

# Linkage Newsletter

Vol. 7 No. 1 February 1993

Published by Jurg Ott, Columbia University, New York.

Editorial Assistant: Katherine Montague

Fax: +1-212-568-2750 Tel. 212-960-2507

e-mail: ott@nyspi.bitnet or jurg.ott@columbia.edu

Postal address: Columbia University, Unit 58

722 West 168th Street, New York, NY 10032

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

## EDITORIAL

As linkage analysts buy more and more powerful computers they also try to run larger problems than before. Under MS-DOS on a PC, one often runs into restrictions imposed by DOS or Turbo Pascal such that no analysis or only an approximate analysis is possible. We keep trying to improve this situation for PC users. The 80486 machines are now quite powerful and it is more a matter of using appropriate software to tap the full potential of these machines. We now have considerable experience with OS/2 and to a lesser degree, with Windows 3.1. OS/2 version 2 clearly seems the more stable platform, particularly now that Corrective Service Diskettes are available from IBM, which eliminate problems encountered in the original version 2 of OS/2.

We started using the NDP Pascal compiler from Microway (see advertisements, for example, in BYTE magazine, Feb. 1993, page 160) and are impressed with its potential. It has none of the restrictions that plague users of Turbo Pascal and, to some degree, also of Prospero Pascal. The company is very responsive to inquiries, which unfortunately seems less the case for the makers of Prospero Pascal in England. Currently, only a beta test version of NDP Pascal is available for OS/2 but the full version might be on the market by the time you read this.

We are also in the process of porting the LINKAGE programs to Windows under Borland Pascal version 7. However, Borland Pascal has restrictions similar to those of Turbo Pascal except that presumably, much more memory can be allocated under Windows than under DOS. On the other hand, arrays still cannot exceed 64KB in size, etc.

Among other developments at Columbia, we are working on setting up an anonymous ftp site for program distribution. Also, programs will be available not only for DOS but also for Unix and VMS machines (see below for some existing ftp sites).

## LINKAGE COURSES

Due to time constraints, no introductory linkage course could be scheduled in Zürich for this spring. However, the following two courses will be held in 1993:

New York (introductory course), at Columbia University: May 17-21, 1993.

Zürich (advanced course), at the Computing Center of the University of Zürich Irchel campus: September 27 - October 1, 1993.

Registration is open for both of these courses. For information and application forms please write to the address above, preferably by fax. The *Columbia University Advanced Course* for the academic year 1993/4 will be held in January of 1994 but a date has not been fixed yet.

## HINTS FOR INSTALLING AND USING OS/2

The following experiences with setting up OS/2 may be useful to the readers of the Newsletter.

On many machines other than IBM's, it is preferable to make changes to the BIOS before installing OS/2. In our experience the most important point is that an external disk cache be disabled (OS/2 provides its own disk cache, which works well). Also, installation is easiest when you allow the installation program to format the hard disk while it installs OS/2. If you want to install OS/2 without reformatting the hard disk, the following steps are recommended: 1) Make a boot disk (floppy) for DOS and try it out; that is, you must be sure you can boot DOS from the floppy drive. 2) Delete any unneeded files. 3) Remove all file fragmentation and make all files contiguous on the disk by using, for example, the Norton SpeedDisk program (use full optimization). 4) Turn off your computer, insert the OS/2 installation disk, and turn your computer on again.

The program will ask you whether you want to install everything or only a selection of features. I would choose the latter. For example, I would NOT install fonts or games. This way you only require approximately 25MB of disk storage for the system. As you select and deselect features, the program displays how much space is available on your disk and keeps a running tally of how much disk space is required for the current selections.

A major decision is whether you want to use the high performance file system (HPFS) or the old-fashioned file allocation table (FAT) system. For compatibility with other programs, particularly when you want to boot native DOS, FAT is preferable although it suffers from the well-known problem of file fragmentation. Once you use OS/2, you may occasionally encounter a problem with extended file attributes that OS/2 uses but DOS does not. For example, you may be unable to delete a file because it is cross-linked with another file's extended attribute. There is an easy solution: just run OS/2's `chkdsk` program as often as is required to get rid of the problem. Some of these problems cannot be fixed by `chkdsk` if OS/2 was booted from the hard disk (as is usually the case). Then, shut down OS/2 (keep the cursor on a free space of the Desktop and press the right mouse button), put the OS/2 installation disk into the A: drive, and reboot. After you insert the second floppy (disk no. 1), press Esc when the program asks whether you want to continue installing. You are then left with a working version of OS/2 and should see the prompt, [A:>]. Now, enter, for example, `C:\OS2\CHKDSK D:` if your OS/2 system resides on the C: drive and you want to check problems on the D: drive. Alternatively, insert disk no. 2 of the installation package (it contains `chkdsk.exe`) into the A: drive and enter `A:CHKDSK D:.` You may need to issue this command repeatedly until `chkdsk` no longer reports a problem (it seems to fix only one problem at a time).

