Linkage Newsletter

No. 1 August 1987

PREFACE

For several years now, some of us working in the area of linkage analysis have been playing with the idea of circulating a newsletter as a means of rapidly exchanging the kind of information that is of importance for one's work but that is not suitable for publication in a regular journal. While previous attempts at issuing a genetics newsletter have not been very fruitful, this represents at least an initial effort at creating and maintaining a viable means of communication among researchers interested in linkage analysis. Presently, this newsletter is mailed to about 200 researchers in all continents.

Below, I propose a general outline for the newsletter; any suggestions from readers will be gladly received.

OBJECTIVES OF NEWSLETTER

1. Basically, the newsletter should readily disseminate pieces of information relevant to genetic analysis that cannot usually be published in scientific journals, where emphasis is on linkage analysis. This comprises, for example, proposals for new methods, notes on computer programs, computer hardware, meetings, experiences with analysis methods and -- last but not necessarily least -- gossip.

2. Readers are encouraged to submit articles or notes to be published in the newsletter, and I will make every effort at quickly including such contributions. I will not exercise any censorship, except that advertisements for commercial products may have to be paid for (fee yet to be determined). When announcing computer programs you will have to specify at what cost they are available.

3. A Question and Answer section is planned in which I will discuss questions I have been asked and what I consider as meaningful answers. Again, contributions from readers to this section are welcome.

4. For electronic communication among geneticists worldwide, the Bitnet/EARN network presently appears most suitable and least expensive (actually, it is free when you have an account at a computer that is a Bitnet node). At regular intervals,

therefore, Bitnet addresses of geneticists will be published in the Newsletter.

5. Funding for such an undertaking is difficult to obtain. Therefore, presumably, I will be forced to ask for a modest contribution, but this and one or two subsequent issues are distributed free of charge. Please feel free to copy the Newsletter and distribute copies to anyone interested. There is no copyright to any article or note appearing in the newsletter. If the Newsletter was not addressed to you but you are interested in receiving it directly from me, please send me your address.

SOFTWARE NOTES

Recently, Dr. Eric Lander kindly gave me a demonstration of his **MAPMAKER** program. It is written in C and presently runs under Unix on the HP9000 minicomputer. It is operated through a shell of commands that make it very easy to select markers, try various gene orders, maximize the likelihood under a given order, etc. I was impressed by the high degree of user-friendlyness of the program. One may remember from Dr. Lander's presentation at the Philadelphia meeting last fall that MAPMAKER allows analyzing large numbers of codominant loci in families of two or three generations.

Regarding the Linkage program package (Lathrop et al.), a new version (4.5) is being distributed. A shell to allow for user-friendly operation will be added in the next few weeks. The whole package will be presented in Paris during the Human Gene Mapping workshop, in the CEPH building. Version 4.5 has faster execution speed than previous versions. The corresponding programs for small family structures allow analyzing many more loci. Again, Turbo versions will be available for the IBM PC and AT. A separate program will be available to carry out error checks, for both datafile and pedfile, and crosschecks between datafile and pedfile.

We have recently been using the **SIMLINK** program of Dr. M. Boehnke quite a bit. It approximates the expected lod score for a pedigree with given disease genotypes but as yet unknown marker genotypes, that is, it indicates the usefulness of a pedigree for linkage analysis prior to collecting marker data. We apply it to determine which of several pedigrees is most suitable for typing; also, the program seems a nice tool for grant proposals. Presently, it does not allow for reduced penetrance so that, for example, with X-linked inheritance, a mother with an unaffected son cannot be a possible gene carrier.

QUESTIONS AND ANSWERS

Q1: I have a marker locus with 2 codominant alleles. For some pedigree members, however, I can only determine whether allele 1 is present or not. In LIPED, no problem exists to accommodate such phenotypes, but is there such a possibility to allow for both kinds of phenotypes (codominant and dominant) in a binary factor locus in the Linkage programs?

A1: Not directly, as far as I know. However, a simple general method of converting any locus in LIPED format to Linkage format is the following. As an example, consider a marker with 2 alleles, A and B, and phenotypes as in the following table:

	Phenotypes				
Geno-	Dominant		Codominant		
type	A+	A-	A	AB	B
A A	1	0	1	0	0
A B	1	0	0	1	0
B B	0	1	0	0	1

The body of this table represents what would go into LIPED as part of the locus description. In the Linkage programs (datafile), define this locus as an affection status locus with as many liability classes as there are columns in the table above, and enter the penetrances appropriately, that is, for each genotype, you will need to furnish 5 penetrances. Each phenotype (in pedfile) is then represented by two numbers, 2 i, where i is the column number in the table above, that is, each individual is defined to be affected, except that the unknown phenotype is coded as 0 i, i being any integer from 1 to the number of liability classes. If the marker is on the X chromosome, the penetrances given for males at the AB genotype are irrelevant. While this scheme is generally applicable, in specific cases it is possible to use coding schemes requiring a much smaller number of penetrance classes.
(JO)

Q2: In LIPED, to express presence or absence of a particular allele of an X-chromosomal marker, I use codes + and - for males, and ++, +- and -- for females. How do I do this in the Linkage programs?

A2: There is no need to distinguish these phenotypes for the sexes, in any program, since these codes refer to phenotypes rather than genotypes. Thus, you may as well use +, +- and - for either sex where, of course, +- does not occur in males. In the Linkage programs, to express presence of allele number 1 at a binary factor locus, the corresponding codes would be 1 0, 1 1 and 0 1, respectively, where, again, 1 1 would not occur in males.

PRACTICAL HINTS

This section might prove useful to many readers. Please submit examples of your own that you feel other researchers could benefit from.

H1: **Running the Ilink program with new data:** With new data and several marker loci, it is often useful to first find or confirm estimates of recombination fractions between adjacent marker loci, that is, to run the Ilink program for the marker loci only. However, before doing that, it is a good idea to do one run with the Mlink program to verify that the likelihood is nonzero in all pedigrees. If the likelihood is zero in one or more pedigrees, for example, due to genotype inconsistencies, then the Ilink program will still try to maximize the likelihood and will, of course, fail but only after running for a possibly very long time.

PERSONAL NOTES

Dr. Lodewÿk Sandkuyl, M.D., from Rotterdam, is spending one year as a visiting scientist at Columbia University Department of Psychiatry.

Jurg Ott
Columbia University, Box 58
722 West 168 Street
New York, NY 10032

Tel. (212) 960-2504
Bitnet: OTTSR@CUVMB

Linkage Newsletter

Vol. 1 No. 2 Dec. 1987

EDITORIAL

The response to the previous, first issue of the Newsletter was extremely positive. I received many supportive letters from various countries -- thank you all for writing. This positive response reinforces my impression that the Newsletter fills a need in our scientific community.

Several readers of the last issue of the Newsletter inquired about availability of the computer programs mentioned. While I will continue discussing software developed by anyone, I can only distribute programs developed by me or my collaborators (a few exceptions are some generally available programs of a more technical nature, such as the Kermit program for computer communication, or the PC-WRITE text editor). Programs developed by other researchers will have to be ordered from them directly. Note that the Mapmaker program, mentioned in the last issue of the Newsletter, has just been described in detail (Genomics 1, 174-181, 1987).

Readers who would like to make a written contribution to the Newsletter are encouraged to write to me (I cannot, however, absolutely guarantee publication). Texts longer than half a typewritten page are requested in the form of an ASCII (text) file on a floppy disk, 5 1/4", IBM PC format. The author's name will appear as given on the text submitted (initials or full name, with or without address).

When the International Workshop on Human Gene Mapping was held in Paris last summer, several useful computer programs developed by human geneticists were demonstrated at the CEPH building. There were database programs for linkage analysis, programs to draw pedigrees on screen or on a matrix printer, and so on. I would like to compile and maintain a list of available programs and publish this list at regular intervals as part of the Newsletter. To be able to do this, I need the cooperation of the authors of such programs -- if you have a computer program you would like to make available to other users please fill out the attached questionnaire and return it to me.

ANNOUNCEMENT

Dr. Sue Hodge submitted the following announcement regarding the QUAD computer program. It is reprinted here as sent:

We have recently identified an error in an earlier version of our computer program QUAD (Hodge et al., 1983, p. 1141). **All the quadratic interpolations themselves, both one- and two-dimensional, are correct.** However, the program calculated the standard errors and the variance-covariance matrix incorrectly. Specifically, the standard error given by the incorrect program is too large by a factor of 1.5174 x $\sqrt{\ln 10}$; that is, the standard error must be multiplied by 0.6590 to be correct. The error is easily corrected by changing two lines in the FORTRAN code. (Write us for details.)

To determine whether your version of QUAD contains the error, simply enter the following lod scores for one-dimensional interpolation:

θ	.01	.05	.10	.20	.30	.40	.50
lod	1.06	1.29	1.26	1.10	.94	.63	0

The correct standard error for this test case is .0555. If your program gives a value of .0842, then the correction needs to be made.

We have notified those individuals and centers who requested our program. However, we are publishing this notice to ensure that everyone using the program has been alerted.

Reference
Hodge SE, Anderson CE, Neiswanger K, Sparkes RS, Rimoin DL: The search for heterogeneity in insulin-dependent diabetes mellitus (IDDM): Linkage studies, two-locus models, and genetic heterogeneity. Am J Hum Genet 35:1139-1155, 1983.

SOFTWARE NOTES

Version 4.6 of the **Linkage** programs has been made available. It represents a major improvement over the previous version (3.5). Included in the package is a command shell program (LCP) which is menu-driven and builds a batch file that can then be run to carry out the calculations defined in the LCP program. Another program included in the package is a report generator. The whole package may be obtained on 18 floppy disks from Dr. Mark Lathrop in Salt Lake City.

For PC/AT users who are interested in analyzing general pedigrees only, the corresponding version 4.6 Linkage programs are available from me on 4 disks (3 without source code) (please write to me for ordering instructions). Unfortunate-

ly, I have so far been unable to compile the new program version on either an IBM mainframe (Pascal/VS) or on the DSI-32 coprocessor board.

Using a pedigree with 91 individuals (no loops) and 3 codominant markers I happened to work on, I ran the new and old versions of Ilink on different machines. For version 4.6, the running times include processing of the input file by the 'Unknown' program. The Ilink program took 7 iterations and a total of 39 function evaluations. On the IBM mainframe, it finished sooner; the time shown below is extrapolated to the number of iterations taken in the other runs.

Program version	Machine	Running time
3.5	IBM AT 8MHz	71 min. 30 sec.
3.5	IBM AT 8MHz using 80287	58 min. 58 sec.
3.5	IBM 4381-11 mainframe	4 min. 12 sec.
4.6	IBM AT 8MHz	9 min. 53 sec.
4.6	IBM AT 8MHz using 80287	9 min. 27 sec.

Mark Lathrop and Peter Cartwright in Salt Lake City are planning on sending me some notes on the new Linkage version for inclusion in the next issue of the Newsletter.

I modernized **PC-LIPED** somewhat and included various example data files but did not change the basic algorithms. The new features are the following:

- Lognormal and straight-line age-dependant penetrance are implemented as special locus types.
- PC-LIPED does most calculations in double precision.
- Output is printed to the screen and optionally to a file so that the program may be interrupted with Ctrl-C without loss of the results obtained up to the break.
- PC-LIPED now comes in 2 versions, one of them does and the other does not use the numeric coprocessor. Previously, the Microsoft library used in the compilation allowed detecting presence of the coprocessor so that it would automatically be used when present. However, sensing presence of the coprocessor apparently fails in some IBM PC clones so that on those machines, PC-LIPED did not run.
- The program is compiled with Microsoft Fortran v.4.01. However, execution speed is about the same as when compiled with version 3.13 of MS-Fortran.

Programs for linkage analysis on microcomputers should be used with a **numeric coprocessor** whenever possible. Both

Turbo Pascal and MS Fortran allow a much larger range of real numbers when the coprocessor is used than when it is not. Without it, with larger family sizes, the value of the likelihood can sometimes become too small resulting in an underflow. While LIPED (MS Fortran) reports the occurrence of an underflow as an error, the Linkage programs (Turbo Pascal) do not and simply set the likelihood (of one or more genotypes) equal to zero. In unfortunate cases, this can lead to the situation that a family appears to contain an inconsistency while in fact it does not, or that it falsely appears to be uninformative for linkage. The Linkage programs as distributed do not make use of the coprocessor which is fine for most applications, but as a safeguard against nondetectable underflows, people who have a coprocessor installed in their microcomputer are encouraged to recompile the programs using the Turbo Pascal compiler with 8087 support.

My collaborators have recently started using the **DESQview** program version 2.0 (Quarterdeck Office Systems, Santa Monica). Although we presently have only 640KB of memory available on the Compaq Deskpro 386, the program behaves beautifully. For example, while one is working with WordPerfect on a manuscript, DESQview allows running in the background a long Mlink job on the DSI-32 coprocessor board and at the same time uploading a file to the mainframe computer. DESQview is inexpensive, we can buy it here in New York for $79.

QUESTIONS AND ANSWERS

Q: The application of results from linkage analyses to risk calculation requires the use of confidence limits for the recombination fraction θ that was estimated. Due to the lack of knowledge about the true distribution of lod scores, $z(\theta)$, no confidence limits for θ can be found in a strictly statistical sense.

One of the approximations proposed to avoid this problem was the use of so-called 'one-unit-down' confidence regions comprising all θ's for which $z(\theta)>z(\theta_{max})-1$.

I have calculated the significance of this method for the simplest case where the observations are distributed as binomial $B(n,\theta)$ variables, that is, recombinants and non-recombinants can definitely be distinguished. I performed these calculations using the normal approximation for the standardized number, S_0, of recombinants:

$$S_0 = (S-n\theta)/\sqrt{[n\theta(1-\theta)]} \approx N(0,1),$$

where S = number of recombinants. The results showed that

for n between 1000 and 2000,

$$c = \min_{\theta<0.5} P_\theta[z(\theta)>z(\theta max)-1] \approx 85\%. \quad (1)$$

If 1.75 instead of 1 is chosen as the 'max-lod' difference then c ≈ 95%. I know that in the case of real linkage data the graph of lod scores vs. recombination fraction does not have peaks as sharp as in the case of $B(n,\theta)$. Therefore, the 'confidence regions' obtained by the 'one-unit-down' method might be much larger. Can you give any hints about how to deal with this problem? Do you know whether and by whom it was treated mathematically? (Michael Krawczak, Institute of Human Genetics, Gosslerstr. 12 d, D-3400 Göttingen, FRG).

A: There is a well-known close relationship between significance tests and confidence intervals in the sense that a confidence interval comprises all those parameter values, θ_0, that do not make the test of $H_0:\theta=\theta_0$ significant. In the present example, a confidence interval has been defined as consisting of all those θ-values for which $z(\theta)>z(\theta_{max})-1$. Denote its endpoints by θ_1 and θ_2, $\theta_1<\theta<\theta_2$.

To avoid problems connected with approximating the binomial by the normal distribution, I calculated the values in the table on the next page using the binomial distribution, where k = number of recombinants, n = total number of recombinants and nonrecombinants, α_1 and α_2 are the "error probabilities" to the left and right of the confidence interval, and $c = 1 - \alpha_1 - \alpha_2$ is the confidence coefficient. The table suggests that the "1-unit-down" method results in reliable confidence intervals, at least for phase-known double backcross families. Analogous calculations for phase-unknown families with 2 offspring each confirm the general findings of the table. The only instance possibly leading to c<0.90 appears to occur with of a maximum lod score smaller than 1 and the restriction, $\theta_2 = 0.5$, but that case is specifically excluded from consideration.

Strictly speaking, the 'one-unit-down' method leads to a so-called support interval as it is based entirely on the log likelihood (= support) and not on any test theory. The notion of support intervals has been advanced by A.W.F. Edwards, "Likelihood", New York: Cambridge University Press, 1972. However, such support intervals may also be interpreted as confidence intervals (see above). The values for α_1 and α_2, above, were calculated based on J. Pfanzagl: "Allgemeine Methodenlehre der Statistik II", Sammlung Göschen, Walter de Gruyter, Berlin 1966, p.114. A more demanding treatment of confidence intervals may be found in statistics textbooks such as C.R. Rao: "Linear statistical inference and its applications", Wiley, New York. (JO)

k	n	θ1	θ2	α1	α2	c
0	10	0.000	0.206	0.000	0.100	0.900
1	10	0.004	0.403	0.040	0.044	0.915
0	20	0.000	0.109	0.000	0.100	0.900
1	20	0.002	0.222	0.039	0.044	0.916
2	20	0.014	0.298	0.031	0.036	0.933
5	20	0.088	0.484	0.026	0.029	0.945
0	50	0.000	0.045	0.000	0.100	0.900
1	50	0.001	0.094	0.039	0.044	0.917
2	50	0.006	0.128	0.031	0.036	0.933
5	50	0.033	0.214	0.025	0.029	0.947
10	50	0.099	0.337	0.022	0.025	0.953
0	100	0.000	0.023	0.000	0.100	0.900
1	100	0.000	0.048	0.039	0.044	0.917
2	100	0.003	0.066	0.030	0.036	0.934
5	100	0.016	0.111	0.024	0.029	0.947
10	100	0.048	0.176	0.022	0.025	0.953
0	500	0.000	0.005	0.000	0.100	0.900
1	500	0.000	0.010	0.049	0.043	0.908
2	500	0.001	0.013	0.037	0.036	0.927
5	500	0.003	0.023	0.026	0.028	0.946
10	500	0.009	0.037	0.022	0.025	0.954
50	500	0.074	0.131	0.018	0.020	0.962

Other questions will be discussed in the next Newsletter.

PERSONAL NOTES

Our New York community will soon lose two prestigious members: Early in January of 1988, Pat Jacobs and Newton Morton will leave us to continue their career in Southampton, England. We wish them well at their new place of work.

Jurg Ott
Columbia University, Box 58
722 West 168 Street
New York, NY 10032
Tel. (212) 960-2504

Q U E S T I O N N A I R E for computer programs

If you want to share/sell your program, please fill out this questionnaire and return it to the address at the bottom. For packages of programs, fill out one questionnaire for each program or, if package comprises many programs, one form only for the whole package.

Program name (and version):

Cost (and currency if other than US$):

Runs on which machines/operating systems:

Distribution medium (floppies, which type, how many, etc.):

Source code and/or run-time version?

If source code – what programming language?

Short description:

Limitations (eg, number of observations, no consanguinity):

Data format required (particularly for family data):

Your name and address:

Please return completed questionnaire to

Jurg Ott
Columbia University, Box 58
722 West 168 Street
New York, NY 10032 USA

Linkage Newsletter

Vol. 2 No. 1 July 1988

EDITORIAL

You may have been wondering what happened to the Linkage Newsletter. It is still very much alive although there has been an unplanned gap since the last issue, largely due to a lack of time on my part. But I hope that in the future, the Newsletter will appear on a more regular basis.

