genotypes for a male
    1000 individuals in all pedigrees combined (maxind)
     150 pedigrees (maxped)
      10 binary codes at a single locus (maxfact)
       3 quantitative factor(s) at a single locus
      20 liability classes
       3 quantitative factor(s) at a single locus
      20 liability classes
      10 binary codes at a single locus
    2.00 base scaling factor for likelihood (scale)
    2.00 scale multiplier for each locus (scalemult)
 0.00000 frequency for elimination of heterozygotes (minfreq)
YOU ARE USING LINKAGE (V5.1 (1-Feb-1991)) WITH  4-POINT
YOU ARE USING FASTLINK (V2.1 (21-Mar-1994)) AUTOSOMAL DATA
Number of alleles at locus 1 is 6
Number of alleles at locus 2 is 5
Number of alleles at locus 3 is 2
Number of alleles at locus 4 is 3
-----------------------------------
-----------------------------------
THETAS  0.112 0.000 0.210
-----------------------------------
PEDIGREE |  LN LIKE  | LOG 10 LIKE
-----------------------------------
PEDIGREE |  LN LIKE  | LOG 10 LIKE
-----------------------------------
        1  -203.121527   -88.214370
-----------------------------------
TOTALS    -203.121527   -88.214370
-2 LN(LIKE) =  4.06243E+02
Maxcensor can be reduced to        -32767
-----------------------------------
-----------------------------------
THETAS  0.112 0.210 0.000
-----------------------------------
PEDIGREE |  LN LIKE  | LOG 10 LIKE
-----------------------------------
        1 -100000000000000000000.000000 -43429355638650388480.000000
-----------------------------------
TOTALS    -100000000000000000000.000000 -43429355638650388480.000000
-2 LN(LIKE) =  2.00000E+20
Maxcensor can be reduced to        -32767
-2 LN(LIKE) =  2.00000E+20
Maxcensor can be reduced to        -32767
*******************************************************************
Run   1 - LINKMAP : p12 p14 p1 p17
*******************************************************************
Program UNKNOWN version 5.1 (1-Feb-1991)
The following maximum values are in effect:
      10 loci
      55 single locus genotypes
      10 alleles at a single locus
     500 individuals in one pedigree
```

6

```
         3 marriage(s) for one male
         3 quantitative factor(s) at a single locus
        20 liability classes
        10 binary codes at a single locus
        20 liability classes
        10 binary codes at a single locus
YOU ARE USING LINKAGE (V5.1 (1-Feb-1991)) WITH  4-POINT AUTOSOMAL DATA
Program LINKMAP version  5.10 (1-Feb-1991)
FASTLINK version  2.10 (21-Mar-1994)
The program constants are set to the following maxima:
         8 loci in mapping problem (maxlocus)
        12 alleles at a single locus (maxall)
       157 recombination probabilities (maxneed)
     50000 maximum of censoring array (maxcensor)
       180 haplotypes = n1 x n2 x ... where ni = current # alleles at locus i
     16290 joint genotypes for a female
     16290 joint genotypes for a male
     16290 joint genotypes for a female
     16290 joint genotypes for a male
      1000 individuals in all pedigrees combined (maxind)
       150 pedigrees (maxped)
        10 binary codes at a single locus (maxfact)
         3 quantitative factor(s) at a single locus
        20 liability classes
        10 binary codes at a single locus
      2.00 base scaling factor for likelihood (scale)
      2.00 scale multiplier for each locus (scalemult)
   0.00000 frequency for elimination of heterozygotes (minfreq)
YOU ARE USING LINKAGE (V5.1 (1-Feb-1991)) WITH  4-POINT
YOU ARE USING FASTLINK (V2.1 (21-Mar-1994)) AUTOSOMAL DATA
Number of alleles at locus 1 is 6
Number of alleles at locus 2 is 5
Number of alleles at locus 3 is 2
Number of alleles at locus 4 is 3
Number of alleles at locus 3 is 2
Number of alleles at locus 4 is 3
-------------------------------------
-------------------------------------
THETAS  0.112 0.000 0.210
-------------------------------------
PEDIGREE |  LN LIKE  | LOG 10 LIKE
-------------------------------------
         1  -203.121527   -88.214370
-------------------------------------
TOTALS    -203.121527   -88.214370
-2 LN(LIKE) =  4.06243E+02
Maxcensor can be reduced to        -32767
-------------------------------------
-------------------------------------
THETAS  0.112 0.105 0.133
