

# Linkage Newsletter

Vol. 10 No. 3 December 1996

Published by Jurg Ott, Rockefeller University, New York  
Email: ott@rockvax.rockefeller.edu  
WWW: <http://linkage.rockefeller.edu>

Postal address:  
Rockefeller University, Box 192  
1230 York Avenue, New York, NY 10021-6399

Editorial Assistant: Katherine Montague  
Email: montagk@rockvax.rockefeller.edu

Fax: 212-327-7996  
Tel: 212-327-7979

(this and all previous newsletters are available on our ftp site/Web page)

## EDITORIAL

My collaborators and I have now completed our move to Rockefeller University (see new telephone and fax numbers above), where we form the new Laboratory of Statistical Genetics. Our web page (<http://linkage.rockefeller.edu>) and ftp site ([linkage.rockefeller.edu](ftp://linkage.rockefeller.edu)) are up and running. We welcome visitors to spend time with us — such interactions tend to be very fruitful.

## LINKAGE COURSES

The next linkage courses will be held as indicated below. Interested researchers are urged to **apply early** as our courses tend to fill up quickly. Dates for advanced courses (fall of 1997) are not yet known (advanced courses are reserved for individuals with a very good background in linkage analysis).

March 10-14, 1997, at University of Zürich, Switzerland (basic course, maximum of 18 participants). Deadline for applications is January 15, 1997, which is when applicants will be assigned space in the course; applications submitted after the deadline will still be considered but have a much smaller chance of success.

June 23-27, 1997, at Rockefeller University (basic course, maximum of 30 participants). Deadline for applications is May 3, 1997.

To obtain information on these courses, please write to Katherine Montague, course coordinator, by email (preferred) or fax.

We will use our book (Terwilliger and Ott, *Handbook of Human Genetic Linkage*, Johns Hopkins University Press, 1994), with supplemental handouts for advanced courses. Participants are expected to buy the book and bring it to the course; in case of problems please contact Katherine Montague in advance of the course. A list of corrections for the book may be downloaded from our anonymous ftp site, [linkage.rockefeller.edu](ftp://linkage.rockefeller.edu) (file *corr\_ter.txt* in directory *book*).

# SOFTWARE

## ftp sites

Please note, as mentioned above, that our current ftp site is

linkage.rockefeller.edu

Our previous site, linkage.cpmc.columbia.edu, is now being used by Dr. Joseph Terwilliger at Columbia University, where he is making his programs available for distribution.

## Bug Report

(Dr. Daniel Weeks, University of Pittsburgh and Wellcome Genome Center, Oxford)

27 Nov 1996

### FASTSLINK and SimIBD bug report

Thanks to the help of Hakan Sakul at Sequana, we have tracked down a subtle bug in FASTSLINK (which was due to a one letter typo). The source code on the ftp server watson.hgen.pitt.edu has been corrected, so you can just re-ftp the software from there (or from the UK mirror site ftp://ftp.ebi.ac.uk/pub/software/linkage\_and\_mapping/hgen\_pitt/).

Since SimIBD uses FASTSLINK to do the simulations, SimIBD is influenced by this bug too. Please re-ftp the SimIBD package also if you use it.

NOTE: We have developed a diagnostic version of FASTSLINK which allows you to easily check if your pedigrees were such that your results would have been influenced by this bug.

### MANUAL CORRECTION OF FASTSLINK CODE

To correct the FASTSLINK code itself, instead of ftp'ing the whole package, you may just correct the one line that is in error as follows:

The bug is a typo at line 741 of slautomodified.c.

The current incorrect code reads:

```
parunk = (((*LINK->p)->geneloc == 0) || ((*LINK->p)->geneloc == 0) || !nolop);
```

It should read:

```
parunk = (((*LINK->p)->geneloc == 0) || ((*LINK->q)->geneloc == 0) || !nolop);
```

### EFFECT OF THE BUG

The error causes the segdown routine to do the wrong thing under some circumstances. At first glance this looks like a bug that should have shown up a lot sooner. However, it won't show up in FASTSLINK if:

1. Spouses always get consecutive numbers.  
or
2. The pedigree has a loop  
or
3. The pedigree is simple and numbered top to bottom in increasing order (i.e., children always have ids larger than their parents).