You may switch between various DOS and OS/2 windows. However, be aware that whatever windows are open consume a certain amount of RAM, which is lost to other windows. Also, the DOS windows by default have 2MB of extended and expanded memory each, which is usually much too much. To adjust these settings for a given DOS window, first exit from this window (it must not be active), then click on its icon using the right mouse button. A small window appears, which says "Open" at the top. With the left mouse button, click on the arrow in that row (be sure it's the arrow). Then click on Settings, then on Session, and then on DOS Settings. The most important DOS settings to change are EMS\_MEMORY\_LIMIT and XMS\_MEMORY\_LIMIT. Also, if you want to operate a modem from this window (it is best to reserve one window for running your communications program such as Kermit), set IDLE\_SENSITIVITY to 100; this will ensure smooth operation of your communications program. Note that these changes cannot be made on a window while it is open.

Under OS/2, one has even more control over programs running in a DOS window than when they run under native DOS. For example, if a program is caught in a loop and you are unable to interrupt it, you can simply close the window in which it is running. Under DOS or Windows 3.1, one must reboot the machine.

In our experience, there are very few DOS programs that do not work properly in DOS windows of OS/2. Some of the newest Norton Utilities do not work properly (but FileFind and FileSize work fine). To occasionally use such programs, one simply boots the machine under DOS from a floppy disk.

While OS/2 emulates DOS version 5 very closely, we have found one difference thus far: Backup and Restore are different enough that files backed up in a DOS window under OS/2 cannot be restored under native DOS, and vice versa.

The current OS/2 version supports Windows 3.0. We have successfully installed several Windows programs in this "Windows" version. One program, SYSTAT, does not work properly this way although it works fine under regular Windows 3.1.

## SOFTWARE NEWS / BUG REPORTS

### Version 5.2 of LINKAGE

Mark Lathrop has released version 5.2 of LINKAGE. The new programs are available from Mark Lathrop and will soon be available from us. We are presently running some tests. Preliminary benchmark runs (see Linkage Newsletter, May 1991) show the following results (times given in seconds for two likelihood calculations, run on an 80486 25MHz):

Version 5.1	Prospero and Turbo Pascal	DOS	121 sec.
Version 5.1	NDP Pascal	OS/2	71 sec.
Version 5.2	NDP Pascal	OS/2	71 sec.

Line 1 versus 2 shows the greater efficiency of NDP versus Prospero and Turbo Pascal. Line 3 says that, for our benchmark data set, the new version is about as fast as version 5.1; it may, however, be faster for other data sets.

Incidentally, we also ran the benchmark data set using the MENDEL program. It was compiled with Microsoft Fortran 5.1 such that it ran under DOS or OS/2. Because of the array sizes, which are required by MENDEL for the given data set, MENDEL was unable to run under DOS. Under OS/2,

it required approximately 12MB of RAM to run and took 1298 seconds to complete (run on an 80486 with 16MB RAM to prevent usage of virtual memory). The MENDEL program can thus be quite slow in the presence of many untyped individuals. On the other hand, it is more flexible than LINKAGE in the problems that a user can address.

In this context, the time requirements for the benchmark data set on three other machines are of interest (reported by Iain Fenton, Cardiff). The following times represent elapsed time, not CPU time:

DEC 5830, running Ultrix 4.2, 3 CPU's, 128 MB memory (with approx. 50 interactive users)	40 sec.
DEC VAX 6000-400 cluster, running VMS 5.4, 2 CPU's, 96+32 MB memory (with approx. 20 users)	96 sec.
Viglen VigI, running MS-DOS 3.30, 9.54MHz 8086, no numeric coprocessor, 640K memory	6.8 hours

### Sensitivity Analysis Programs

(contributed by Drs. David A. Greenberg & Susan E. Hodge)

SENSEN and SENPED are short Fortran programs designed to facilitate basic sensitivity analyses of families, as described in Hodge and Greenberg [1].

SENSEN takes a standard LIPED input file with data for one family and prepares equivalent LIPED input files, reversing the affectedness status at the main trait for each family member, one at a time.