People interested in contributing news and comments are welcome to do so. Please submit your contributions on a diskette, either 5¼" (360KB or 1.2MB) or 3½" (DD or HD). The authors name will appear as given on the text submitted (initials or full name, with or without address).

ANNOUNCEMENT

The computer programs LODSTAT and SIMLINK (Boehnke, 1986; Boehnke and Ploughman, 1986) were developed to permit estimation of the power of a linkage study based on a set of pedigrees of known structure. The method used by these programs assumed that trait phenotype implies trait genotype and, hence, limited their usefulness to dominant, fully-penetrant traits. We have recently developed a new method (Ploughman and Boehnke, 1988) that allows estimation of the power of a linkage study for traits with more complex modes of inheritance. This method simulates genotypes at the trait locus, correctly taking into account trait information on all pedigree members. The method can be used for arbitrary pedigrees and a broad class of genetic models. These include models allowing incomplete or age-dependent penetrance, or a major locus and individual-specific environment. An updated version of the SIMLINK program employing our method is in preparation and will be made available this fall. Current users of SIMLINK and LODSTAT will be sent the updated version as soon as it is avail-

able. Other interested individuals may request the programs by writing:

Michael Boehnke, Department of Biostatistics
School of Public Health, University of Michigan
Ann Arbor, Michigan 48109

Boehnke M (1986) Estimating the power of a proposed linkage study: a practical computer simulation approach. Am J Hum Genet 39:513-527.

Boehnke M, Ploughman LM (1986) Estimation of the power of a proposed linkage study: computer programs for simulation. Am J Hum Genet 39:A148.

Ploughman LM, Boehnke M (1988) In preparation.

Submitted by Michael Boehnke and Lynn M. Ploughman
Department of Biostatistics, University of Michigan

COMMENTS FROM READERS

From Dr. A.W.F. Edwards, Department of Community Medicine, University of Cambridge, Fenner's, Gresham Road, Cambridge, CB1 2ES, England:

In QUESTIONS AND ANSWERS (Linkage Newsletter 1/2, Dec. 1987) it was stated that 'The application of results from linkage analysis to risk calculation requires the use of confidence limits for the recombination fraction that was estimated'. This is not so. Any consistent theory for the prediction of risk in terms of probabilities must invoke the likelihood principle (for which see my book Likelihood*, p. 30). Confidence limits are not appropriate; the correct procedure is to use the complete likelihood function and derive from it the corresponding likelihood function for the risk probability. A relevant genetical example is given on pp. 61-63 of Likelihood; an example involving the calculation of the risk of nuclear accidents may be found in Nature 324, 417-8 (1986). Of course, if a decision has to be made, decision theory tells us that there is then no alternative to assuming a Bayesian prior for the recombination fraction.

*) Ref. in Linkage Newsletter 1/2; also now in paperback (1984, 1987). ■

From Dr. Martin Farrall, M.D., London, England:

It is good practice to determine the "power" of a sample of families once thay have been collected (and before any typings are undertaken) in anticipation of a linkage study. Analytic methods are suitable for simple structures (see J. Ott, "The

Analysis of Human Genetic Linkage") and simulation methods may be used for more complex pedigrees (e.g. M. Boehnke's SIMLINK program).

Another simple method involves first calculating the maximum lod score for the family (usually with the aid of LIPED or LINKAGE) by 'making' up phenotypes that are fully informative and cosegregate with no cross-overs. This maximal lod score (calculated with $\theta = 0.0$) is used to estimate the equivalent number of phase-known meioses (John Edwards has provided suitable formulae in the past). It is then trivial to calculate the lod scores under imaginary conditions, e.g. the maximal lod score if two thirds of the meioses are informative and there's 10% recombination.

This should give a 'feel' for the "power" of the study which although rough and ready, is very easy to do. ■

A similar view has been expressed by Dr. Aravinda Chakravarti, Pittsburgh, who suggests first calculating the maximum lod score. If that is only equal to 1, for example, there might be little need to go into more detailed analyses.

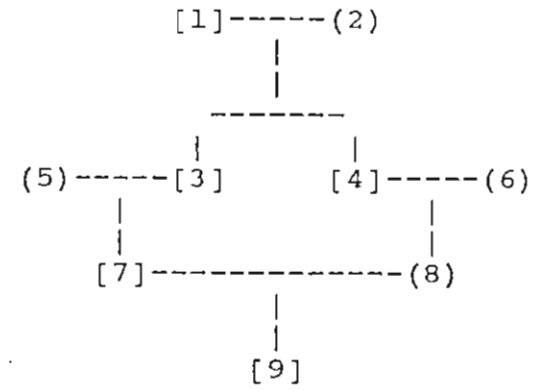
SOFTWARE NOTES

My Linkage Utility programs (version 2.1, list available on request) have been modernized and updated to run under Turbo Pascal version 4. Two bugs have been eliminated. One was in the CHIPROB program which calculates the p-value for a chi-square variable with given number of degrees of freedom; when chi-square was larger than 169, due to a programming error, the p-value returned may have been larger than 1. The other bug was in the ASSOC program which partitions the usual chi-square for the association between the phenotypes at two codominant loci into two portions, one due to allelic association and the other due to all other interactions; the algorithm used in the previous program version did not always produce reliable results, particularly for small numbers of observations. Users of the ASSOC program are encouraged to ask for the new version (please send a disk, 5¼" or 3½").

In the LINKAGE program package, several improvements have been made in the transition from version 4.6 to 4.7, but the small number of remaining bugs can still make life difficult for the unwary. Here is a list of the problems I know about.

- When you invoke the programs through LCP and want to calculate genetic risks, the risk locus must not be the last locus, otherwise the SETUP program aborts with an error message. You may, however, use any of the calculating programs (eg, MLINK) without problem.
- In MLINK, risks for homozygous normal and affected are

reversed on screen but correct in the output file (FINAL.OUT).
- In ILINK, the female/male distance ratio cannot be kept fixed unless it is equal to 1.
- When more than 9 loci are present, the PREPLINK program will write the 2-digit locus numbers without intervening spaces in the line for locus order. The analysis programs (MLINK etc.) are then unable to read the locus numbers properly and (may) report error messages difficult to interpret. The cure is to manually insert spaces in the output file produced by PREPLINK (the present version sent out by me does things right).
- The MAKEPED program cannot handle ID numbers larger than 700. It is generally preferable to use names rather than ID numbers in which case this problem is avoided.
- MAKEPED does not object when in a pedigree two unrelated sets of individuals exist (it does detect a single unrelated individual). MLINK, on the other hand, will use only one set of related individuals and completely disregards the others. Often, presumably, it is due to an error that sets of unrelated individuals occur in the same pedigree. Note that the LIPED program takes any number of unrelated sets of relatives in a single pedigree into account but it does not make the user aware of the presence of unrelated individuals.
- MAKEPED may not properly brake loops. The following example shows two bugs in the program.



When the individuals in the pedigree above are given in the order of their ID numbers and the MAKEPED program is asked to break the loop at individual no. 7, it produces the following pedigree file:

```
1 1 0 0 3 0 0 1 1  1 1 2  Ped: 1  Per: 1
1 2 0 0 3 0 0 2 0  1 2 2  Ped: 1  Per: 2
1 3 1 2 7 4 4 1 0  2 1 2  Ped: 1  Per: 3
1 4 1 2 8 0 0 1 0  1 2 2  Ped: 1  Per: 4
1 5 0 0 7 0 0 2 0  1 1 2  Ped: 1  Per: 5
1 6 0 0 8 0 0 2 0  1 1 1  Ped: 1  Per: 6
1 7 3 5 0 0 0 1 2  2 1 2  Ped: 1  Per: 7
110 0 0 9 0 0 1 2  2 1 2  Ped: 1  Per: 7
1 8 4 6 9 0 0 2 0  1 1 2  Ped: 1  Per: 8
```

```
1 9 7 8 0 0 0 1 0  1 1 2  Ped: 1  Per: 9
```

Individual 9 is now said to have mother "8", which is fine, and father "9" which is wrong - it should be "10", the 'double' of individual no. 7 without parents in the pedigree. In addition, the ID number of individual 10 is written without a space after the pedigree number which leads to the analysis program to read a pedigree number of 110. In contrast, the PEDPOINT program, version 3.5, processes the pedigree correctly:

```
1   5   3   4   0   0   0 1 2   2 1 2  id=   7
1   4   0   0   5   0   0 2 0   1 1 2  id=   5
1  10   9   8   0   0   0 1 0   1 1 2  id=   9
1   9   0   0  10   0   0 1 2   2 1 2  id=   7
1   8   6   7  10   0   0 2 0   1 1 2  id=   8
1   7   0   0   8   0   0 2 0   1 1 1  id=   6
1   6   1   2   8   0   0 1 0   1 2 2  id=   4
1   3   1   2   5   6   6 1 0   2 1 2  id=   3
1   2   0   0   3   0   0 2 0   1 2 2  id=   2
1   1   0   0   3   0   0 1 0   1 1 2  id=   1
```

USEFUL HINTS FOR LINKAGE ANALYSTS

The ILINK program in the LINKAGE package contains a constant, NBIT, which before compilation has to be set to the "number of bits of machine precision", that is, to the length of the mantissa of real (floating point) numbers. As distributed, NBIT has been set equal to 23 which in several compilers is the mantissa length (excluding the sign bit) for single precision variables. However, in Turbo Pascal, the mantissa is 39 bits long without 8087 support and 52 bits long with 8087 support. Thus, it will be beneficial to adjust NBIT accordingly. In the ILINK versions presently sent out by me, NBIT is a variable whose value is determined by a small routine, 'precision', inside the program (within the 'initilink' overlay procedure). That routine reads as follows:

```
procedure precision(VAR nbit:integer);
{ Determines number of bits of machine precision, ie, length of
mantissa.  Based on BYTE magazine, Feb. 1985, page 231 }
CONST
  one = 1.0;
  zero = 0.0;
  minusone = -1.0;
VAR
  radix,precision,width,wide : real;
  x,y,z : real;
BEGIN {precision}
  wide := one;
  REPEAT
```

```
    wide := wide + wide;
    x := wide + one;
    y := x - wide;
    z := y - one;
  UNTIL (minusone + ABS(z)) >= zero;
  y := one;
  REPEAT
    radix := wide + y;
    y := y + y;
    radix := radix - wide;
  UNTIL radix <> zero;
  precision := zero;
  width := one;
  REPEAT
    precision := precision + one;
    width := width * radix;
    y := width + one;
  UNTIL (y-width) <> one;
  nbit := round(precision);
END; {precision}
```

A bivariate table of lod scores for the male and female recombination fraction, m and f, is easy to obtain in LIPED. The structure of the LINKAGE programs, however, makes this more difficult as these programs work in terms of the male recombination fraction, m, and the ratio, r, of the female to male map distance (Haldane measure). The following conversion formula yields the female recombination value, f, from given m and r:

$$f = \tfrac{1}{2}[1-(1-2m)^r].$$

Conversely, for given male and female recombinations, m and f, the corresponding ratio of female to male map distance is obtained as

$$r = \log(1-2f)/\log(1-2m),$$

where log is the logarithm to any base. For example, the following table shows the r values required for each given pair of male and female recombination fractions:

	Female recombination fraction, f				
m	0.01	0.10	0.20	0.30	0.499
0.499	0.0032	0.036	0.082	0.147	1
0.30	0.022	0.244	0.557	1	6.782
0.20	0.040	0.437	1	1.794	12.17
0.10	0.091	1	2.289	4.106	27.85
0.01	1	11.04	25.3	45.4	307.6

QUESTIONS AND ANSWERS

Q: How many loci can be analyzed at once with ILINK on a 256K PC? How do you increase the number to 5 or more? (JS)

A: Parameters such as the max. number of loci, individuals, pedigrees etc., have to be set as constants at the beginning of the program, source code before compilation. The values of these constants depend on each other in the sense that one is able to run a larger number of loci with only a few pedigrees than with a large number of families/individuals. One will simply have to try out what works and what does not. Two hurdles will have to be taken: (1) Some constellations of constant declarations will make it impossible to even compile the program, usually because the arrays would occupy more than 64K of space. (2) After successful compilation, a set of data may not run because there is not enough memory (RAM) in the computer; this latter problem can often be overcome by reducing the number of indivuals per run. ■

Q: How do -2 Ln(L) and location score in LINKMAP relate to likelihood? (JS)

A: It is somewhat confusing that various different transformations of the likelihood are in use, and it is important to distinguish them carefully. The common denominator is the likelihood, L, which, formally being a probability, is a number between 0 and 1. Below, log denotes the logarithm to base 10, and Ln denotes the natural log (base e).

The MLINK program, when analyzing 2 loci, reports the lod score defined in the standard manner, $Z = \log[L(\theta)/L(\frac{1}{2})]$. With more than 2 loci, it reports a quantity called LOG LIKE DIFFERENCE which is equivalent to the former LOCATION SCORE in LINKMAP and is obtained by varying one of the recombination fractions, θ, in a given interval while holding the recombination fractions in the other intervals constant. Specifically, the LOG LIKE DIFFERENCE is the difference, D, obtained by subtracting $-2\ \mathrm{Ln}[L(\theta)]$ from $-2\ \mathrm{LN}[L(\frac{1}{2})]$. Note that this is <u>not</u> a lod score which for this situation may be defined as $Z = \log[L(\theta)] - \log[L(\frac{1}{2})]$ so that it represents one of the possible extensions to the multipoint case of the standard lod score. The two quantities, D and Z, are related by the formula $Z = D/4.6$. ■

Q: How do you account for interference in the calculations when using the LINKAGE programs? (JS)

A: Interference can only be allowed for when 3 loci are analyzed. There are two possible options, (1) one either invokes a mapping function that is one of the first procedures in the programs, or (2) interference is defined through 3 recombination fractions, that is, the 2 θs in the two intervals and the θ between the two flanking loci. One chooses between these options through a switch (1 or 2) on one of the last lines in the data-

file input file (see documentation) where a switch value of 0 specifies absence of interference. With more than 3 loci, absence of interference is assumed. It is, of course, always possible to convert resulting θ estimates to map distances using a mapping function such as Kosambi's that has interference built in, but the programs will carry out internal calculations (probabilities of haplotypes, joint recombination events) under absence of interference whenever the number of loci is larger than 3. ■

COMPARISONS AMONG SOME OF THE FASTER MS-DOS COMPUTERS

To compare execution speed between several of the newer microcomputers regarding linkage analyses, the example pedigree in Lathrop and Lalouel, Am J Hum Genet 42/3, 1988, p.502, was used for test runs. No special approximations or other measures to reduce execution time were taken. In individual 32, the third marker phenotype was changed form 11 to 12 (incompatibility). The MLINK program was used with the {R+} switch which allows range checking in arrays but adds 15-20% execution time (compiled with 8087 support). 4 loci total were used. All machines, below, had an 80287 coprocessor installed except for the Model 70 which had an 80387. The execution times given in the following table reflect the time required by MLINK per likelihood calculation (the UNKNOWN program was run prior to the test runs).

Microcomputer	Execution time	Relative speed
IBM PS/2 Model 70, 20 MHz	42 sec.	3.4
PC Designs GV386, 16 MHz	79 sec.	1.8
Compaq 386, 16 MHz	73 sec.	1.9
Compaq 286, 8 Mhz	141 sec.	1

On the IBM PS/2, the calculations were also carried out with MLINK compiled by Turbo Pascal version 4. They took 12 seconds only. However, under Turbo version 4, the programs sometimes yield wrong results whose causes are still unknown.

COMPUTER PROGRAMS

Several descriptions of computer programs were received in response to the questionnaire in the last Newsletter. Due to lack of time, a list of available programs will appear in the next issue of the Newsletter only.

Jurg Ott
Columbia University, Box 58
722 West 168 Street
New York, NY 10032

Tel. (212) 960-2504
FAX (212) 795-5886
EARN/Bitnet: OTT@NYSPI

Linkage Newsletter

Vol. 3 No. 1 May 1989

EDITORIAL

The present issue of the Linkage Newsletter mainly focuses on the LINKAGE computer programs, the major contribution being a report by Dr. Sandkuyl on a significant upgrade of these programs. The newest PC-version of the LINKAGE programs is 5.03; please notice that I no longer support any of the LINKAGE programs before version 5.03.

Also in this newsletter is a practical example on how to carry out linkage analyses with loci in the pseudoautosomal region.

On March 10-12, 1989, I held the first 3-day course on linkage analysis at Columbia University. While I had sent out notices to people in the U.S. and Canada, I felt it would be too expensive for people from abroad to attend the course. Nevertheless, we had several participants from Europe. As I do not (yet) have any grant support for conducting such courses, it was necessary to charge an admission fee. It is planned to hold at least one linkage course each year (3 days in March). There may be the possibility for two courses, a beginners' course in spring and an advanced course in the fall, but this year I may not find the time to organize another course.

Details and registration procedures for the next linkage course (in March of 1990) will be announced in a later issue of this newsletter.

LINKAGE programs version 5.03

(Lodewijk A. Sandkuyl

Recently, Borland International introduced version 5 of Turbo Pascal. As this version has significant advantages over version 3 which previously was used to compile the LINKAGE programs for the PC, we have adapted the LINKAGE programs version 4.8 to become version 5.03. Here, the major differences will be discussed in detail.

UNKNOWN program

Memory management was made more efficient: it is now virtually impossible to run out of memory, even with many large pedigrees.

An option, unknown to most users, to eliminate rare genotypes from subsequent calculations was deleted. Thus, the number of required calculations is reduced considerably, which speeds up performance. Under some circumstances the program is six times faster than the previous version. A procedure was build in that checks for impossible genotypes (e.g., a child has marker alleles 1 and 2 while both parents are homozygous for allele 1). When an impossible combination of genotypes is found, the program will show on screen in which pedigree and for which marker the problem was detected. Inconsistencies for disease loci will also be detected.

In addition, Jurg incorporated a large number of checks on the datafile and pedigree file in the UNKNOWN program. When one of these procedures detects an error, the program will give an error message to the user and then abort (when the program was called from a batch procedure made with LCP, that procedure will also abort).

MLINK, ILINK and LINKMAP programs

Loops: Three bugs were fixed that led to incorrect results of likelihood calculations for pedigrees with loops. These bugs have been present in all versions 4 that I have checked, but were absent in version 3.5. Depending on the order in which the two copies of the loop person (the person at which the loop is broken) occurred in the pedigree file, the population gene frequencies were sometimes taken into account -incorrectly- for the loop person. For some situa-

tions, this must have produced completely incorrect results. Especially users of the risk calculation option should check their old calculations for pedigrees with loops, and repeat them with the new version.