The problematic line of code is trying to determine if both parents have already had their genotype selected. In segdown, one updates probabilities of a child based on parents and siblings. If both parents have their genotype known, then the siblings can be ignored. This is advantageous because the code for the one child case (whether in the original or fast versions) is faster than for the many child case.

In segdown p is always the father and q is always the mother.

The problem arose only in the loopless case. The effect of the typo was to go into the one child case if the father already had a genotype assigned. This would be erroneous if the mother did not have a genotype yet.

The effect of the error depends on the random seed, which is why Hakan Sakul's pedigree showed different behavior with different seeds.

In some cases, ignoring the effect of the siblings causes an incompatible genotype to be assigned, which would eventually lead to a crash.

In other cases, a compatible-with-siblings genotype would be assigned by chance and the pedigree replicates would come out compatible, but sampled from a wrong distribution, which is extremely unfortunate.

Because FASTSLINK fills in genotypes in increasing order of ids, the only way the problematic code could have an effect is if there is a nuclear family where the father has a lower id than the mother. BUT, the pedigree traversal algorithm ensures that the call to fill in the proband is a segup call, not segdown. Therefore, if mates always have consecutive numbers each traversal has at most one nuclear family with father filled in and mother not AND the mother is the proband for that call so she gets filled in in segup and not segdown.

If the pedigree is simple and numbered from top to bottom, segdown is never used.

For SimIBD, some people are skipped when simulating, changing the normal order in which slink does the simulations. Therefore, even in a nuclear family, if p is the mother and if the children are affected but the mother is unaffected (and skipped), then the bug is hit. However, if the father is unaffected, then the bug is not hit. Therefore, because SimIBD forces FASTSLINK to do the calculations in a different order from the original FASTSLINK, SimIBD results might be affected in other situations as well.

#### **A DIAGNOSTIC VERSION**

We have constructed a diagnostic version of FASTSLINK, which you can make by typing

```
make slinkdiag
```

Then you can run the diagnostic version on each pedigree you used to see if any would hit the bug. To do this, set up as you would for a normal SLINK run, but type 'slinkdiag' instead of 'slink'. Note that it is only necessary to generate 1 replicate per pedigree to test for the effect of the bug.

If the bug is hit, then you will see a message appear on the screen like:

```
BUG WAS HIT: Pedigree 3 Person 255
```

## ANSI.SYS driver in Windows 95 and NT

We keep receiving requests for support because “the LCP program produces only gibberish on the screen”. The reason is a missing *ansi.sys* driver and not, as some users concluded, that the programs don't run under Windows 95 or NT. Currently, to run the LINKAGE programs under Windows 95 or NT, a DOS window must be opened and the programs run there. We are in the process of adapting them to Windows NT using the corresponding new NDP Pascal compiler. Readers of this newsletter will be notified when work is completed.

*Windows 95:* As is mentioned in the *msdosdrv.txt* file in the *windows* directory, the *ansi.sys* driver must be installed by inserting a line in the *config.sys* file as follows:

```
device=c:\windows\command\ansi.sys
```

The *config.sys* file is in the root (top) directory and may initially be empty.

*Windows NT:* In version 3.51 of NT, you open the Control Panel, double click on System, and click on Environment. Then, you define a new variable, *device*, and give it the value *c:\winnt351\system32\ansi.sys* (this is how I remember it). In version 4.0, according to the documentation, the *config.nt* file must be changed in a similar manner as for Windows 95.

## Computer program for identification of relationship errors in sib pair studies

(Harald H. Göring)

In affected sib pair studies, the power to detect linkage will generally be decreased if non-sibs, which were erroneously believed to be sibs, are present in the collected families. While such relationship errors will often be detected through mendelian inconsistencies, this is typically not the case in studies of late onset diseases when the parents of affected individuals are already deceased. To detect non-sibs also in this situation, I wrote a computer program called *relative*. A paper describing the method has been submitted for publication and an abstract has been published: Göring HHH & Ott J (1995) Verification of sib relationship without knowledge of parental genotypes. *Am J Hum Genet* 57, A192.