SENPED takes the lod file output of LIPED runs on all the sensitivity files for a single family (and a single marker) and creates an input file for the Pedigree/Draw program [2], showing the original lod score and the difference in lod score caused by each change in affectedness status.

SENSEN and SENPED are available from  
David A. Greenberg, Ph.D.  
Department of Psychiatry, Box 1229  
Mt. Sinai Medical Center  
1 Gustave Levy Place  
New York, NY 10029

e-mail: miriam@onion.salad.mssm.edu

---

[1] Hodge SE and Greenberg DA (1992): Sensitivity of lod scores to changes in diagnostic status. *Am. J. Hum. Genet.* 50:1053-1066.

[2] Pedigree/Draw is a set of shareware programs for the Apple Macintosh used to prepare genetic pedigrees. For further information about this program (and how to obtain a copy), contact:

Paul Mamelka  
Department of Genetics  
Southwest Foundation for Biomedical Research  
P.O. Box 28147  
San Antonio, TX 78284  
Internet: paul@darwin.sfbr.org

### Iterating on $x_f/x_m$ in ILINK

With only two loci, when both male and female recombination fractions should be iterated on, ILINK will work in the usual way with the variables  $\Theta_m$  (male recombination fraction) and  $R=x_f/x_m$  (female-to-male map distance ratio). There are now two ways of treating  $R$ : as a fixed ratio (the same in all intervals) or as a variable ratio. With only one interval, it might appear that it does not make any difference what one chooses, and this is the case on Vaxstations and on the Sparcstation. On the PC, however, *variable ratio* should be chosen. If the female recombination fraction estimate turns out to be equal to zero, ILINK reports an incorrect lod score on the PC. In the example that occurred in one of our courses, the incorrect lod score reported was 0.86, but with a variable  $R$ , ILINK gave the correct lod score of 4.67. The difference must be due to the way the LINKAGE programs compute likelihoods under the hypothesis of no linkage. With  $R$  exactly equal to zero, the female recombination rate is evidently always set equal to zero even under the assumption of no linkage.

### Bug in LSP under DOS

With a large number of codominant loci (more than about 15), when both Allele Numbers and Binary Factors locus types occur in the datafile, the LSP program produces an erroneous datafile output. For example, towards the end of the new datafile created, there should be a line containing as many recombination fractions as there are locus intervals; that number is not right, which will cause a linkage run to abort. The problem does not occur with the LSP versions on DEC or SUN machines and is restricted to the DOS version. Peter Cartwright has been looking into the problem but thus far has not seen a solution to it. We are planning to compile LSP with different C compilers to see whether that might cure the problem.

### New programs

The programs listed below have recently been developed by Xiaoli Xie and may be of interest to linkage analysts. For a detailed description, please ask for our list of programs.

*TypeNext* implements a special version of the SLINK program. For a number of untyped individuals in a pedigree, it estimates which individuals should be typed next to gain the most informativeness for linkage analysis (*Am J Hum Genet* 51 (suppl), A197).

*VaryPhen* varies the phenotype (affected/unaffected) for each individual and reports the change in maximum lod score (*Am J Hum Genet* 47, A205, 1990).

*LOOPS* checks for undetected loops remaining in the data after a pedigree file has been processed by the *MAKEPED* program. The *LOOPS* program is now part of the *LINKAGE* package and is automatically invoked whenever one calls *MAKEPED* (*Am J Hum Genet* (suppl) 51, A206, 1992).

## USEFUL E-MAIL ADDRESSES

We frequently receive requests for information on how to obtain Unix versions of the *LINKAGE* programs. The ftp site mentioned below contains these and other programs for Sun machines. To download programs using ftp, proceed as follows (directions taken from document obtained from that site):

```
ftp corona.med.utah.edu or ftp 128.110.231.1
```

When prompted to provide a user name, enter "anonymous". As the password, give your last name.

Then, issue the commands

```
cd pub/linkage/sun
```

```
binary
```

```
get linkage.tar.Z
```

```
quit
```

This ends your ftp session. On your Sun machine, issue the commands

```
uncompress linkage.tar.Z
```

```
tar xvf linkage.tar
```

```
rm linkage.tar
```

We also keep getting requests for information about pedigree drawing programs for the PC. Several programs exist, some have been discussed in this newsletter. As an example, the *PEDRAW* program by Dr. David Curtis ([dcurtis@crc.ac.uk](mailto:dcurtis@crc.ac.uk)) may be obtained from various anonymous ftp sites such as <ftp.embl-heidelberg.de> or <ftp.bio.indiana.edu>.