While fixing these bugs, I added an option to all three programs to allow for more than one loop in a pedigree. The maximum number of loops per pedigree is now determined by a constant called 'MAXLOOP' (together with all other constants in LINKAGEC.PAS). To accommodate more than one loop, break each loop by duplicating an individual in the known manner, and give the two duplicate individuals a 2 each in the proband field for the first loop, a 3 each for the second loop, etc. If a duplicate individual is to be the proband, this individual must correspond to the first loop to be broken, and the proband field for the two duplicates has to contain a 1 and a 2 (same rule as before for only a single loop).

The user should be aware of three problems:

- Computing times will increase considerably as more loops are added. Always try to break a loop at a person with known genotypes for all markers and homozygous for as many markers as possible. It is now possible to break a loop at a person without parents in the pedigree (previously not allowed). This may be relevant, for instance, when a woman has married the brother of her deceased first husband. This option gives more flexibility to choose the optimal place to break a loop.

- Since I don't have access to the source of the MAKEPED program, that program has not been modified to allow for more than one loop. If you want to use MAKEPED to produce the pedigree file, you should duplicate the loop persons before invoking MAKEPED. Then answer 'no' when the program asks whether the pedigree contains any loops. Change the proband field for the loop persons after MAKEPED is finished. As far as I have checked, MAKEPED will process a pedigree as usual, even when there appears to be no connection between certain sections of the pedigree (due to the breaking of multiple loops).

- When multiple loops have to be broken, it is conceivable that certain sections of the pedigree will have no apparent connection with the part in which the proband occurs. In previous versions of the programs these parts of the family pedigree were simply ignored. A constant called 'FINDLOST' has been added to LINKAGEC.PAS. When this constant is set to TRUE, apparently 'lost' persons will be detected and taken

into account. To maintain compatibility with previous versions, 'FINDLOST' is set to FALSE by default.

Allelic association (linkage disequilibrium): One of the problems affecting the likelihood calculations for loops also had considerable impact in cases of allelic association. We are convinced that this bug has now been fixed, and we urge everyone to repeat calculations with linkage disequilibrium using this new version (especially the risk calculations).

Memory management: Here, again, memory management was made more efficient. This means that larger problems can be analyzed on the PC. If the computer runs out of memory during calculations, some data that are not immediately needed will be written to disk (swapped out of memory). By default, a file called LINKSWAP.TMP is made in the current directory for this purpose (it is deleted when the program is finished). This file may become large, depending on the particular problem analyzed. As soon as the computer starts using this file, a message will be displayed on the screen. In some cases, the writing and reading of this file will take considerable extra time. In that case, the user can indicate that the 'swapfile' should be made on a virtual disk (extended memory that is used as a very fast disk) when available. This can be done by invoking MLINK (or one of the other programs) followed by a disk and filename, e.g., MLINK E:LINKSWAP.TMP. If LCP is used to make a batch procedure, it is still possible to specify the drive and filename for the swapfile by modifying the PEDIN.BAT file before calling it.

With these modifications, the size of the largest sibship will now determine how much (HEAP) memory is needed during execution. When you have a large sibship in your pedigree, the programs may still run into HEAP problems despite these modifications.

A new constant 'FEEDBACK' was added to the constants in LINKAGEC.PAS. If FEEDBACK is set to TRUE before compilation, the program will show on the screen which individuals it is processing as it traverses the pedigree. This may be useful when otherwise no output is expected for a longer time. One may then see that the program is actually at work rather than being stuck somewhere. Also, if the computer runs short in memory, one can see which parts of the pedigree could be processed without problems, and where the actual problems occurred. - All of these changes have been communicated to Dr. Lathrop, who will fix these problems in versions of the programs for computers other than PC's.

Compilation: Turbo Pascal 5 requires large programs to be divided into so-called units. We have made 5 separate units, some of which are common to all linkage programs. One common unit, LINKAGEC.PAS, contains all constants that one could previously find at the beginning of each program. Changes in the constants should always be made in this file (they will then affect all three programs).

Unit LINKAGE1.PAS contains all the procedures that read the pedigree and the datafile; it is used by all three programs.

Unit LINKAGE2.PAS contains procedures specific to MLINK and LINKMAP, while LINKAGE3.PAS contains the GEMINI section of the ILINK program.

LINKAGE4.PAS contains the procedures that carry out the likelihood calculations. The actual programs MLINK.PAS, ILINK.PAS and LINKMAP.PAS are now very small. You only need to modify those files when you want to change the HEAP size or the STACK size. Vital in the whole compilation is a file called SWITCHES.PAS.

To summarize:
- Make changes in constants in LINKAGEC.PAS
- Make changes in compiler options in SWITCHES.PAS,
- Make changes in HEAP/STACK size in main programs.

When memory problems occur: One can distinguish three kinds of memory problems: 'RANGE CHECK ERROR', 'HEAP OVERFLOW', and 'STACK OVERFLOW'.

'RANGE CHECK ERROR': indicates that one of the constants in LINKAGEC.PAS has to be increased. In general, try to set these constants to the smallest possible value when you experience memory problems. It will save memory, and execution will be faster.

'STACK OVERFLOW': try to increase stack size in the main program (MLINK.PAS, ILINK.PAS or LINKMAP.PAS). When problems still occur, try to set the constants to their smallest values, and remove the statements DEFINE APPROXIMATE and DEFINE RISKCALC from SWITCHES.PAS. In that case, risk calculation will not be possible. Try not to use double precision, instead increase SCALE and SCALEMULT to prevent underflow.

'HEAP OVERFLOW': try reducing the STACK size, and check whether HEAP size is set to its maximum in the main program. Check the constants: MAXPED can be important. Set FEEDBACK to TRUE to monitor the allocation of memory. Remove persons from the section of the pedigree where the problem occurs. Analyze each family separately.

When you detect new problems, please let us know. (LS) ■

FURTHER COMMENTS ON THE LINKAGE PROGRAMS

In addition to Dr. Sandkuyl's report, above, I would like to point out that the PREPLINK program has been made more user-friendly. It no longer crashes when an input error is made, for example, when accidentally a letter is entered instead of a number. Also, when you try saving a file by writing over an existing file, PREPLINK will ask you to confirm this.

The old version 3.5 of the LINKAGE programs has been converted to run under Turbo Pascal 5. With version 5.03 available, however, there appears little need for using version 3.5.

The MAKEPED program still has a few bugs. Dr. Weeks is working on the source code and should have an updated version ready in about two months from now. (JO) ■

Linkage between an X-linked and a pseudoautosomal locus (J. Ott)

Assume a locus 1, X-linked, and a locus 2, in the pseudoautosomal region, each with 2 alleles. To carry out a linkage analysis between these two loci using generally available computer programs, one may code the inheritance of the loci as previously described (Ott J, "Y-Linkage and Pseudoautosomal Linkage", Am J Hum Genet 38:891-897, 1986). For example, the following scheme may be used:

Locus 1 X-linked, 2 codominant alleles and 1 dummy allele

Alleles	Frequency
A	0.3
a	0.7
y	0.5

Genotypes	Penetrances for Female phenotypes				Male phenotypes		
	AA	Aa	aa	unknown	AA or A	aa or a	unknown
A/A	1	0	0	1	0	0	0
A/a	0	1	0	1	0	0	0
A/y	0	0	0	0	1	0	1
a/a	0	0	1	1	0	0	0
a/y	0	0	0	0	0	1	1
y/y	0	0	0	0	0	0	0
L.Phen.	2 1	2 2	2 3	2 4	2 5	2 6	2 7

In the LINKAGE programs, this locus is best coded as an "Affection Status" locus, and no other phenotypes should be used than those given in the line denoted by L.Phen., above. This scheme of using in the LINKAGE programs the general penetrance representation given above (LIPED format) has been mentioned in a previous Linkage Newsletter (August 1987, page 3). It works as follows: Each of the phenotypes in the above table is taken to represent a different liability class in the LINKAGE programs (7 classes total); as the penetrances in the LINKAGE programs are probabilities of being affected, each individual must therefore be taken to be affected, that is, have phenotype 2 followed by the number of the corresponding liability class.

Locus 2 pseudoautosomal, 2 codominant alleles, assumed to be in linkage equilibrium with locus 1.

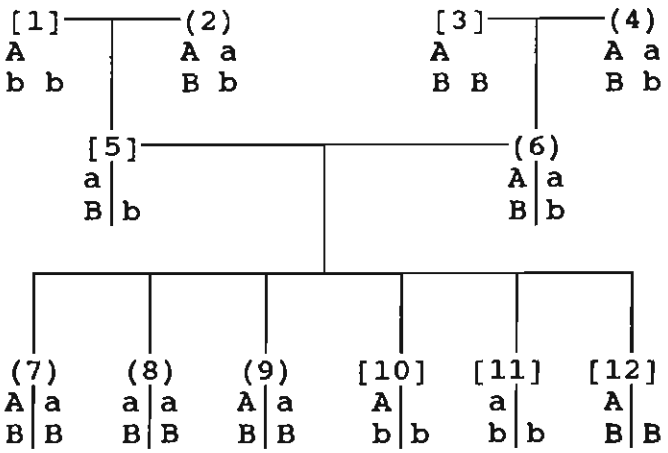
Alleles	Frequency
B	0.4
b	0.6

Genotype	Male or female phenotypes		
	BB	Bb	bb
B/B	1	0	0
B/b	0	1	0
b/b	0	0	1
L.Phen	1 1	1 2	2 2

In the LINKAGE programs, this locus may be used just like any autosomal codominant locus ("Numbered Alleles" type). Notice that the female recombination fraction measures a larger genetic distance than the male recombination, as pointed out in the paper referenced above.

Needless to say, one has to be extremely careful when using such a scheme for carrying out linkage analyses involving pseudoautosomal or Y-linked loci. The linkage programs have no way of detecting an error in coding, for example, when a male phenotype is used for a female phenotype or vice versa, but such errors easily lead to wrong results.

As an application of the method outlined above, consider the following example pedigree in which phenotypes were chosen such that they allow unequivocal determination of recombination events in the offspring.



The last two lines, above, indicate whether a recombination (R) or nonrecombination (N) has occurred in the father (m, male parent) or mother (f, female parent). The lod score is then given by

$$Z(r,s) = 12 \log(2) + \log(r) + 5 \log(1-r) + 2 \log(s) + 4 \log(1-s),$$

where r and s stand for male and female recombination fractions, respectively.

In the MLINK program of the LINKAGE package, the two input files then look as follows.

Pedigree file

```
1  1 0 0 1  2 5  2 2
1  2 0 0 2  2 2  1 2
1  3 0 0 1  2 5  1 1
1  4 0 0 2  2 2  1 2
1  5 1 2 1  2 6  1 2
1  6 3 4 2  2 2  1 2
1  7 5 6 2  2 2  1 1
1  8 5 6 2  2 3  1 1
1  9 5 6 2  2 2  1 1
1 10 5 6 1  2 5  2 2
1 11 5 6 1  2 6  2 2
1 12 5 6 1  2 5  1 1
```

Data file

```
  2 0 0 5  << NO. OF LOCI, RISK LOCUS, SEXL. (IF 1) PROGRAM
 0 0.0 0.0 0  << MUT LOCUS, MUT RATE, HAPLOT. FREQU. (IF 1)
```

```
  1  2
1   3  << AFFECTION, NO. OF ALLELES
 0.3  0.7  0.5   << GENE FREQUENCIES
 7 << NO. OF LIABILITY CLASSES
 1.0  0.0  0.0  0.0  0.0  0.0
 0.0  1.0  0.0  0.0  0.0  0.0
 0.0  0.0  0.0  1.0  0.0  0.0
 1.0  1.0  0.0  1.0  0.0  0.0
 0.0  0.0  1.0  0.0  0.0  0.0
 0.0  0.0  0.0  0.0  1.0  0.0
 0.0  0.0  1.0  0.0  1.0  0.0 << PENETRANCES
3   2  << ALLELE NUMBERS, NO. OF ALLELES
 0.6  0.4   << GENE FREQUENCIES
 0 0  << SEX DIFFERENCE, INTERFERENCE (IF 1 OR 2)
 0.0 << RECOMBINATION VALUES
 1  0.1  0.45 << REC VARIED, INCREMENT, FINISHING VALUES
(JO) ■
```

PERSONAL AND OTHER NEWS

A list of Bitnet addresses of researchers in population genetics/evolution is being maintained, updated and regularly sent by electronic mail to individuals on the list. For information, send a Bitnet message to Brian Golding, FS00047 @ YUSOL.

A reminder: The MENDEL and related programs for multipoint linkage analysis are now available without cost [Lange K, Weeks D, Boehnke M (1988) Programs for Pedigree Analysis: MENDEL, FISHER, and dGENE. Genet Epidemiol 5, 471-472].

I would like to welcome the following newcomers to New York:

Dr. Susan Hodge has joined the faculty in Biostatistics at Columbia University, and Dr. David Greenberg accepted a position with Mount Sinai Medical Center.

Dr. Daniel Weeks obtained his Ph.D. in 1988 from the Department of Biomathematics at UCLA and is now a Research Scientist with me at Psychiatric Institute and Columbia University.

Jurg Ott, Ph.D.
Columbia University, Box 58
Departments of Psychiatry and of Genetics & Development
722 West 168 Street
New York, NY 10032 Tel: 212-960-2504 FAX: 212-960-5604

Linkage Newsletter

Vol. 3 No. 2 December 1989

EDITORIAL

The description of the new PC version of the LINKAGE programs in the last Newsletter has apparently caused some confusion, and I would like to set the record straight. The changes made by Dr. Sandkuyl were soon also implemented by Dr. Lathrop and, due to an unfortunate lack of coordination between Paris and New York, we independently started mailing out upgraded program versions. This is to let users know that Dr. Lathrop's version 4.9 and our version 5.03 are essentially the same. We will jointly issue a new version, 5.1, which will be slightly different from the present versions.

I sometimes get letters from readers who are concerned that they might not have received all issues of the Newsletter. The proper explanation is usually that the Newsletter has not appeared as regularly as one might wish. The following volumes/issues have so far been sent out: 1/1, 1/2 (1987); 2/1 (1988); 3/1, 3/2 (1989).

LINKAGE COURSES

The second linkage course at Columbia University is a *Beginner's Course* and will be held *March 21-24, 1990* (Wednesday through Saturday). Registration is now open until the course is full. An application form is attached to this newsletter. *Please pass on copies of the attached announcement to anyone interested.* In the fall of 1990, an advanced course will be offered (3 days) but the date has not been set yet.

COMPARISON OF COMPUTERS FOR LINKAGE ANALYSES

While various benchmark programs are used to compare computers, the most useful criterion for a linkage analyst is how quickly a computer can run a typical linkage analysis. Dan Weeks and I have, therefore, used the same benchmark problem as the one discussed in the 2/1 Newsletter and had MLINK calculate 10 likelihoods for the first 4 loci in the family published by Lathrop and Lalouel, Am J Hum Genet 42/3, 1988, p. 502 (in individual 32, the third marker phenotype was changed from 11 to 12).

For the present runs, the program constant maxcensor was set equal to 30,000 which generally improved computing speed by approximately 20% as compared with maxcensor=1000. The times indicated in the table, below, are the number of seconds required to calculate the ten likelihoods. On multiuser systems, the CPU time is given rather than elapsed time, which gives these systems an advantage of perhaps 10% over single user systems. The computers are identified by manufacturer, model, processor (clock speed in MHz given after hyphen), numeric coprocessor if present, operating system in parentheses, and compiler. On the machines running under MS-DOS/PC-DOS ("DOS"), the compiler used was Turbo Pascal version 5.0.

The times shown below not only measure machine performance but also, for example, some aspects of the efficiency of the compiler used. Other compiler characteristics are not given. For example, the Pascal compiler on the Silicon Graphics machine took less than 30 seconds to compile the MLINK program yet carried out some checks not done by other compilers.

Computer	Seconds
DEC DECstation 3100 RISC (Unix)	17
Silicon Graphics Personal Iris 4D/25, RISC-20 (Unix)	20
Sun SPARCstation 1 (Unix)	37
DEC VAXstation 3100 (VMS)	62*
Proteus, 80386-25, 80387-25 (DOS 4.0)	96
Dell System 310, 80386-20, 80387-20 (DOS 4.0)	116
IBM 4381-11 mainframe (CMS), Pascal/VS	119
Compaq Deskpro, 80386-20, 80387-16 (DOS 3.3)	147
Mac SE/30, 68030-16, 68882, MPW Pascal	162
IBM PS/2 Model 70, 80386-20, 80387-20 (DOS 3.3)	178
DEC Microvax II	191
Compaq Deskpro, 80386-16, 80287 (DOS 3.3)	374
PC Designs, 80386-16, 80287-10 (DOS 3.3)	459
Toshiba 3100e, 80286-12, 80287-10 (DOS 3.3)	486
IBM AT, 80286-8, 80287-5 (DOS 3.3)	658
Compaq Deskpro, 80286-8, 80287-5 (DOS 3.3)	658

*Compiled with VMS Pascal with the /NOCHECK/OPTIMIZE=ALL option and linked with the /NOTRACEBACK option.

SOFTWARE NOTES

Dr. Kenneth Lange announced at the Baltimore meeting of the American Society of Human Genetics November 11-15 that the full source code of the *Programs for Pedigree Analysis (MENDEL, FISHER, SEARCH, and dGENE)* is now available effective immediately. Interested researchers should contact Dr. Lange (Tel. 213-825-5018) and will have to sign a user agreement.

Dr. Daniel Weeks is working on implementing the *LINKAGE programs on the Macintosh.* Test versions of the programs run fine; see computer comparison chart, above. Generally available versions should be ready in a few months. These programs will run best on a Macintosh with a math coprocessor. The Macintosh programs will have an advantage over the MS-DOS programs since they can use all available memory.

I started adapting the *LINKAGE programs to OS/2* using the Prospero OS/2 Pascal compiler. Potentially, using the linkage programs under OS/2 could be an attractive solution (low cost, multitasking) for researchers who need to run more than the relatively small maximum number of loci under MS-DOS but do not want to purchase a workstation running Unix or VMS. OS/2 has virtually the same command set as DOS, so a minimal adaptation is required when changing from DOS to OS/2.

To compile the LINKAGE programs with Prospero Pascal requires splitting them into separately compilable so-called segments. A test version of the MLINK program runs fine, but the following compiler error was encountered. The R compiler switch controls whether the floating-point coprocessor will be used (R1) or not (R0). With R0, compiling is no problem but with R1, the Likelihood procedure cannot be compiled due to an 'internal compiler error'. The Prospero company cannot duplicate the error on an IBM clone, but the error definitely occurs under IBM OS/2 on an IBM PS/2 model 70 with 80387 coprocessor. Progress in this matter will be reported in the next Newsletter. Prospero OS/2 Pascal might be quite useful as its present implementation allows dynamically addressing up to 4MB of memory compared to approximately ½MB with Turbo Pascal under DOS.