-------------------------------------
PEDIGREE |  LN LIKE  | LOG 10 LIKE
-------------------------------------
PEDIGREE |  LN LIKE  | LOG 10 LIKE
-------------------------------------
         1  -201.053315   -87.316159
-------------------------------------
TOTALS    -201.053315   -87.316159
-2 LN(LIKE) =  4.02107E+02
Maxcensor can be reduced to        -32767
-------------------------------------
-------------------------------------
THETAS  0.112 0.210 0.000
-------------------------------------
PEDIGREE |  LN LIKE  | LOG 10 LIKE
-------------------------------------
         1 -100000000000000000000.000000 -43429355638650388480.000000
-------------------------------------
TOTALS    -100000000000000000000.000000 -43429355638650388480.000000
-2 LN(LIKE) =  2.00000E+20
Maxcensor can be reduced to        -32767
*********************************************************************
Run   1 - LINKMAP : p12 p14 p1 p17
*********************************************************************
Program UNKNOWN version 5.1 (1-Feb-1991)
The following maximum values are in effect:
        10 loci
        55 single locus genotypes
        10 alleles at a single locus
       500 individuals in one pedigree
```

7

```
      3 marriage(s) for one male
      3 quantitative factor(s) at a single locus
     20 liability classes
     10 binary codes at a single locus
YOU ARE USING LINKAGE (V5.1 (1-Feb-1991)) WITH  4-POINT AUTOSOMAL DATA
Program LINKMAP version  5.10 (1-Feb-1991)
FASTLINK version  2.10 (21-Mar-1994)
The program constants are set to the following maxima:
      8 loci in mapping problem (maxlocus)
     12 alleles at a single locus (maxall)
    157 recombination probabilities (maxneed)
  50000 maximum of censoring array (maxcensor)
    180 haplotypes = n1 x n2 x ... where ni = current # alleles at locus i
  16290 joint genotypes for a female
  16290 joint genotypes for a male
   1000 individuals in all pedigrees combined (maxind)
  16290 joint genotypes for a male
   1000 individuals in all pedigrees combined (maxind)
    150 pedigrees (maxped)
     10 binary codes at a single locus (maxfact)
      3 quantitative factor(s) at a single locus
     20 liability classes
     10 binary codes at a single locus
   2.00 base scaling factor for likelihood (scale)
   2.00 scale multiplier for each locus (scalemult)
0.00000 frequency for elimination of heterozygotes (minfreq)
YOU ARE USING LINKAGE (V5.1 (1-Feb-1991)) WITH  4-POINT
YOU ARE USING FASTLINK (V2.1 (21-Mar-1994)) AUTOSOMAL DATA
Number of alleles at locus 1 is 6
Number of alleles at locus 2 is 5
Number of alleles at locus 3 is 2
Number of alleles at locus 4 is 3
---------------------------------
Number of alleles at locus 4 is 3
---------------------------------
---------------------------------
THETAS  0.112 0.000 0.210
---------------------------------
PEDIGREE |  LN LIKE  | LOG 10 LIKE
---------------------------------
        1  -203.121527   -88.214370
---------------------------------
TOTALS    -203.121527   -88.214370
-2 LN(LIKE) =  4.06243E+02
Maxcensor can be reduced to       -32767
---------------------------------
---------------------------------
THETAS  0.112 0.070 0.163
---------------------------------
PEDIGREE |  LN LIKE  | LOG 10 LIKE
---------------------------------
        1  -201.247913   -87.400672
---------------------------------
        1  -201.247913   -87.400672
---------------------------------
TOTALS    -201.247913   -87.400672
-2 LN(LIKE) =  4.02496E+02
Maxcensor can be reduced to       -32767
---------------------------------
---------------------------------
THETAS  0.112 0.140 0.097
---------------------------------
PEDIGREE |  LN LIKE  | LOG 10 LIKE
---------------------------------
        1  -201.272873   -87.411512
---------------------------------
TOTALS    -201.272873   -87.411512
-2 LN(LIKE) =  4.02546E+02
---------------------------------
---------------------------------
THETAS  0.112 0.210 0.000
---------------------------------
THETAS  0.112 0.210 0.000
---------------------------------
PEDIGREE |  LN LIKE  | LOG 10 LIKE
---------------------------------
        1  -305.564464  -132.704678
---------------------------------
TOTALS    -305.564464  -132.704678
-2 LN(LIKE) =  6.11129E+02
```