The *relative* program computes probabilities that two stated sibs are in reality unrelated, half-sibs or sibs. This is accomplished by computing the likelihoods for the three relationships from the observed genotypes of the two individuals. For this estimation, the same genotypes may be used that were obtained by marker typing for linkage analysis. The more markers are available, the better the estimation procedure performs. Linked markers are handled correctly by the program, and a single parent may be typed alone. In addition to standard LINKAGE-format input files, the user must supply prior probabilities for relationship errors. The program then identifies in the pedigree file all typed sib pairs that do not have both parents typed and carries out computations for each sib pair in turn. After non-sibs are identified by the program, the investigator may remove them from the sample of collected families. This tends to increase power for linkage detection. Note that the program can also compute likelihoods for some other common relationships.

The *relative* program is distributed on our ftp site (linkage.rockefeller.edu) in directory /software/relative.

## **Request for Information**

(Dr. Wentian Li)

Under the web page of our laboratory at Rockefeller University, there is an “alphabetic list of genetic analysis software” which lists more than 110 linkage analysis, genetic map construction, and pedigree drawing programs. The URL (uniform resource locator) is:

<http://linkage.rockefeller.edu/soft/list.html>

For each program, I try to include information on author, ftp site, web home page, source code language, references, etc. For many programs I cannot find all of this information. To keep the list up-to-date, it would be very helpful if the authors of these programs could check the list and send us any missing information, missing links, and new release information (please send email to wli@linkage.rockefeller.edu).

The current format is the following:

1. Full program name: e.g. if you have a program called APM, identify it as “Affected Pedigree Member (method)”.
2. Version: e.g. v3.8 (Dec 1996)
3. Description: any description of what the program does.
4. Authors: ...
5. Web: home page if any.
6. FTP: ftp site if any.
7. Source code language: e.g. Fortran, C, C++...
8. Operating systems: e.g. DOS, Windows 95, Windows NT, UNIX (SunOS, OSF1). This list need not be complete, but at least list the operating systems on which the program is known to be working.
9. Executables: if the executable name is different from the program name it is advantageous to have these names.
10. References: major publications relevant to the program such as the program announcement or the theoretical paper on which the program is based.
11. Price: if it is a commercial package.

A more up-to-date list will benefit everybody, and your help will surely be appreciated.

## **Genehunter update — version 1.1**

Several changes have been made to the Genehunter software and are included in version 1.1 -

currently available via anonymous ftp at [genome.wi.mit.edu](ftp://genome.wi.mit.edu) in `distribution/software/genehunter`.

\* We implemented an algorithmic improvement which should increase analysis speed by a factor of 5-10 in many cases. The speed-up was motivated by the work of Ramana Idury, *Genetic Epidemiology* 13:A311, 1996.

\* We implemented a fix to allow compilation under AIX.

In addition, users should note that the use of the KOSAMBI map function can produce slight numerical errors in LOD score values in intervals between markers. We recommend that the HALDANE map function, which is multilocus feasible, be used for all analyses and that distances between markers be input as recombination fractions or in Haldane centiMorgans. Thanks to Dan Weeks for bringing this to our attention in an example where a pedigree with all members designated "phenotype unknown" produced a LOD score of 0.03 (instead of 0) between two markers.

The Haldane map function is the default in the new version of Genehunter. Users of the old version, in which Kosambi is the default, should download the new version as soon as possible. In the meantime, they are advised to use the Haldane map function in all analyses by issuing the 'map function haldane' command when the program is started up.

Information regarding Genehunter can be obtained at the above ftp site or by reading our article in the *American Journal of Human Genetics* (58:1347-1363 June 1996). If you have problems or questions regarding the software contact:

Leonid Kruglyak      (leonid@genome.wi.mit.edu)  
Mark Daly            (mjdaly@genome.wi.mit.edu)

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