The standard Pascal version of the LINKAGE programs is available from Dr. Mark Lathrop in Paris. It may easily be adapted to Pascal/VS on *IBM mainframes running under VM/CMS.* The following considerations apply.

- All Pascal file names, eg, DATAFILE, will be connected to disk files with the same name preceded by 'file', eg, FILE DATAFILE.

- To assign the standard output file to the terminal, the statement TERMOUT(OUTPUT) should be added at the beginning of the main program body.

- In Pascal/VS, real variables cannot be smaller than 10**(-78) in absolute value, or else an underflow occurs which will be indicated on the screen. If this happens, the results calculated after the underflow occurred are not reliable and should not be used. The possibility of underflows occurring limits the size of the problem you may analyze.

- Each occurrence of the up arrow sign, ^, has to be replaced by the @ sign.

I am sometimes asked to send source code of the LINKAGE programs by *Bitnet*. While this might be an attractive way of distributing source code, it is presently too complicated for my collaborators in that it requires special attention to each such request. I am sorry that I cannot at this time provide this service, but it may become available at a later date. We will explore the possibility of using a server so that interested researchers can request programs or other text by Bitnet without us having to upload and send them.

Jurg Ott, Ph.D.
Columbia University, Box 58
Departments of Psychiatry and of Genetics & Development
722 West 168 Street
New York, NY 10032

Tel: 212-960-2504

FAX: +1-212-568-2750 or 212-960-5604

Jurg Ott
Columbia University, Box 58
722 West 168 Street
New York, NY 10032
Tel. (212) 960-2504

December 1989
FAX (212) 568-2750
or (212) 960-5624
Bitnet: OTT@NYSPI

Beginners' Linkage Course

Wednesday through Saturday, March 21-24, 1990

Course Description and Prerequisites

This course is for researchers with little or no experience in using linkage programs. A basic knowledge of linkage analysis is, however, required. *Topics:* Introduction to linkage analysis; practical aspects of data collection; strategies and methods of linkage analysis; reduced penetrance; application of computer programs (LINKAGE version 5.1; Programs for Pedigree Analysis [MENDEL]; Linkage Utility Programs); affecteds-only analyses; inbreeding loops; simple risk calculations. Computer programs can be taken home (on 3½" or 5¼" diskettes). A copyrighted course manual will be mailed for study before the course begins. Please distribute copies of this course announcement to any interested persons.

The course will be held in the microcomputer classroom of the Health Sciences Library (Columbia-Presbyterian Medical Center, 701 West 168th Street) which is equipped with 20 IBM PS/2 machines. The number of participants is *limited to 25.* People should plan on arriving in New York in the evening of Tuesday, March 20, 1990. For telephone information, please call Katherine Montague at (212) 960-2504 Mon/Tue/Thu or Fri, between 8:30 am and 12:30 pm New York time.

An advanced course will be held in the autumn of 1990 (date not yet fixed). The topics will comprise linkage with pseudoautosomal loci; estimating and testing for heterogeneity; linkage disequilibrium; risk calculations, also under linkage disequilibrium and locus heterogeneity.

Course Fee

The fee for the 4-day-course is *$400* for researchers at an academic institution, and *$500* for individuals from private (for profit) companies. It may be paid by check made payable to Columbia University Dept. of Psychiatry, or by Government pay order, but send no money now -- applicants will receive a bill and information regarding cancellation policy. As there is presently no support for this course from sources other than the course fee (a grant application is under review), no reduction of the cost to applicants is possible. This fee covers tuition and course related expenses (handouts, diskettes, etc.) but not room, board or meals. Course participants will receive a list of good and moderately priced hotels in New York and will have to make their own arrangements (except foreign participants). A small number of guest rooms are available in Bard Hall next to the Health Sciences Library (double rooms at $60 per room per night, single rooms at $55).

Application for Linkage Course by Dr. Jurg Ott

Please fill out this page and send it by mail or, preferably, by FAX. People interested in staying in a double room at Bard Hall should indicate a preferred roommate, or else we will match applicants.

Your name: _____

Affiliation: _____

Address: _____

_____

_____

Tel. number: _____

FAX number: _____ Bitnet: _____

Interested in staying at Bard Hall? YES / NO DOUBLE / SINGLE

If yes, which nights? Tue Wed Thu Fri Sat Sun

What size diskettes do you use? 3½" 3½"HD 5¼"HD (please circle)
(we prefer 3½"; we have no easy way of writing to low-density 5¼" disks)

For applicants from abroad: Do you want us to make hotel reservations? What accommodations?

Below, please describe which linkage programs you have used if any, how many families you have analyzed, and other experience in linkage analysis you might have:

Signature: _____ Date: _____

Linkage Newsletter

Vol. 4 No. 1 August 1990

LINKAGE COURSES

The third linkage course at Columbia University is an *Advanced Course* and will be held *November 8-10, 1990* (Thursday through Saturday). Registration is now open until the course is full. An application form is attached to this newsletter. *Please pass on copies of the attached announcement to anyone interested.* Unfortunately, there is presently no possibility for travel stipends or for reduced admission fees.

BUG REPORT

I am grateful to Drs. Bertram Müller and Tiemo Grimm (Würzburg, F.R. Germany) for making me aware of the following bugs in the LINKAGE programs. The bugs have an effect only when one works with different mutation rates in the two sexes. In PREPLINK, the order of male and female mutation rates is reversed, and in the LINKAGE analysis programs, the likelihood is calculated incorrectly when male and female mutation rates are different. In the versions currently being mailed out, these errors have provisionally been corrected by Dr. L. Sandkuyl and Joseph Terwilliger.

SOFTWARE NOTES

Version 5.1 of the LINKAGE programs is still in preparation. It should be available in September or October. A test version of it runs about 10% faster than version 5.03, and the release version should be somewhat faster still. Version 5.1 will also be available under OS/2 (see below).

A serious **potential pitfall** in the LINKAGE programs was recently uncovered by one of my collaborators, Dr. Chantal Mérette. She analyzed a pedigree for linkage between a dominant disease and a marker with 8 alleles. The data clearly showed a recombination, but the MLINK program (version 5.04) did not report a lod score of minus infinity at $\theta=0$. The explanation was as follows. The marker was coded as an *allele numbers* locus type with 8 alleles. In LINKAGEC.PAS, there is a program constant, MAXFACT, which gives the maximum number of

Support by the National Center for Human Genome Research for circulating this newsletter is gratefully acknowledged.

binary codes at a single locus. MAXFACT was left at its former value of 5 since no *binary factors* loci were to be analyzed. As it turns out, MAXFACT is also relevant for *allele numbers* loci and must be at least as large as the number of alleles, MAXALL. In future versions, a check will be incorporated to verify that this is indeed the case. For now, users should verify manually that MAXFACT≥MAXALL in LINKAGEC.PAS.

Conversion of the LINKAGE programs to the **Macintosh** (Dr. D. Weeks) is still in progress but has been relatively slow due to time constraints. These programs should become available towards the end of 1990. They will be announced in this newsletter. It should be noted, however, that presently there are no plans to also have LCP (Linkage Control Program) running on the Macintosh, since converting LCP by us would be too much of a time commitment.

Adapting the LINKAGE programs to **OS/2** is going smoothly. The problems with the Prospero Pascal compiler reported in the previous issue of this newsletter have been resolved. The current version of the compiler, version iio 5.202, is free from the bugs seen in previous versions. It is available from Prospero Software, 100 Commercial Street Suite 306, Portland, Maine 04101, and is highly recommended. The purchase price ($253 after educational discount, plus $5 handling charge) includes both the MS-DOS and OS/2 versions. It has several advantages over Turbo Pascal. For example, programs need not be broken into smaller units for compilation; file buffers can be set for any files, not just text files as in Turbo Pascal; up to 4 MB of memory can be addressed dynamically under OS/2 so that larger problems can be run (Turbo Pascal is not available under OS/2). ProPascal shares with Turbo Pascal the restriction that no Pascal procedure can be larger than 64KB in code (none of the procedures in the LINKAGE programs exceeds that limit). On the other hand, our test problem runs 34% faster when compiled with Turbo Pascal than with ProPascal (for DOS). We may be able to increase speed under ProPascal somewhat but probably not up to the level of Turbo Pascal. More details will be given in the next issue of the newsletter.

Several of the **HOMOG** programs have been updated. The HOMOG2 program has been changed to report likelihood ratios instead of p-values for the test of homogeneity versus heterogeneity. Since even the asymptotic distribution of the test statistic is unknown for the case addressed by the HOMOG2 program (when going from two to one component, this single restriction automatically eliminates two parameters), the formal p-values seem unreliable. Also, a specialized version of the HOMOG3 program has been added: HOMOG3R assumes that a locus is linked to either of two markers on different chromosomes, or is unlinked with the two markers. As usual, new program versions are obtained by requesting from us a list of programs with ordering instructions.

Jurg Ott, Ph.D.
Columbia University, Box 58
722 West 168 Street
New York, NY 10032 Tel: 212-960-2504 FAX: +1-212-568-2750

Support by the National Center for Human Genome Research for circulating this newsletter is gratefully acknowledged.

722 West 168 Street Tel. (212) 960-2507
New York, NY 10032 Bitnet: OTT@NYSPI

Advanced Linkage Course

Thursday through Saturday, November 8-10, 1990

Course Description and Prerequisites

This course is intended for researchers who already have experience with linkage analysis and would like to become more proficient in analyzing linkage data and handling unusual problems. A working knowledge of IBM PC's is assumed. The course will begin with a brief introduction to theory and techniques. Its main part will consist of problem-solving sessions with use of computer programs (LINKAGE version 5.04 or, hopefully, version 5.1; MENDEL; HOMOG; Linkage Utility Programs) and general question-and-answer sessions. The topics to be covered comprise heterogeneity, pseudoautosomal linkage, inbreeding loops, risk calculations, sex-specific recombination fractions, etc. Computer programs can be copied (on 3½" or 5¼" diskettes) and taken home. A manual will be mailed before the start of the course.

The course will be taught by myself and my collaborators. It will take place in the classroom of the Health Sciences Library (701 West 168th Street) which is equipped with 20 microcomputers of type IBM PS/2 (3½" diskette drives). Due to space limitations, course attendance is *limited to 30 participants*. Participants should plan on arriving in New York in the evening of Wednesday, November 7, 1990.

Course Fee

The fee for the 3-day-course is *$400* for researchers at an academic institution, and *$500* for individuals from private (for profit) companies. It may be paid by a check drawn on a U.S. bank, made payable to Columbia University Dept. of Psychiatry, by Government pay order, or by Travellers checks, but send no money now -- applicants will receive a bill and information regarding cancellation policy. As there is presently no support for this course from sources other than the course fee, no reduction of the cost to applicants is possible. This fee covers tuition and course related expenses (handouts, diskettes, etc.) but not room, board or meals. Course participants will receive a list of good and moderately priced hotels in New York and will have to make their own arrangements (except foreign participants). A small number of guest rooms are available in Bard Hall next to the Health Sciences Library (double rooms at $60 per room per night, single rooms at $55).

Application for Linkage Course by Dr. Jurg Ott

Please fill out this page and send it by mail or, preferably, by FAX. People interested in staying in a double room at Bard Hall should indicate a preferred roommate, or else we will match applicants.

Your name: _____

Affiliation: _____

Address: _____

_____

_____

Tel. number: _____

FAX number: _____ Bitnet:_____

Interested in staying at Bard Hall? YES / NO DOUBLE / SINGLE

If yes, which nights? Wed Thu Fri Sat

What size diskettes do you use? 3½" 3½"HD 5¼"HD (please circle)
(we prefer 3½"; we have no easy way of writing to low-density 5¼" disks)

For applicants from abroad: Do you want us to make hotel reservations? What accommodations?

Below, please describe which linkage programs you have used if any, how many families you have analyzed, and other experience in linkage analysis you might have:

Signature: _____ Date: _____

Linkage Newsletter

Vol. 5 No. 1 January 1991

Published by Jurg Ott, Columbia University, New York (support through grant HG00008 from the National Center for Human Genome Research is gratefully acknowledged)

Editorial Assistant: Katherine Montague
Tel. (212) 960-2507 Fax: +1-212-568-2750 Bitnet/EARN: OTT@NYSPI

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

1. EDITORIAL

This is the fifth year that the Linkage Newsletter is being published. It appears two to three times a year (but has occasionally also been issued only once) free of charge. This is the first time that it is also being distributed by e-mail over the *Bionews* electronic network (a brief description of *Bionews* is attached to the newsletter when mailed through the postal system). Individuals wishing to receive the newsletter by postal mail should write to us.

The purpose of this newsletter is to foster communication among researchers active in the analysis of human genetic linkage. The following three paragraphs are quotes from the first issue in 1987:

1. Basically, the newsletter should readily disseminate pieces of information relevant to genetic analysis that would not usually be published in scientific journals, where emphasis is on linkage analysis. This comprises, for example, proposals for new methods, notes on computer programs, computer hardware, meetings, experiences with analysis methods and — last but not necessarily least — gossip.

2. Readers are encouraged to submit articles or notes to be published in the newsletter, and I will make every effort at quickly including such contributions. I will not exercise any censorship, however advertisements for commercial products may have to be paid for (fee yet to be determined). When announcing computer programs you will have to specify at what cost they are available.

3. A Question and Answer section is planned in which I will discuss questions I have been asked and what I consider as meaningful answers. Again, contributions from readers to this section are welcome.

2. LINKAGE COURSES

The third linkage course at Columbia University (an advanced course) was held November 8-10, 1990. Of the 30 participants, 17 were from the United States and 13 came from other countries (Australia 1, Belgium 3, Denmark 1, England 3, Finland 2, France 1 Germany 2).

The fourth linkage course at Columbia University will be an *Introductory Course* held from May 15-18, 1991 (Wednesday through Saturday). *The previously scheduled dates of April 3-6 had to be changed because the microcomputer classroom was unavailable then.* Registration is now open and applications will be accepted until February 28, 1991. An application form is attached to this newsletter. *Please pass on copies of the attached announcement to anyone interested.* There is presently no possibility for travel stipends or for reduced admission fees. In the fall of 1991, the usual Advanced Linkage Course will be taught at Columbia University (date not yet set). The topics will include estimating and testing for heterogeneity; linkage disequilibrium; linkage with pseudoautosomal loci; risk calculations (also under linkage disequilibrium and locus heterogeneity).

Linkage courses will also be given in Europe on a more regular basis. The next European course (an introductory course) will be held July 9-12, 1991, at the Association Française contre les Myopathies (AFM) in Paris. Tuition is FR 6 000 and the language of instruction is English. Participants (maximum 24) must meet the following criteria: (i) be currently affiliated with an European laboratory, (ii) be familiar with a PC (i.e. word processing), and (iii) be able to exhibit an immediate need for linkage mapping. For further information and application forms, please contact:

Dr. Lynn Davis; AFM; 1 rue de l'Internationale
BP 59 - 91002 EVRY cedex, FRANCE;
telephone: (33)(1) 69 47 28 28, FAX: 60 77 12 16.

3. SOFTWARE NOTES

3.1 Bug in the SLINK program

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

University of Pittsburgh
Department of Human Genetics
130 DeSoto Street, A300 Crabtree Hall
Pittsburgh, PA 15261
Tel. (412) 624-5388 FAX: (412) 624-3020
WEEKS@PITTVMS.BITNET WEEKS@VMS.CIS.PITT.EDU

1/7/91

To: All users of the SLINK package
From: Daniel E. Weeks
Re: a bug in MSIM, ISIM, and LSIM

Instructions for correcting MSIM, ISIM, and LSIM:

We have discovered another bug in three programs of the SLINK package. In order to correct this bug, one line in the PROCEDURE readspseg has to be changed in each of the programs MSIM, ISIM, and LSIM:

IF j>0 THEN

has to be changed to

IF (j > 0) AND (i <= segperson) THEN

Using a text editor, find the appropriate line in each of your three source code files (as indicated below) and then change it. Then you must recompile each corrected program before you use it. If you have any problems, please feel free to contact me.

Old version:

```
**********************************************
 PROCEDURE readspseg;
 {Reads from the speedfile in appropriate segments}

  VAR
   i, j, a, b, sys: integer;
   ch           : char;

  BEGIN                                    {readspseg}
  { Note that each person[]^.unknown is set to FALSE as the
person is read in}
   i:=lastspeed;
   j:=lastspeed-lastseg;
   IF j>0 THEN
    BEGIN
     person[j]^.unknown:=TRUE;
      NEW(person[j]^.store);
    { person[j]^.store:=infoptr(NewPtr(SizeOf(information)));}
     WITH person[j]^.store^ DO
      FOR sys:=1 TO nlocus DO
       FOR a:=1 TO maxall DO
        FOR b:=1 TO maxall DO
         possible[sys, a, b]:=FALSE;
    END;
**********************************************
```

Corrected version:

```
**********************************************
 PROCEDURE readspseg;
 {Reads from the speedfile in appropriate segments}

  VAR
   i, j, a, b, sys: integer;
   ch: char;

 BEGIN                                    {readspseg}
  { Note that each person[]^.unknown is set to FALSE as the
```

```
person is read in)
  i := lastspeed;
  j := lastspeed - lastseg;
  IF (j > 0) AND (i <= segperson) THEN
   BEGIN
    person[j]^.unknown := TRUE;
     NEW(person[j]^.store);
   (  person[j]^.store := infoptr(NewPtr(SizeOf(information)));)
    WITH person[j]^.store^ DO
     FOR sys := 1 TO nlocus DO
      FOR a := 1 TO maxall DO
       FOR b := 1 TO maxall DO
       possible[sys, a, b] := FALSE;
   END;
************************************************
```

3.2 Computer programs for human genetic linkage analysis

There has not been much change regarding the LINKAGE programs since the last Newsletter, except that version 5.1 has just arrived from Mark Lathrop. We are in the process of carrying out a few tests and hope to have this latest version ready for distribution soon.

A particular version of the LINKAGE programs is now available, which allows for two loci to be jointly responsible for occurrence of a disease (DISK22 in list of programs). To receive a list of all our programs please write, fax, or send an e-mail message.

We are planning on making our programs available by e-mail. One possibility of doing this would be to install a file server on our IBM mainframe running under VM/CMS, which is a Bitnet node. However, we have no file server available and might have to write one. Does anyone have experience with such programs? The other possibility is to run a microcomputer under Kermit in server mode so that people can call up and download information and program files. The latter solution would be relatively easy to implement but requires people to pay telephone tolls.

Appendix to Linkage Newsletter of Jan. 1991, based on a document distributed through Bionews by David Kristofferson, PhD, GenBank On-line Service Manager, Sat Apr 28 1990

THE BIOSCI BULLETIN BOARD NETWORK

The BIOSCI bulletin board network was developed to allow easy worldwide communications between biological scientists who work on a variety of computer networks. By having distribution sites or "nodes" on each major network, BIOSCI allows its users to contact people on other networks 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 (see list below) and copies of the message will be distributed automatically to all other subscribers on all of the participating networks.

The following is a list of newsgroup topics available for distribution to sites on the ARPANET/Internet, BITNET, EARN, NETNORTH, HEANET, and JANET as well as UUCP sites. For each of these bboards, BITNET name abbreviations and analogous UNIX USENET newsgroup names are provided in a second list below. Also below is a list of the various sites (nodes) that distribute the bboards, the address format for posting messages, and the addresses used for subscription requests.

BBOARD NAME	TOPIC
AGEING	Scientific Interest Group
AGROFORESTRY	Scientific Interest Group
BIONAUTS	Address & other info about biologists on networks
BIONEWS	General announcements of widespread interest to biologists
BIOTECH	Biotechnology issues
BIO-CONVERSION	Scientific Interest Group
BIO-JOURNALS	Tables of Contents of biological journals
BIO-MATRIX	Applications of computers to biological databases
BIO-SOFTWARE	Information on software for the biological sciences
EMBL-DATABANK	Messages to and from the EMBL database staff
EMPLOYMENT	Job opportunities
GENBANK-BB	Messages to and from the GenBank database staff
GENOMIC-ORGANIZATION	Scientific Interest Group
HUMAN-GENOME-PROGRAM	NIH-sponsored newsgroup on human genome issues
METHODS-AND-REAGENTS	Requests for information and lab reagents
MOLECULAR-EVOLUTION	Scientific Interest Group
PIR	Messages to and from the PIR database staff
POPULATION-BIOLOGY	Scientific Interest Group
PROTEIN-ANALYSIS	Scientific Interest Group
RESEARCH-NEWS	Research news of interest to the community
SCIENCE-RESOURCES	Information from/about the funding agencies
SWISS-PROT	Messages to and from the SWISS-PROT database staff

The following list includes BITNET newsgroup names (<= 8 characters) and also the names of the corresponding UNIX USENET newsgroups.

Messages can be posted directly to any of these newsgroups without editorial intervention. Use the address format

bboardname@nodename

where "bboardname" comes from either the left or middle column below and "nodename" is found at the bottom of the respective column.

For example, the BIOSCI node at SERC Daresbury in the U.K. utilizes the BITNET names and is thus listed below the middle column in the following list. To post a message to the METHODS-AND-REAGENTS newsgroup, U.K. users would mail to methods@uk.ac.daresbury.

USENET users can also post news directly to the USENET groups listed below by using the "postnews" software on their local UNIX computer. Be sure to set the message distribution to "world." USENET groups are read using, e.g., the "readnews," "rn," or "vnews" software on UNIX systems. USENET news software is in the public domain and is available for most UNIX systems. A public domain USENET news software package named ANU-NEWS is also available for VAX/VMS systems. Your local BIOSCI node can point you towards acquiring the software for use on your computer system (BIOSCI addresses below).

BBOARD NAME	BITNET/EARN Name	USENET Newsgroup Name
AGEING	AGEING	bionet.molbio.ageing
AGROFORESTRY	AG-FORST	bionet.agroforestry
BIONAUTS	BIO-NAUT	bionet.users.addresses
BIONEWS	BIONEWS	bionet.general
BIOTECH	BIOTECH	bionet.technology.general
BIO-CONVERSION	BIO-CONV	bionet.technology.conversion
BIO-JOURNALS	BIO-JRNL	bionet.journals.contents
BIO-MATRIX	BIOMATRX	bionet.molbio.bio-matrix
BIO-SOFTWARE	BIO-SOFT	bionet.software
EMBL-DATABANK	EMBL-DB	bionet.molbio.embldatabank
EMPLOYMENT	BIOJOBS	bionet.jobs
GENBANK-BB	GENBANKB	bionet.molbio.genbank
GENOMIC-ORGANIZATION	GENE-ORG	bionet.molbio.gene-org
HUMAN-GENOME-PROGRAM	GNOME-PR	bionet.molbio.genome-program
METHODS-AND-REAGENTS	METHODS	bionet.molbio.methds-reagnts
MOLECULAR-EVOLUTION	MOL-EVOL	bionet.molbio.evolution
PIR	PIR-BB	bionet.molbio.pir
POPULATION-BIOLOGY	POP-BIO	bionet.population-bio
PROTEIN-ANALYSIS	PROTEINS	bionet.molbio.proteins
RESEARCH-NEWS	RESEARCH	bionet.molbio.news
SCIENCE-RESOURCES	SCI-RES	bionet.sci-resources
SWISS-PROT	SWISSPRT	bionet.molbio.swiss-prot

Node Addresses for left column	location	Node Addresses for middle column	location
bmc.uu.se	(Sweden)	irlearn.ucd.ie	(Ireland)
genbank.bio.net	(U.S.A.)	uk.ac.daresbury	(U.K.)
		genbank.bio.net	(U.S.A.)

Subscription Requests and other Information

There are FOUR main BIOSCI nodes. The BIOSCI nodes in Sweden, UK, and USA all have managers who can add new members to the Bboards, whereas the BIOSCI node in Ireland allows for automatic subscription at LISTSERV@IRLEARN (see detailed instructions on LISTSERV use below). To make the management/debugging of the BIOSCI network easier and to minimize network traffic, it would be most helpful if the following guidelines could be adhered to:

Your Geographical location	subscription address	BIOSCI node location
Scandinavia & Cont. Europe	biosci@bmc.uu.se	Sweden (Internet)
Ireland and Cont. Europe	LISTSERV@IRLEARN	Ireland (EARN/BITNET)
United Kingdom	biosci@uk.ac.daresbury	UK (JANET)
North and South America	biosci@genbank.bio.net	USA (Internet/BITNET)

NOTE: The procedure for automatic subscription through LISTSERV@IRLEARN (primarily for EARN users) is described further below. The BIOSCI manager at IRLEARN may be reached by mailing to BIOSCI@IRLEARN.

Canceling Subscriptions

If you have subscribed to a Bboard and are now leaving an institution or changing your e-mail address, then it is IMPERATIVE that you send a note to one of the addresses above and cancel your subscription. Non-existant addresses or overflowing mailboxes cause computer mail programs to send back "daemon" messages which bother everybody on the Bboard. We will immediately remove any address causing such a problem, but would prefer it if you would notify us in advance as a courtesy to the rest of the user community.

NOTE TO EARN (European Academic Research Network) USERS

Currently the LISTSERV service which allows automatic subscription is available from IRLEARN (NOT from any of the other BIOSCI nodes!). To subscribe to any of the Bboards at IRLEARN, send a message to LISTSERV@IRLEARN containing one of the following lines in the body of your mail message to select the Bboard of interest. If you want to belong to many different Bboards then send all the subscriptions at once in a BATCH mail file.

Replace your_personal_name in the list below with your own name. For example:

SUBSCRIBE +BIONEWS John A. Doe
SUBSCRIBE BIO+SOFT John A. Doe etc....

```
*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#
\              Mail the message to LISTSERV@IRLEARN             /
 *   PLEASE DO NOT SUBSCRIBE DIRECTLY TO THE NEWSGROUP ADDRESS   *
  #*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#
```

If you wish to cancel your subscription to a bulletin board, please send a similar message to LISTSERV@IRLEARN replacing SUBSCRIBE with the word SIGNOFF and omit your name, e.g., use simply

SIGNOFF +BIONEWS

It is also possible to cancel all LISTSERV subscriptions with a single command:

SIGNOFF * (NETWIDE

```
#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#
 * PLEASE DO NOT SEND YOUR SIGNOFF MESSAGE TO THE NEWSGROUP ITSELF *
  #*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#*#
```

Both subscription and cancellation requests are to be sent to LISTSERV@IRLEARN.

Bboard Name	Subscription/	Cancellation Command
AGEING	SUBSCRIBE +AGEING your_personal_name	SIGNOFF +AGEING
AGROFORESTRY	SUBSCRIBE AG+FORST your_personal_name	SIGNOFF AG+FORST
BIONAUTS	SUBSCRIBE BIO+NAUT your_personal_name	SIGNOFF BIO+NAUT
BIONEWS	SUBSCRIBE +BIONEWS your_personal_name	SIGNOFF +BIONEWS
BIOTECH	SUBSCRIBE +BIOTECH your_personal_name	SIGNOFF +BIOTECH
BIO-CONVERSION	SUBSCRIBE BIO+CONV your_personal_name	SIGNOFF BIO+CONV
BIO-JOURNALS	SUBSCRIBE BIO+JRNL your_personal_name	SIGNOFF BIO+JRNL
BIO-MATRIX	SUBSCRIBE +BIOMATR your_personal_name	SIGNOFF +BIOMATR
BIO-SOFTWARE	SUBSCRIBE BIO+SOFT your_personal_name	SIGNOFF BIO+SOFT
EMBL-DATABANK	SUBSCRIBE EMBL+DB your_personal_name	SIGNOFF EMBL+DB
EMPLOYMENT	SUBSCRIBE +BIOJOBS your_personal_name	SIGNOFF +BIOJOBS
GENBANK-BB	SUBSCRIBE +GENBANK your_personal_name	SIGNOFF +GENBANK
GENOMIC-ORGANIZATION	SUBSCRIBE GENE+ORG your_personal_name	SIGNOFF GENE+ORG
HUMAN-GENOME-PROGRAM	SUBSCRIBE GNOME+PR your_personal_name	SIGNOFF GNOME+PR
METHODS-AND-REAGENTS	SUBSCRIBE +METHODS your_personal_name	SIGNOFF +METHODS
MOLECULAR-EVOLUTION	SUBSCRIBE MOL+EVOL your_personal_name	SIGNOFF MOL+EVOL
PIR	SUBSCRIBE PIR+BB your_personal_name	SIGNOFF PIR+BB
POPULATION-BIOLOGY	SUBSCRIBE POP+BIO your_personal_name	SIGNOFF POP+BIO
PROTEIN-ANALYSIS	SUBSCRIBE +PROTEIN your_personal_name	SIGNOFF +PROTEIN
RESEARCH-NEWS	SUBSCRIBE +RESEARC your_personal_name	SIGNOFF +RESEARC
SCIENCE-RESOURCES	SUBSCRIBE SCI+RES your_personal_name	SIGNOFF SCI+RES
SWISS-PROT	SUBSCRIBE +SWISSPR your_personal_name	SIGNOFF +SWISSPR

Introductory Linkage Course

Wednesday through Saturday, May 15-18, 1991

Course Description and Prerequisites

This course is intended for researchers without or with only little prior knowledge of linkage programs. However, familiarity with the principles of linkage analysis and with the use of IBM PCs will be required. The course will begin with a brief introduction to theory and techniques. The main part will consist of problem-solving sessions with use of computer programs (LINKAGE, Linkage Utility Programs) and general question-and-answer sessions. The topics to be covered include strategies and methods of linkage analysis; reduced penetrance; application of computer programs (LINKAGE, Linkage Utility Programs); affecteds-only analyses; inbreeding loops; simple risk calculations. Computer programs can be taken home (on 3½" or 5¼" diskettes). A copyrighted course manual will be mailed for study before the course begins. Please distribute copies of this course announcement to anyone interested.

The course will be taught by myself and my collaborators. It will take place in the computer classroom of the Health Sciences Library (701 West 168th Street) which is equipped with 20 microcomputers of type IBM PS/2 (3½" diskette drives). Due to space limitations, course attendance is *limited to 30 participants*. Participants should plan on arriving in New York on the evening of Tuesday, May 14, 1991.

Course Fee

The fee for the 4-day course is *$500* for researchers at an academic institution, and *$600* for individuals from private (for profit) companies. It may be paid by check drawn on a U.S. bank made payable to Columbia University Dept. of Psychiatry, by Government pay order, or by Travellers checks, but send no money now — applicants will receive an invoice with detailed instructions. As there is presently no support for this course from sources other than the course fee, no reduction of the cost to applicants is possible. This fee covers tuition and course related expenses (handouts, diskettes, rental of classroom etc.) but not room, board or meals. Course participants will receive a list of good and moderately priced hotels in New York and will have to make their own arrangements (except foreign participants). A small number of guest rooms are available in Bard Hall next to the Health Sciences Library (double rooms at $65 per room per night, single rooms at $60).

Application for Linkage Course by Dr. Jurg Ott

Please fill out this page and send it by mail or, preferably, by FAX (see below). People interested in staying in a double room at Bard Hall should indicate a preferred roommate, or else we will match applicants. **PLEASE PRINT.**

Your name: _____

Affiliation: _____

Address: _____

_____

_____

Tel. number: _____

FAX number: _____ Bitnet: _____

Interested in staying at Bard Hall? YES / NO DOUBLE / SINGLE

If yes, which nights? Tue Wed Thu Fri Sat Sun

What size diskettes do you use? 3½" 3½"HD 5¼"HD (please circle)
(we prefer 3½"; we have no easy way of writing to low-density 5¼" disks)

For applicants from abroad: Do you want us to make hotel reservations? What accommodations?

Below, please describe which linkage programs you have used if any, how many families you have analyzed, and other experience in linkage analysis you might have:

Signature: _____

Date: _____

Katherine Montague, Course Coordinator
Tel. (212) 960-2507
FAX +1 (212) 568-2750 or 960-5624
Bitnet: OTT@NYSPI

Linkage Newsletter

Vol. 5 No. 2 May 1991

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

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 todays 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 2×2 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.

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 News letter, 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 micro-computer 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,l]);
```

(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, $L(\Theta, \Theta_2) / L(½, ½)$. 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

Linkage Newsletter

Vol. 6 No. 2 July 1992

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

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

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

EDITORIAL

This summer issue of the newsletter is rather short. It is again being distributed via e-mail (to the GENETIC-LINKAGE news-group of the BIOSCI network and to individual subscribers) as well as by postal mail. Hope everyone has a nice summer!

LINKAGE COURSES

The following linkage courses are scheduled:

Zurich (Advanced course): October 19-23, 1992, in the Computer Center at the University of Zürich, Irchel campus (see full-page course announcement attached).

New York (Advanced course): January 11-15, 1993, in the microcomputer classroom of the Health Sciences library, Columbia University. Course fee $100 (supported by National Center for Human Genome Research; only a small number of participants from outside the U.S. can be admitted). Registration is open; topics covered are as in the Zürich course. A formal course announcement will be made later.

In the spring of 1993, there will also be introductory courses in New York and Zürich but dates have not yet been fixed. For information and application forms for any course, please write (preferably by fax) to Katherine Montague, course coordinator, at the address given above.

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

SOFTWARE NOTES

Bug in Vax version of MLINK?

Joseph Terwilliger recently noticed the following discrepancy between the Vax version and the PC version of the MLINK program. Consider two parents and their child who are typed for two 2-allelic markers. Both parents are heterozygous, 1/2, at each locus and the child is homozygous, 1/1, at each locus. There is complete linkage disequilibrium, P(1-1) = P(2-2) = 0.5, and P(1-2) = P(2-1) = 0. Without allelic association, of course, this family is uninformative for linkage. Under complete disequilibrium, however, given the offspring's phenotype, parental phases are known and lead to two nonrecombinations. Thus, the lod score is given by $Z(\theta) = \log[4(1-\theta)^2]$, which has a maximum of 0.60 at $\theta=0$. This is also the result furnished by the PC version of MLINK, but the Vax version reports zero lod scores at any θ. We are investigating this matter and will report results as soon as possible.

Undeclared loops

In the LINKAGE programs, when loops (marriage or consanguinity) are present in a pedigree, this must be specified in the pedigree file (the safest method is to use the MAKEPED program). If a loop remains undeclared, depending on the type of loop, the analysis programs may terminate with an error or, worse, they may appear to terminate normally but give incorrect results. Xiaoli Xie has now written a program to detect loops (Xie X, Ott J [1992] *Am J Hum Genet*, abstr, in press). It is best run after MAKEPED in which case it catches any loops the user failed to declare. This LOOP program is based on the depth-first search algorithm in graph theory and is freely available.

Simulating under heterogeneity in SLINK

Both the SIMLINK (M. Boehnke) and SLINK (D. Weeks) programs allow for simulating pedigree data under heterogeneity, that is, with a given proportion of families simulated without linkage between trait locus and marker loci. In the analysis, however, there is a major difference between the two programs: The SIMLINK program analyses the data under heterogeneity while the analysis programs of the SLINK package do not. Some users have previously been unaware of this and have analyzed the data under homogeneity even though they had been generated by SLINK under heterogeneity (the resulting expected lod scores are too small). The program ELODHET was developed to allow analysis under heterogeneity for data generated by the SLINK program. It is part of the current package of SLINK.

Change in the LINKLODS program

The LINKLODS program, which comes with the PC version of LINKAGE, calculates lod scores for individual families from the output of MLINK or LINKMAP. Dr. Chantal Mérette recently pointed out that occasionally the LINKLODS program gives incorrect "total" lod scores. This occurred when a large number of families was analyzed and the total log likelihood was smaller than the program constant *lowlod*, which previously was set to -500. This constant has now been changed to a value -10000 so that the error should no longer occur. The current program version is 1.70 (6 July 1992).

Linkage analysis with highly polymorphic markers

Large numbers of alleles can pose problems in linkage analysis. Various exact and approximate ways of overcoming these problems have been proposed. Before discussing an overview of these possibilities, I would like to ask the readers if anyone has experience with the URP program (Michael S. Braverman: "An algorithm to improve the computational efficiency of genetic linkage analysis," *Comp Biomed Res* **18**, 24-36, 1985). Please let me know ─ I'm sure many researchers will be interested in this topic.

CORRECTIONS IN *ANALYSIS OF HUMAN GENETIC LINKAGE*

Below, the currently known corrections to this book (J. Ott, 1991, Johns Hopkins University Press, Baltimore) are listed.

Page 14, line 4 up: Assumption (2) is sufficient for that statement; (2) implies (1).

Page 18, line 8: Replace (1.3) by (1.2).

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

Page 44, line 8 below table 3.1 should read: "Generally, for phase known data, if T=k/n is the value of...". Also, line 12 should read: "Since T is unbiased, ..."

Page 47, lines 5-8: These two sentences are clearer when worded as follows: "Consider now our previous hypothetical example of one recombinant and four nonrecombinants and test H_0:θ=½ against H_1:θ=0.1. For these data, the likelihood ratio is calculated as $T_{obs} = [0.1\times(0.9)^4]/(0.5)^5$."

Page 48, line 16: Replace A≈(1-β) by $A\approx(1-\beta)/\alpha$.

Page 59, line 3 from the bottom: ..., P(0≤θ<½) = *1/22*, ...

Page 60, lines 6 and 7 should read: The *i*th segment (i = 1..s), of length b_i, then contains the likelihood ratio, $L^*(\theta_i)$, where $b_1 = ½(\theta_2+\theta_1)$, $b_i = ½(\theta_{i+1}+\theta_i)-½(\theta_i+\theta_{i-1}) = ½(\theta_{i+1}-\theta_{i-1})$, $b_s = 0.5 - ½(\theta_s+\theta_{s-1})$; $\Sigma b_i = 0.5$.

Line 17 should read: 52.672, resulting in a value of 0.71 for Smith's (1959) posterior...

Table 4.1: The values of b_i for i=1 (now 0.025) and i=2 (now 0.050) should be 0.030 and 0.045, respectively. This way, they are consistent with the definition of the b_i's further up on page 60.

Page 34, lines 17 and 18 up are clearer when formulated as follows: "... often used before linkage analysis as a preliminary test of paternity."

Page 45, lines 12 and 13 should be phrased more exactly as follows: "..., which allows the calculation of approximate confidence intervals from asymptotic variances... ."

Page 68, last line before section 4.5: Replace 11.7 by 11.6.

Page 74, line 3: Replace Z() and $Z(_f)$ by $Z_1(_m)$ and $Z_2(_f)$.

Page 75, line 5: Replace $(1-\alpha_1)^n$ by $(1-\alpha_1)^g$.

Page 92, line 3: Replace *A1* by *A2*.

Page 93, table 5.3, line i=4: Replace AB-22 by AB-11.

Page 101, after equation (5.15): Replace $1/[n\times i(r)]$ by $1/[n\times i(r)]^{½}$.

Page 101, line 6 in section 5.9 should read: "type 1 is a recombinant under one of the parental phases (phase I, say) but a nonrecombinant under the other, ..."

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 139, Table 6.10, line R: Replace "$444\theta_{BC}$" by "$444\theta_{AB}$".

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 the line above it).

Page 216, Problem 9.2, line 2: Replace "table 9.6" by "table 9.7".

Page 250: The last sentence of the top paragraph contains a typo: -2 should be 2, and Z(α,x) was not defined. For better clarity, the last two sentences in that paragraph should read: "In practice, this means that one evaluates Z(,x) at each map position, *x*, where Z(α,x) is analogous to (9.9) with θ_1 replaced by *x*, and is determined by the maximum of Z(α,x) at the given *x* value. Only those points *x* are then excluded for which Z(,x)<2 and Z(x)<-2, where Z(x) is the lod score under homogeneity."

Page 268, Solution 9.2, line 2: Replace "table 9.6" by "table 9.7".

Page 270, line 1: Replace _ by _. Line 3: Replace "with that mutation" by "without that mutation".

Page 279, ref. Hall et al. (1990): Replace "Anserson" by "Anderson".

Page 294, line 2 up should read: "...tetraploid..."

Page 302, Support interval: Replace 110 by 55.

Advanced Linkage Course

October 19-23, 1992 University of Zürich, Switzerland (Irchel Campus Computer Center, 12 IBM PS/2s), Proff. Eric Kubli (Zürich) and Jürg Ott (New York).

Tuition for the 5-day course is $600.
Maximum number of participants is 12.
Application deadline: August 25, 1992.

TOPICS include: Theory of linkage programs. Practical exercises using the LINKAGE and other programs. Handling of inbreeding loops, age-dependent penetrance, and sex-specific recombination fractions. Problems of interference in multipoint mapping. Models of genetic heterogeneity. Calculation of genetic risks, also under allelic and nonallelic association. Linkage analysis with quantitative trait loci, biological covariates, and pseudoautosomal loci. Computer simulation methods. Gene mapping in CEPH reference families.

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.

To obtain further information and an application form, contact:

Katherine Montague, Course coordinator, Columbia University, Unit 58, 722 West 168th Street, New York, NY 10032
Fax +1-212-568-2750 Tel. 212-960-2507

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 Θ_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) may be obtained from various anonymous ftp sites such as ftp.embl-heidelberg.de or ftp.bio.indiana.edu.

Linkage Newsletter

Vol. 7 No. 2 June 1993

Published by Jurg Ott, Columbia University, New York

Editorial Assistant: Katherine Montague
Fax: +1-212-568-2750
Tel. 212-960-2507 or 718-960-2507 (there have been some problems recently with our telephone and we are not quite sure which one works best right now)
e-mail: ott@nyspi.bitnet or jurg.ott@columbia.edu

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

EDITORIAL

Finally - we have made a major step into the electronic age. Up until now, our programs have been distributed only on diskettes. We have now set up an anonymous ftp site from which interested researchers can obtain programs and other materials (see below for details).

LINKAGE COURSES

There are still a few places available for the next Advanced Linkage Course. It will be held in Zürich (Switzerland), at the Computing Center of the University of Zürich Irchel campus, September 27 - October 1, 1993. Maximum number of participants is 12. 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.

MEETING ANNOUNCEMENT

(contributed by Dr. Robert Elston)

Second Annual meeting of the INTERNATIONAL GENETIC EPIDEMIOLOGY SOCIETY, October 10-11, 1993 - New Orleans, Louisiana. The meeting will be held in the Monteleone Hotel (French Quarter) and will begin on Sunday, Oct. 10th with a joint symposium with the American Society of Human Genetics on "Genetic & Environmental Aspects of Disease." For further information and abstract forms please contact Louisiana State University Medical Center Foundation, Division of Professional Education, 433 Bolivar Street, New Orleans, LA 70112 - Phone #(504) 568-3712/FAX #(504) 568-3460.

NEW ANONYMOUS ftp SITE

An anonymous ftp site (Login: anonymous; no password) is being made available for use by people interested in genetic linkage analysis. We are in the process of uploading all our programs to this ftp site. All programs distributed by us will eventually be made available on this site. The DOS version of LINKAGE version 5.20 is now also available from us (see below). For those people who are unable to access our ftp site, we will still distribute programs on floppy disks.

The ftp site presently works fine as far as we can tell. However, because of local network congestion in the afternoon, please access the site during off-peak hours as transfer may be impossible during periods of heavy network activity. To access our site, use the command

ftp york.ccc.columbia.edu or ftp 128.59.97.32

When asked for a login name, enter ANONYMOUS. A password is not required. For reports on problems please send e-mail to

joe@york.ccc.columbia.edu (Joe Terwilliger).

The host computer, YORK.CCC.COLUMBIA.EDU, is running Desktop-VMS Version 1.2, so the directory syntax must be specified in VMS format, which is different from UNIX-based ftp sites. When you reach our computer, you will be in directory *[anonymous]*, which is the root directory for the ftp site. All higher directories are off-limits, and any attempt to access such directories will block any further activity for you. That root directory contains a file, README.TXT, which contains a description of the directory structure and a list of all files available from the ftp site.

Each of the subdirectories in the ftp site contains a file with the directory name and extension .TXT. This file briefly describes the contents of the subdirectory. As an example, the directory [anonymous.pub.newsletter] contains the current and past issues of the *Linkage Newsletter*. Eventually, it will contain all newsletters published by us since the first issue of August 1987. Please note: We are still sending out the newsletter by e-mail. If you want to receive it by e-mail as soon as a new issue is ready, send a message to:
jurg.ott@columbia.edu

When you logon, you will notice that there are some files in the root directory. These are all account specific files, which are read/write protected. You must go to the subdirectory *PUB.DIR*, with VMS syntax *[anonymous.pub]*. You can get there by typing *CD [.pub]*, for example, if your computer is running UNIX ftp software. A directory name preceded by a "." means a subdirectory. If the "." is not given, the computer will think you are trying to access a higher-level directory, and will block further access to the system. If, for example, you are in the directory *[anonymous.pub.linkage.dos]*, and you want to move to the directory *[anonymous.pub.linkage]*, you could enter *CD [-]*, where the "-" sign means to go back one directory level. To move to the directory *[anonymous.pub.tlinkage.dos]*, you could enter *CD [--.tlinkage.dos]*, meaning back up two levels, and then move forward to subdirectory *[.tlinkage.dos]*. If you have any further questions about VMS directory syntax, or are having difficulty navigating around the ftp site, please send e-mail to Joseph Terwilliger (JOE@YORK.CCC.COLUMBIA.EDU).

To get any files other than ASCII files (*.EXE or *.W51 files, for example, are non-ASCII), please make sure you do this in binary mode. Typically, one should type *SET TYPE BINARY* at the ftp prompt, though the exact syntax may vary from system to system. If you further must use communications software like KERMIT to get the binary files from a UNIX or VMS machine to your PC, be sure to do those transfers in BINARY format as well.

SOFTWARE NEWS

PC version 5.2 of LINKAGE

The PC version of LINKAGE version 5.20 is now also available from us. It is essentially the same as the one you may obtain from Paris with the following exceptions: We have added a few more error checks, the program manual is included not for Postscript printers but as ASCII and WordPerfect files, and the COMPILE batch file of version 5.10 has been adapted to version 5.20. We are also in the process of adapting the Turbo Pascal version to Borland Pascal such that, for example, it can run under Windows and use more memory than under DOS. This adaptation, however, may take some time to become available.

The OS/2 version of LINKAGE version 5.2 is in preparation and should be available in August. It is being adapted to NDP Pascal for OS/2, which is just now becoming available from Microway, Inc. (tel. (508 746 7341, fax (508) 746 4678). The previous, pre-release version already had excellent qualitites (see Linkage Newsletter of February 1993).

Modified LINKAGE programs in C

Robert Cottingham at Baylor College of Medicine has translated some of the LINKAGE version 5.1 programs into C and rewritten several procedures resulting in a program version with much improved execution speed. A paper describing these modifications will appear in the *Amer J Hum Genet.* The programs may be obtained from the anonymous ftp site, gc.bcm.tmc.edu. Robert Cottingham may be reached at bwc@bcm.tmc.edu.

Estimating allele and haplotype frequencies - BEWARE!

As more and more people estimate allele and haplotype frequencies using the LINKAGE programs, in particular, the ILINK program, it is important to be aware of the following two steps required for reliable estimation:

1) The program constant, *fitmodel*, must be set to *true* and the program recompiled.

2) Whenever all individuals in a pedigree are untyped for a marker, the UNKNOWN program makes those individuals homozygous for the *1* allele at that marker. While this speeds up calculations considerably, it leads to wrong results in the estimation of allele and haplotype frequencies. A possible solution is simply not to use the UNKNOWN program. We have slightly amended the UNKNOWN program such that it contains a program constant, *makehomozygous*, which should be set to *false* if no *unknown* phenotypes should be turned into homozygotes *1 1*. A compiled version of UNKNOWN with makehomozygous set to false is furnished in executable form as UNKNOWN1.EXE.

FORTRAN programs

While most of our programs are written in Pascal, a few programs are in FORTRAN. This includes PC-LIPED and NOCOM (estimation of mixture of distributions). All our FORTRAN programs are now compiled with Microsoft FORTRAN 5.1 and bound in such a way that the same executable code runs under DOS or OS/2 (the LIPED version allowing for 12 alleles runs only under OS/2).

URP program for downcoding alleles

URP is a computer program which reduces the number of alleles at a marker locus by re-labeling ("down-coding") them. The original version of this program was designed and written by Michael Braverman in PL/I computer language in the early 1980's and was

implemented on an IBM Series/1 minicomputer in Dr. Kenneth K. Kidd's lab at Yale University. In 1984 Randy Holmes converted this program into primary PASCAL language, and in 1985 Matthew Hawley adapted it to a VMS/VAX machine. We recently obtained a copy of URP from Dr. Andrew Pakstis at Yale University, and Xiaoli Xie converted it to Turbo Pascal. Xiaoli Xie tried it out on nine CEPH families with chromosome 13 marker data; her experiences with the program are given below.

The input file to the URP program has a format similar to the one used in the LIPED program. However, genotypes can only be coded with alphabet letters, so there are up to 52 different allele codes as upper and lower case letters distinguishable. Besides, the URP program only recodes one marker at a time. Since genotypes in input files to the LINKAGE programs are coded with numbers, one must replace these numbers by letters. As the URP program only recodes one marker at a time, repeatedly modifying the input file for URP is inevitable. If one wants to use the output file resulting from URP as input to a linkage program other than LIPED, one must change the letters back to numbers. Since the whole procedure is very time consuming, we only tested the program with two-point analysis.

We used nine families in the pedigree file, each family with approximately 14 individuals. There were two markers with 12 and 9 alleles, respectively. All but seven individuals were untyped. We ran MLINK (PC version) and LIPED based on the original data-set, and then reduced the number of alleles by running the URP program, which reduced the number of alleles in the first marker from 12 to 4, and in the second one from 9 to 3. We assigned equal allele frequencies to the recoded makers and then ran MLINK and LIPED again. The results turned out to be identical. The elapsed time for the recoded dataset was much shorter than for the original data set, for example, 3 seconds versus 125 seconds. Note, however, that not knowing the allele frequencies can create problems for the analysis (J. Ott, Am J Hum Genet 51:283-290, 1992).

DISCLAIMER: The above investigation was carried out because we keep receiving requests regarding the availability of such programs. We do not endorse the URP program in any way. Its application may be problematic in several respects that we are not pursuing any further. For example, with untyped individuals present in a pedigree, exact solutions become extremely cumbersome (Ott [1979] *Cytogenet Cell Genet* 25:196). Also, the algorithm applied in the URP program focuses on twopoint analysis, but cases are known in which pairwise comparisons are uninformative yet the multipoint analysis is not. The most urgent need for downcoding alleles is in multipoint analysis, but it is not clear whether the procedure employed in URP consistently works fine for multipoint analysis. For the time being, we choose to carry out downcoding by hand even though it may be tedious.

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

Advanced Linkage Course

Sept. 27-Oct.1, 1993 University of Zürich, Switzerland (Irchel Campus Computer Center, 12 IBM PS/2s), Proff. Eric Kubli (Zürich) and Jürg Ott (New York). Tuition for the 5-day course is $600.
Maximum number of participants is 12.
Application deadline: July 26, 1993.

TOPICS include: Theory of linkage programs. Practical exercises using the LINKAGE and other programs. Handling of inbreeding loops, age-dependent penetrance, and sex-specific recombination fractions. Problems of interference in multipoint mapping. Models of genetic heterogeneity. Calculation of genetic risks, also under allelic and nonallelic association. Linkage analysis with quantitative trait loci, biological covariates, and pseudoautosomal loci. Computer simulation methods. Gene mapping in CEPH reference families.

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

(see reverse side for application form)

Katherine Montague, Course coordinator, Columbia University,
Unit 58, 722 West 168th Street, New York, NY 10032
Fax +1-212-568-2750 Tel. 212-960-2507

Application for Advanced Linkage Course by Dr. Jurg Ott - Zürich -

Please fill out this page and submit it by FAX (our preference) or by mail.

Your name: _____

Affiliation: _____

Address: _____

_____

_____

Tel. number: _____

FAX number: _____ E-mail: _____

On a separate sheet, please describe in detail what kind of linkage analyses you have done, including which programs you used, how you used them, how many families you have analyzed, etc.

Below, please describe your experience with microcomputers and DOS (Which programs used? Know how to copy files, make directories, etc?)

Signature: _____ Date: _____

Katherine Montague
Columbia University, Unit 58
722 West 168th Street, New York NY 10032
Fax (212) 568-2750 Tel. (212) 960-2507

Linkage Newsletter

Vol. 8 No. 1 January 1994

Published by Jurg Ott, Columbia University, New York

Editorial Assistant: Katherine Montague
Fax: 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

EDITORIAL

Our newsletter, now in its eighth year, is sent to many hundred recipients world-wide. More and more people receive it by e-mail, but we are still sending several hundred copies by postal mail. Please, if you receive this by postal mail, send us your e-mail address if possible - we would like to further reduce the labor intensive task of sending the newsletter by postal mail. If you are afraid of not receiving the newsletter by e-mail, we can send you a test message, or you may ask us to send you the newsletter twice, or to different e-mail addresses.

LINKAGE COURSES

The next two linkage courses are Introductory Courses. They will take place as follows:
June 6-10, 1994, at Columbia University, New York (maximum of 30 participants), and
June 13-17, 1994, at the University of Zurich, Irchel Campus Computer Center (Switzerland; maximum number of participants is 18).

These courses are for researchers with little or no experience in using linkage programs. A basic knowledge of linkage analysis is, however, required. *Topics:* Introduction to linkage analysis; practical aspects of data collection; strategies and methods of linkage analysis; incomplete penetrance; inbreeding loops; simple risk calculations; introduction to computer simulation. As usual, the main focus will be practical exercises and linkage analyses carried out by the 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 as the course textbook instead of mimeographed handouts (Terwilliger and Ott, "Handbook for Human Genetic Linkage," Johns Hopkins University Press, to become available in May 1994).

COURSE/MEETING ANNOUNCEMENTS

(Contributed by Dr. Pericak-Vance)

COURSE ANNOUNCEMENT
GENETIC ANALYSIS METHODS
FOR MEDICAL RESEARCHERS

Description: A four day intensive course on mapping human genetic diseases. The concentration is on the entire disease mapping process, including clinical classification, pedigree collection, molecular genetic analysis, statistical analysis, and gene characterization. The emphasis is on the global decision-making process, rather than details of specific techniques. Participants will be expected to discuss their own research projects.

Date and Location: May 15-18, 1994
Duke University
Durham, NC

Deadline for completed application: March 1, 1994.

For information contact: Dr. Margaret Pericak-Vance
Division of Neurology, Box 2900
Duke University Medical Center
Durham, NC 27710
or E-MAIL: genclass@genemap.mc.duke.edu

(Contributed by Dr. A.W. Eriksson)

INTERNATIONAL SYMPOSIUM
GENETIC EPIDEMIOLOGY OF TWINS & TWINNING

On April 22 and 23, 1994, the International Symposium 'Genetic Epidemiology of Twins & Twinning' will be organized at the Free University in Amsterdam, The Netherlands. At this symposium two main issues will be addressed. First, causes and consequences of the recent epidemic of multiple gestations as well as the possibilities to repel current high multiple birth rates. Second, twins as a methodological instrument in genetic epidemiology and behavior genetics.

For information contact: Bureau PAOG-Amsterdam
Tafelbergweg 25
1105 BC Amsterdam
The Netherlands
Tel: 31-20-566 4801 Fax: 31-20-696 3228

SOFTWARE NEWS

PC version 5.2 of LINKAGE

Two bugs were reported to me by Drs. Ken Morgan (McGill U.) and Marcella Devoto (Columbia U.). Both resulted in wrong lod scores in a LINKAGE program that seemed to run fine otherwise. The first bug was fixed by Dr. Mark Lathrop; the updated LINKMAP program is being mailed out since about November 1992. The second bug is caused by an overflow of an integer variable in Turbo Pascal, where integers are only 2 bytes long. Until a better solution can be found, the problem may be cured by the statement, TYPE integer=longint, at the beginning of the Turbo Pascal LINKAGE programs (this has been implemented in the currently mailed versions).

Turbo Pascal versions of SLINK and TLINKAGE

The Prospero Pascal compiler for a while seemed a good choice for PC's because it can generate code to run under DOS or OS/2. Because of availability problems and because it does not seem to do well with version 5.2 of LINKAGE, we are converting all our Prospero Pascal programs to Turbo Pascal (for DOS) and NDP Pascal (for OS/2). For a few weeks now, we have been mailing Turbo Pascal versions of SLINK and TLINKAGE. We just became aware of problems with these program versions. These programs are based on earlier LINKAGE versions, in which some dynamically allocated arrays had not been initialized. We have now taken corrective action - if you are experiencing problems with your Turbo Pascal version of SLINK or TLINKAGE, please let us know, we'll gladly send you an amended version.

SLINK will soon be available in a fast version. Drs. Alejandro Schaffer and Daniel Weeks have incorporated some of the features of FASTLINK in SLINK. More information will be available in the next Newsletter.

Turbo Pascal for OS/2?

An article in the German computer magazine -ct- reported on a rather easy patch to Turbo Pascal with the effect that code generated will run under OS/2. I do not yet know much more about this interesting possibility but hope to be able to provide detailed information in the next Newsletter. If anyone has experience with this, please let us know so all linkage analysts can profit from it.

Two potential problems with NDP Pascal

End-of-file character: In DOS, the character with ASCII no. 26 (Ctrl-Z) serves as an end-of-file mark. Our OS/2 NDP Pascal version, however, does not recognize this character and its function in DOS, rather it treats it just like any other character. This creates a problem in pedigree input files for which the LINKAGE programs apply an end-of-file test too see whether an input line is the last line in the file. If Ctrl-Z is present on the next line, the end-of-file test is "false" and the program attempts to read another input line (a pedigree number). It then issue the error message "invalid real" and aborts. Obviously, the problem may be overcome by removing the Ctrl-Z character before the file is used by the LINKAGE programs. Neither PREPLINK nor MAKEPED append a Ctrl-Z but, unfortunately, both editors supplied with OS/2 do. Thus, if you modified one of the input files for the LINKAGE programs with an OS/2 editor, you must remove the trailing Ctrl-Z character. This may be achieved by using an editor that does not append a Ctrl-Z, for example, DOS EDIT or the old WordPerfect Program Editor. If you use DOS EDIT, press Ctrl-End to move the cursor to the end of the file - the cursor should then be at the beginning of the line immediately following the last input line. If it does not, press the backspace key as often as necessary until the cursor is at the end of the last input line (on that line); then save the file.

Numbers with preceding 0's: Most Pascal programs read a number such as 023 simply as 23. Not so programs compiled with NDP Pascal. They treat numbers with preceding 0's as octal numbers. For example, 023 is interpreted as the decimal number 19. Numbers with preceding 0's that contain 9's are interpreted as 0. Consequently, if two pedigrees carry the ID numbers 08 and 09, they will in NDP Pascal programs wind up with the same ID number (that is, 0). We have thus added a small program to the MAKEPED batch file. It replaces the first 0 in a pedigree ID by P0. That way, pedigree ID's are treated by MAKEPED as characters rather than numbers, and the pedigree ID's output by MAKEPED are simple numbers.

LSP for the DOS version of CLINKAGE

As reported over a year ago, for the CEPH families version of the LINKAGE programs (eg. CILINK), the LSP program created a faulty datafile when the number of loci was larger than about 20. This problem has now been resolved by Ms. Xiaoli Xie - it was due to an inappropriate array bound in LSP.

LCP and LSP for OS/2

The OS/2 versions of LCP and LSP until recently ran only in full screen mode. If a user started one of these programs in a window the screen temporarily switched to full screen until the program finished. Ms. Xie has now made these programs window-compatible.

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

Linkage Newsletter

Vol. 8 No. 2 August 1994

Published by Jurg Ott, Columbia University, New York

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 erroneous likelihoods. The bugs and their suggested corrections are given below.

== 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.

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

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 -1000000000000000000.000000 -434293555638650388480.000000
---------------------------------
TOTALS    -1000000000000000000.000000 -434293555638650388480.000000
-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
```

```
       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 -10000000000000000000.000000 -43429355638650388480.000000
---------------------------------
TOTALS    -10000000000000000000.000000 -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
```

```
       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
```

Linkage Newsletter

Vol. 9 No. 1 February 1995

Published by Jurg Ott, Columbia University, New York

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

LINKAGE COURSES

In 1995, the following linkage courses will be held:

February 27 - March 3: Introductory course at the University of Zurich (Switzerland), Irchel Campus Computer Center. This course is full.

June 12-16: Introductory course at Columbia University, New York (maximum of 30 participants). We often receive many more applications than we can accept and will carry this course twice in case of high demand (date not yet set).

October 2-6: Advanced course at the University of Zurich (Switzerland), Irchel Campus Computer Center.

The next Advanced course at Columbia University will be held early in January 1996.

To obtain information on these courses, please write to Katherine Montague, course coordinator, by e-mail or fax. The textbook to be used for theoretical background and course exercises is Terwilliger JD & Ott J: *Handbook of Human Genetic Linkage*, Johns Hopkins University Press, 1994 (additional material for advanced courses will be handed out at the course). Course participants are expected to bring this book with them to the course; in case of problems please contact Katherine Montague.

UPGRADING OUR ftp SITE

Our anonymous ftp site (york.cpmc.columbia.edu) runs on an old Vaxstation 3100 with ftp software that is not well adapted to the system. Unfortunately, the Vaxstation occasionally freezes up and must be rebooted -- there is evidently little that can be done about this. So, if you experience problems accessing the ftp site, please try again later. Also, accessing it at night or in the morning tends to be more successful than at peak hours.

A new member of our team, Dr. Wentian Li, started installing a new anonymous ftp site on our Sparcstation IPC. It is expected to be much more stable then the current one. Furthermore, it runs under Unix rather than under VMS as on the Vaxstation. As soon as a larger hard disk is installed, we will transfer all files from the Vaxstation to the Sparcstation. When ready, the anonymous ftp site can be reached as **linkage.cpmc.columbia.edu** and the old site (york.cpmc.columbia.edu) will be taken out of service. The new ftp site is expected to be up by mid-March.

SOFTWARE NEWS

ESPA program

We regularly receive requests for the ESPA program, which carries out affected sib pair analysis with estimation of unobserved parental marker genotypes. This pro-gram is not available from us -- please address any requests to the developer of ESPA, Dr. Lodewijk Sandkuyil (sandkuyl@rullf2.LeidenUniv.nl), Voorstraat 27 a,
2611 JK Delft, The Netherlands. Tel: 011,31-15-123 638 Fax: 011,31-15-143-925.

Bug in LINKAGE regarding loops

Dr. Alejandro Schaffer submitted the following report:

> The purpose of this message is to report a dangerous bug in LINKAGE. The bug is that if maxloop is set to a number strictly lower than the actual number of loops, then LINKAGE gives no warning. In some cases it will give plausible but incorrect results. In contrast, in FASTLINK if maxloop is set to a number strictly lower than the actual number of loops, the program gives a warning and exits with instructions on how to fix the problem, without computing anything.

Editor's note: Readers of the Newsletter (January 1992) have been made aware of this problem. However, thus far no test had been implemented in LINKAGE to catch the situation that maxloop is smaller than the actual number of loops. This test has now been implemented in the DOS version of LINKAGE (available on our ftp site) and will be implemented shortly in other LINKAGE versions.

Newer versions of FASTLINK

The following announcement was submitted by Dr. Alejandro Schaffer:

Since May 1993, we have been distributing faster versions of the genetic linkage analysis programs in LINKAGE 5.1. Several users have dubbed the new code "FASTLINK".

Version 2.2 of FASTLINK is now ready and available. The changes from version 2.1 (distributed in March 1994) include

1. Bug fixes to bugs introduced in FASTLINK and bugs inherited from LINKAGE
2. More dynamic memory allocation
3. More diagnostics to detect user errors politely
4. Crash-recovery is now possible for LINKMAP and MLINK, in addition to LODSCORE and ILINK
5. Lots of information about portability of FASTLINK to different operating systems is included
6. A document entitled "The Mystery of the Unknown" explaining how the preprocessor program UNKNOWN works among other goodies. See the file README.updates, which comes with the distribution, for details.

Like FASTLINK 2.1, this version is being distributed from a computer at Rice University. Here are the instructions for retrieving the code:

ftp softlib.cs.rice.edu

Login as anonymous and leave your full e-mail address as password.

cd pub/fastlink

In that directory you will find various files. You can get everything at once by retrieving the file:

fastlink.tar.Z

and then (outside of ftp) doing the commands:

uncompress fastlink.tar.Z
tar xvf fastlink.tar

If you prefer to get your files piecemeal, instead of getting fastlink.tar.Z, start by getting README* The file README (with no extension) will give you a roadmap to all the documentation. I am maintaining a mailing list of FASTLINK users. If you have retrieved the code and would like to be on the mailing list, send me e-mail at the address below.

Special thanks to many FASTLINK users including: Lucien Bachner, Alan Cox, David Featherstone, Sandep Gupta, Victoria Haghighi, Carol Haynes, Jerry Halpern, Kimmo Kallio, Luc Krols, Shriram Krishnamurthi, Tara Cox Matise, Ken Morgan, Jurg Ott, Steve Roberts, Joe Terwilliger, Gerard Tromp, Ellen Wijsman, Xiaoli Xie, who provided bug reports, suggestions for improvements, guidance on documentation, and assistance with portability. I could not have prepared FASTLINK 2.2 without your help!

We are also distributing executable versions of FASTLINK for DOS. The ftp instructions are similar. Instead of doing
cd pub/fastlink

do

cd pub/fastlink/dos
binary

In that directory you will find 13 files. One is an executable for UNKNOWN (called unknown.exe). We have 3 versions each of LODSCORE, ILINK, LINKMAP, and MLINK with the constant maxhap set to 48, 96, 250 respectively. For example, the file li96.exe is LINKMAP with maxhap set to 96 and the file il250.exe is ILINK with maxhap set to 250. maxhap is the maximum product allowed for the number of alleles at each locus. For example, if you want to a 3-locus analysis with 2x4x8=64 alleles, you should not use the versions with maxhap set to 48, but you can use either the 96 or 250 versions.

Alejandro Schaffer, PhD
Department of Computer Science
Rice University
Houston, TX 77251
schaffer@cs.rice.edu
Phone: (713) 527-8101 x3813
FAX: (713) 285-5930

Compiling with the 'colig.bat' batch program

Dr. Joseph Terwilliger made me aware of a problem with the 'colig' batch program that we distribute to facilitate recompiling the LINKAGE programs for DOS (Turbo/Borland Pascal):

I have a question for you about the turbo pascal colig.bat file. Why does it have the line del *.tp* at the end? This seems to delete turbo.tpl, which is needed to compile the programs here on the PC's in Finland. Is it okay to change that line to DEL *.TPU instead of TP*, to preserve the *.TPL files?

Response: When the LINKAGE programs are compiled with Turbo/Borland Pascal, intermediate files are created and are left on the hard disk. These files are named *.tpx, where x=U for regular Turbo Pascal, x=P when compiled for protected mode, and x=W when compiled to run under Windows. The statement 'del *.tp*' deletes any of these files that might have been created. This procedure was chosen under the assumption that the compiler would reside in a different directory. If it resides in the same directory as the programs to be compiled, one of the compiler files, 'turbo.tpl', will also be deleted. Thus, the 'colig.bat' and 'colit.bat' batch programs have now been restructured so that 'turbo.tpl' will no longer be deleted.

Meeting Announcement

The following announcement has been submitted to the Newletter by the meeting organizers:

Fourth Annual meeting of the INTERNATIONAL GENETIC EPIDEMIOLOGY SOCIETY, June 20-22, 1995 - Snowbird, Utah. The meeting will be held in the Cliff Lodge (just 29 miles from Salt Lake City International Airport) in conjunction with the Society of Epidemiological Research. Abstracts for consideration of oral presentations and posters are due February 10, 1995. For abstract forms, contact Michele Brown, University of Utah, (801) 581-558099 or fax (801) 581-3165. For program information contact Melissa Austin, University of Washington, Seattle, (206) 685-9384, fax (206) 685-3407.

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

Linkage Newsletter

Vol. 9 No. 2 July 1995

Published by Jurg Ott, Columbia University, New York Postal address:
Email: jurg.ott@columbia.edu Columbia University, Unit 58
WWW: http://linkage.cpmc.columbia.edu 722 West 168 Street, New York, NY 10032

Editorial Assistant: Katherine Montague
Email: km165@columbia.edu
Fax: 212-568-2750
Tel: 212-960-2507

World Wide Web server

Dr. Wentian Li has installed a WWW server for our group. It may be reached at http://linkage.cpmc.columbia.edu. Our anonymous ftp site (linkage.cpmc.columbia.edu) is now fully functional and readers are encouraged to download programs and other files.

LINKAGE COURSES

The next two linkage courses are advanced courses and will take place as follows:

October 2-6, 1995, at the University of Zurich, Irchel Campus Computer Center, Switzerland (maximum number of participants is 14).
January 8-12, 1996, at Columbia University, New York (maximum of 20 participants).

To obtain information on these courses, please write to Katherine Montague, course coordinator, by email 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. Course work will be divided between theoretical introductions to topics, practical exercises carried out on IBM PC's using the LINKAGE and other programs, and informal discussions. We will use our book (Terwilliger and Ott, *Handbook of Human Genetic Linkage,* Johns Hopkins University Press, 1994) and some handouts for topics not covered in the book. 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. Participants are encouraged to bring their own data.

The next introductory courses will be held in the spring or early summer of 1996, one or two at Columbia University New York and one at the University of Zurich (dates not yet set).

SOFTWARE NEWS

PC versions of LINKAGE

In an earlier issue of this newsletter, we reported problems with version 5.2 of LINKAGE for DOS. This is to point out to users that similar problems may also occur on other platforms. For example, on our Sparcstation IPC, when version 5.2 is compiled without optimization, it yields the same results as version 5.1 and FastLINK. However, it tends to yield different results when compiled with optimization. Since a completely new version of LINKAGE (written in C) will be available soon, there would be no point in trying to fix these problems.

Haplotyping programs

The following announcement was contributed by Dr. Daniel Weeks:

We are pleased to announce that our haplotyping programs SIMCROSS and SIMWALK are now available in the pub directory by anonymous ftp to watson.hgen.pitt.edu (url: ftp://watson.hgen.pitt.edu/pub). These are programs for generating optimal haplotype configurations on general pedigrees using a likelihood-based approach to correctly take intermarker recombination fractions into account.

These two programs are described briefly in our Letter to the Editor which just appeared in the June issue of the American Journal of Human Genetics:

Weeks DE, Sobel E, O'Connell JR, Lange K (1995) Computer programs for multilocus haplotyping of general pedigrees. American Journal of Human Genetics 56:1506-1507.

and are described in depth in the forthcoming article:

Sobel E, Lange K, O'Connell JR, Weeks DE (1995) Haplotyping algorithms. In: Speed TP, Waterman MS (eds) Genetic mapping and DNA sequencing. Springer-Verlag, New York, in press.

If you have any questions or comments regarding these two programs, please feel free to contact me.

-- Dan Weeks --

Daniel E. Weeks, Ph.D.

The Wellcome Trust Centre for Human Genetics	Department of Human Genetics
University of Oxford	University of Pittsburgh
Windmill Road	Crabtree Hall, Room A310
Oxford OX3 7BN	130 DeSoto Street
United Kingdom	Pittsburgh, PA 15261
	USA
(+44) 1865 740 043 (desk)	1 412 624-5388 (desk)
(+44) 1865 742 441 (main)	1 412 624-3066 (secretary)
FAX: (+44) 1865 742 196	FAX: 1 412 624-3020
daniel.weeks@well.ox.ac.uk	dweeks@watson.hgen.pitt.edu

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

Linkage Newsletter

Vol. 9 No. 3 November 1995

Published by Jurg Ott, Columbia University, New York
Email: jurg.ott@columbia.edu
WWW: http://linkage.cpmc.columbia.edu
Postal address:
Columbia University, Unit 58
722 West 168 Street, New York, NY 10032

Editorial Assistant: Katherine Montague
Email: km165@columbia.edu or kath@linkage.cpmc.columbia.edu
Fax: 212-568-2750
Tel: 212-960-2507

LINKAGE COURSES

The next four linkage courses have been scheduled as follows:

January 8-12, 1996, at Columbia University, New York (advanced course, maximum of 20 participants, deadline for application is Nov. 20; course announced in previous Newsletter).

February 26 - March 1, 1996, at the University of Zurich, Irchel Campus Computer Center, Switzerland (basic course, maximum of 18 participants).

June 10-14, 1996, at Columbia University, New York (basic course, maximum of 30 participants).

June 24-28, 1996, at Columbia University, New York (basic course, maximum of 30 participants).

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.cpmc.columbia.edu (file *corr_ter.txt* in directory *book*).

SOFTWARE NEWS

Bug in FastLINK

Rita Kruse, Bonn Germany, made me aware of the following problem in FastLINK: For a pedigree with seven loops, the lod score was different depending on the order in which "loop individuals" were defined. For one order, a maximum lod score of 6.50 was obtained while under another order, with the same data, the maximum lod score was only 0.66. This difference occurred only with FASTLINK while the LINKAGE programs consistently furnished a lod score of 6.50. Dr. Alejandro Schaffer confirmed this bug and has already corrected it in version 2.3P of FASTLINK.

SIMLINK and RHMAP on the Web

The following announcement was contributed by Dr. Michael Boehnke, Ann Arbor:

For some time, I have been distributing two programs for genetic analysis: RHMAP and SIMLINK. RHMAP is a program package for the statistical analysis of radiation hybrid mapping data. SIMLINK is a program for evaluating the statistical power of a proposed linkage study. Both RHMAP and SIMLINK are written in FORTRAN 77, and both previously were distributed free of charge on floppy diskettes after completion of a user agreement.

Beginning October 31, 1995, the primary method of distribution for RHMAP and SIMLINK will be via the World Wide Web (WWW). To access the current versions of RHMAP and SIMLINK (2.01 and 4.11, respectively), you will need a WWW browser such as Mosaic or Netscape. Using a browser, you can download the software packages from http://www.sph.umich.edu/group/statgen/software.

Both analysis packages, containing source code and manuals, are available as archives for Unix and DOS systems. Once you download the program package archives, you will need to dearchive the files. On Unix computers the "tar" utility should be used, e.g., "tar -xvf filename". (For other computer systems, e.g., VMS, etc., one can usually obtain public-domain utilities which can read "tar" files. Links to several such utilities are provided on the web page.) Simply compile the FORTRAN 77 code on your computer to obtain an executable. For DOS computers the archives are self-extracting and contain DOS executables of the programs. Simply issue the name of the archive as a DOS command, e.g., rhmapzip.exe, and all the files in the package, including the executable(s), will be extracted.

Also beginning October 31, the primary method for notification of software updates for RHMAP and SIMLINK will be via e-mail. If you already have or do in the future obtain one or both of the programs, I urge you to add yourself to the appropriate mailing list by sending a short e-mail note to me at boehnke@umich.edu indicating which package(s) you use and your preferred e-mail address for notification. This will insure that you continue to get program updates, and will be helpful for me at grant renewal time.

For those who lack Internet access, I will continue to distribute RHMAP and SIMLINK by floppy diskette. However, I ask you to make use of this option only if it truly is necessary. In this case, you will need to request a user agreement which you will need to fill out and send back.

Please feel free to contact me if you have any questions.

Michael Boehnke — Phone: 313-936-1001
Department of Biostatistics — FAX: 313-763-2215
University of Michigan — E-Mail: boehnke@umich.edu

Announcement of GenoCheck program

The following announcement was contributed by Dr. Margaret Gelder Ehm, Houston:

GenoCheck, version 1.0

Meg Gelder Ehm
Rice University
Department of Statistics
Houston, Texas USA
gelder@stat.rice.edu

New error detection capabilities are available for genetic linkage data!

GenoCheck, 1.0 is an error checking program designed to identify individuals and loci that are likely to contain errors. The statistical method was designed to identify typing error, but is general enough to pinpoint any unlikely genotype still consistent with Mendelian inheritance.

GenoCheck was developed using FASTLINK 2.2 (modified version of LINKAGE 5.1) and uses a similar file configuration and installation procedure. The code contains checkpointing facilities that allow users to recover from crashes without having to rerun the program and comes with documentation.

The instructions for retrieving the code are given below:

```
ftp softlib.cs.rice.edu
```

Login as anonymous and leave your full e-mail address as the password.

```
cd pub/GenoCheck
```

In that directory you will find various files. You can get everything at once by retrieving:

```
genocheck.tar.Z
```

and then (outside of ftp) doing the commands:

```
uncompress genocheck.tar.Z
tar xvf genocheck.tar
```

If you prefer to get the files piecemeal, instead of getting genocheck.tar.Z start by getting README*. The file README (with no extension) will describe all of the documentation.

The statistical algorithm implemented in GenoCheck, version 1.0 is described in the papers:

M. G. Ehm, R. W. Cottingham Jr., and M. Kimmel. Error Detection in Genetic Linkage Data Using Likelihood Based Methods. Journal of Biological Systems, Vol. 3, No. 1 (1995) 13-25.

M. G. Ehm, R. W. Cottingham Jr., and M. Kimmel. Error Detection in Genetic Linkage Data Using Likelihood Based Methods. American Journal of Human Genetics, Vol. 58, No. 1 (1996) (to appear).

Please e-mail any questions or suggestions about GenoCheck to gelder@stat.rice.edu.

INFOBIOGEN server on the Web

Dr. Lucien Bachner (bachner@infobiogen.fr) requested that information about his server be included in this newsletter. Here is a brief description: The server may be accessed at URL http://www.infobiogen.fr/vjf/server_vjf.html. Academic users may obtain an account at Infobiogen (after returning a form) which enables them to access a large number of programs and databases useful for gene mapping and DNA analysis. Interaction with the server is in French.

Dr. Bachner is also organizing linkage training courses at the beginners level. The first was held October 24-25, he is planning the second for end of May 96 or June. Language is French. The same course is also scheduled at Institut Pasteur, December 19-20.

New version of ANALYZE package

A new version of the ANALYZE program package has been released. It simplifies the performance of a large array of parametric and nonparametric tests for linkage and association on data entered in LINKAGE format pedigree and parameter files. The programs are distributed at two different sites. You can obtain them either from

ftp.well.ox.ac.uk in directory software/analyze

or from

linkage.cpmc.columbia.edu in directory software/analyze.

Users are advised to download and study the README files. A DOS version of the whole package is in preparation and should be available on the latter ftp site by the end of November.

SURVEY

We occasionally hear from people that they are having problems obtaining our books through book stores. Typically these books are not stocked in regular book stores but those outlets designated by the Johns Hopkins University Press should always have books on stock. Readers who had problems obtaining our books (either through a book store or directly from the publisher) are invited to let us know; please also indicate the date when you had problems. We would like to have customer service improved in this area.

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

Jurg Ott
Columbia University
722 West 168 Street
New York, NY 10032

FAX +1 (212) 568-2750
Tel. (212) 960-2507

Introductory Linkage Course

Feb. 26-
March 1, 1996

University of Zürich, Switzerland (Irchel Campus Computer Center 12 IBM PS/2s, Proff. Eric Kubli (Zürich) and Jürg Ott (New York). Course fee: $600 (US) Maximum number of participants: 18 The course will be taught in English. Application deadline: December 20, 1995

Course Descriptions and Prerequisites

This course is for researchers with little or no experience in using linkage programs. A basic knowledge of linkage analysis is, however, required. *Topics:* Introduction to linkage analysis; practical aspects of data collection; strategies and methods of linkage analysis; incomplete penetrance (narrow and wide definition); inbreeding loops; simple risk calculations: introduction to computer simulation. The major part of the course will consist of carrying out exercises using the LINKAGE programs.

People should plan an arriving in Zürich no latter than Sunday evening prior to the course (February 25). For additional information, please fax a request to +1 (212) 568-2750 or call Katherine Monataque at (212) 960-2507.

Course Fee

The fee for the 5-day-course is *$600* for researchers at an academic institution, and *$700* for individuals from private (for profit) companies. For applicants from Switzerland: course fee -700 Fr.or Fr.560 for people affiliated with ETH or University of Zürich. Send no money now --applicants will receive a bill and information regarding cancellation policy. As there is presently no support for this course from sources other than course fee, no reduction of the cost to applicants if possible. This fee covers tuition and course related expenses (diskettes, etc.) but not room, board or meals. Course participants will receive a list of good and moderately priced hotels in Zürich and will not have to make their own arrangements.

Participants are expected to purchase the textbook which will be used in the course (Terwilliger and Ott, "Handbook for Human Genetic Linkage," John Hopkins Univ. Press). Our office can give you the address of representatives of John Hopkins University Press for your country upon request.

Jurg Ott
Columbia University, Unit 58
722 West 168 Street
New York, NY 10032

FAX +1 (212) 568-2750
Tel. (212) 960-2507

Introductory Linkage Course

Course #1 - Monday through Friday, June 10-14, 1996

Course #2 - Monday through Friday, June 24-28, 1996

Course Description and Prerequisites

This course is for researchers with little or no experience in using linkage programs. A basic knowledge of linkage analysis is, however, required. *Topics:* Introduction to linkage analysis; practical aspects of data collection; strategies and methods of linkage analysis; incomplete penetrance (narrow and wide definition); inbreeding loops; simple risk calculations; introduction to computer simulation. The major part of the course will consist of carrying out exercises using the LINKAGE programs.

The course will be held in the microcomputer classroom of the Health Sciences Library (Columbia-Presbyterian Medical Center, 701 West 168th Street) which is equipped with 20 IBM PS/2 machines. The number of participants is *limited to 30*. People should plan on arriving in New York no later than in the evening of Sunday, June 9 or June 23, 1996. For additional information, please fax a request to +1 (212) 568-2750 or call Katherine Montague at (212) 960-2507.

Application deadline: March 11, 1996. Participants are accepted on a "first come, first served" basis for the beginning course, so applications should be sent in as soon as possible.

Course Fee

The fee for the 5-day-course is *$600* for researchers at an academic institution, and *$700* for individuals from private (for profit) companies. It may be paid by check made payable to Columbia University Dept. of Psychiatry, or by Government pay order, but send no money now -- applicants will receive a bill and information regarding cancellation policy. As there is presently no support for this course from sources other than the course fee, no reduction of the cost to applicants is possible. This fee covers tuition and course related expenses (handouts, diskettes, etc.) but not room, board or meals. Course participants will receive a list of good and moderately priced hotels ($65+) in New York and will have to make their own arrangements (except foreign participants).

Participants are expected to purchase the textbook which will be used in the course (Terwilliger and Ott, "Handbook for Human Genetic Linkage," Johns Hopkins Univ. Press). Our office can give you the address of representatives of Johns Hopkins University Press.

Linkage Newsletter

Vol. 10 No. 1 March 1996

Published by Jurg Ott, Columbia University, New York
Email: ott@rockvax.rockefeller.edu
WWW: http://linkage.cpmc.columbia.edu

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

Editorial Assistant: Katherine Montague
Email: km165@columbia.edu
Fax: 212-568-2750
Tel: 212-960-2507

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

EDITORIAL

After ten years with Columbia University, the 'genetic linkage' group will be moving to Rockefeller University, New York, where a new *Laboratory of Statistical Genetics* is being formed. The move will take place throughout this summer and will be completed in November of this year. A new ftp site and Web page will be set up at Rockefeller but addresses are not known yet (note Dr. Ott's new email address above). They will be posted on our current ftp site/Web page, which will continue to exist for some time (at least one year).

LINKAGE COURSES

The next courses in genetic linkage analysis have been scheduled as follows, and interested researchers are urged to **apply early** as our courses tend to fill up quickly:

June 10-14, 1996, at Columbia University, New York (basic course, maximum of 30 participants).

June 24-28, 1996, at Columbia University, New York (basic course, maximum of 30 participants). The second course will be held only with a sufficient number of applications (which we expect). Deadline for applications has been extended to April 8.

October 7-11, 1996, at University of Zurich, Switzerland (advanced course, maximum of 12 participants). Deadline for applications is August 30, 1996.

October 14-18, 1996, at Rockefeller University (advanced course, maximum of 20 participants, with preference to participants at U.S. institutions and companies). Deadline for applications is August 30, 1996.

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.cpmc.columbia.edu (file *corr_ter.txt* in directory *book*).

SOFTWARE NEWS

Fortran and Windows 95

The only Fortran program currently distributed by us is LIPED. If compiled with Microsoft (MS) Fortran version 5.10, various compiler switches allow it to run under DOS, OS/2 and Windows 3.1. Under Windows 95, programs compiled for DOS still run in a DOS window; however, programs compiled for Windows 3.1 will not run in Windows 95. Presumably, for Windows 95 an updated Fortran compiler must be obtained from Microsoft. If anyone has experience in this matter, suggestions would be appreciated.

OS/2 and Windows 95

Even though I was an early fan of OS/2 ever since its version 1.1 came out, "market forces" made me decide that we no longer want to support OS/2 versions of our programs; these versions are still available on our ftp site but they will not be updated. Users with a need for running large programs may find it convenient to turn to our programs compiled with NDP Pascal for DOS. This compiler does not impose any limitations on array sizes, code or data segments, etc.

FASTLINK 3.0P

(submitted by A. Schäffer)

I am pleased to announce that FASTLINK 3.0P is now available. FASTLINK is a faster version of the main programs in LINKAGE. As of version 2.3P, FASTLINK runs in parallel on either UNIX multiprocessors or networks of UNIX workstations.

As with previous versions, the code is available by ftp from:

softlib.cs.rice.edu

Login as anonymous, leave full e-mail address as password.

cd pub/fastlink

If you are a UNIX user:

get fastlink.tar.Z

If you are a DOS user:

cd dos

<retrieve everything in that directory>

If you are a VMS user:

get README

get README.VMS

and follow the instructions there for what you need

Retrievers in Europe may find it more convenient to retrieve from the mirror site at:

ftp.ebi.ac.uk

The instructions are similar, but instead of:

cd pub/fastlink (for softlib)

do

cd pub/software/linkage_and_mapping/FASTLINK/fastlink (for ebi)

Among the improvements in FASTLINK 3.0P (compared to 2.3P) are:

--The code runs faster on data sets that have loops or unused alleles. If you used FASTLINK 2.3P, you may have noticed that it printed diagnostics when a data set had unused alleles, but it did not take advantage of the situation.

--UNKNOWN now detects violations of Mendelian rules of inheritance in looped pedigrees. It never did this before.

--maxhap and maxfem are no longer in the code as constants. If you were resetting maxhap each time and recompiling, you won't have to do that any more. If you were setting maxhap unnecessarily high to avoid recompiling, you may perceive some speedup.

See the file README.updates for more details on recent code changes in FASTLINK.

Some algorithmic aspects of the most recent improvements can be found in:

A. A. Schaffer, Faster Linkage Analysis Computations for Pedigrees with Loops or Unused Alleles, *Human Heredity*, to appear.

This paper can be found as paper5.ps at the ftp sites.

Alejandro Schaffer
schaffer@nchgr.nih.gov

Two problems in LINKAGE programs

(submitted by A. Schäffer)

Two LINKAGE problems have arisen recently in FASTLINK bug reports that users should be aware of. I call them “problems” and not “bugs” because one has to abuse LINKAGE to get them to happen.

Problem 1. If a loop involves someone who is multiply married, and it is unbroken, then LINKAGE might not go into an infinite loop. If it doesn't, plausible but wrong results are printed. I have now seen 3 pedigrees with this problem. This is not a “bug” because if one uses MAKEPED with LOOPS embedded as one should, the loop will be caught there.

Problem 2. If an allele frequency is specified as 0.0 various problems can arise. The one I saw is:

--user specified frequency of first allele is 0
--data set had a pedigree in which nobody was typed at that locus
--unknown converted everyone to homozygous 1 1 at that locus
--because the allele frequency was 0.0, the likelihood also came back as 0.0, making the log likelihood -infinity.

Both problems are also present in all versions of FASTLINK until the most recent (3.0P). I fixed problem 1 with a new diagnostic in UNKNOWN in the initial 3.0P release. I addressed problem 2 with a new warning in the main programs.

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

Linkage Newsletter

Vol. 10 No. 2 July 1996

Published by Jurg Ott, Columbia University, New York
Email: ott@rockvax.rockefeller.edu
WWW: http://linkage.cpmc.columbia.edu

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

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

Fax: 212-568-2750
Tel: 212-960-2507

(this newsletter and all previous issues are available on our ftp site/Web page: linkage.cpmc.columbia.edu)

EDITORIAL

This 'summer' edition of the Newsletter is very short — a more extensive edition will follow in the autumn.

LINKAGE COURSES

The last linkage course at Columbia University was held June 24-28. The next linkage courses will be at Rockefeller University (and, as in the past, at the University of Zürich). Dates are given below. Interested researchers are urged to **apply early** as our courses tend to fill up quickly:

October 7-11, 1996, at University of Zürich, Switzerland (advanced course, maximum of 12 participants). Deadline for applications is August 20, 1996.

October 14-18, 1996, at Rockefeller University (advanced course, maximum of 20 participants, with preference to participants from U.S. institutions and companies). Deadline for applications is August 30, 1996.

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.cpmc.columbia.edu (file *corr_ter.txt* in directory *book*).

OPEN POSITIONS

In our new *Laboratory of Statistical Genetics* at Rockefeller University, New York, we are looking for colleagues to join us in our efforts to develop, implement, and apply methods of linkage analysis and association to find genes underlying complex traits. Positions are open at the postdoctoral level, and faculty positions may be discussed at any time. We are particularly interested in researchers with a good background in statistical genetics. Rockefeller University is located in a prime area in Manhattan, situated in a garden-like campus. The University's *Starr Center for Human Genetics* provides diverse opportunities for application of new methods and active collaboration on interesting data sets.
Valeria Brancolini, Marcella Devoto, Suzanne Leal, Tara Matise, Xiaoli Xie, Thomas Lehner, Wentian Li, Jurg Ott

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

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) 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 (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) || !noloop);
```

It should read:

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

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 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.

Linkage Newsletter

Vol. 12 No. 1 February 1998

Published by Jurg Ott, Rockefeller University, New York
Email: ott@rockefeller.edu
URL: http://linkage.rockefeller.edu

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

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

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

(this and all previous newsletters are available at our web site, http://linkage.rockefeller.edu/newsletter/)

EDITORIAL

For various reasons, no newsletter was written last year. The last published newsletter is dated December 1996 and was mailed just a little over one year ago. New issues of the newsletter will be published depending on need and how much material we receive for inclusion.

LINKAGE COURSES

The next basic linkage courses will be held as indicated below. Interested researchers are urged to **apply early** as our courses tend to fill up quickly. To obtain information on these courses, please write to Katherine Montague, course coordinator, by email (preferred) or fax.

March 16-20, 1998, at University of Zürich, Switzerland (maximum of 18 participants). Deadline for applications was January 15, 1998, but we still have one or two open slots.

June 15-19, 1997, at Rockefeller University (maximum of 30 participants). Deadline for applications is May 1, 1998.

We will use our book (Terwilliger and Ott, *Handbook of Human Genetic Linkage,* Johns Hopkins University Press, 1994). 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. For a list of corrections see *http://linkage.rockefeller.edu/ott/corr-ter.htm.*

The date for this year's advanced course (summer/fall of 1998) is not yet known. Please watch our web page at *http://linkage.rockefeller.edu/course/here.html* for dates. Advanced courses are reserved for individuals with a very good background in linkage analysis). Our advanced courses have been greatly modernized and now include presentations by outside scientists who are experts in areas such as genomic databases or particular computer programs. Therefore, we can no longer offer advanced courses outside of New York. Depending on demand, we will offer more than one advanced course per year. This will be particularly important for foreign participants since for the advanced course supported by the NIHGRR, the number of participants from outside the U.S. is very limited.

SOFTWARE

New program

(submitted by Dr. Alejandro Schaffer)

We announce the availability of CASPAR (Computerized Affected Sibling Pair Analyzer and Reporter), a software package for studying the genetics polygenic diseases. CASPAR's main novel feature is conditional linkage analyses, in which the population can be subdivided according to criteria at some loci and analyzed for linkage at other loci. CASPAR uses simulation to overcome the problems inherent in such multiple testing.

To retrieve CASPAR:
ftp fastlink.nih.gov
cd pub/caspar
binary
get caspar.tar.Z

Then on your computer:
uncompress caspar.tar.Z
tar xvf caspar.tar

Please send comments, questions, complaints, and problem reports to:
Richa Agarwala (richa@helix.nih.gov) and
Alejandro Schaffer (schaffer@helix.nih.gov)

Other participants in CASPAR development are:
Jeremy Buhler, Kenneth Gabbay, Marek Kimmel, David Owerbach

FastLINK

(submitted by Dr. Alejandro Schaffer)

FASTLINK version 4.0P is now available. The principal improvement over version 3.0P is a method to automatically select loop breakers that addresses a problem posed in 1975 by Ken Lange and Robert Elston. The new code was developed in collaboration with Ann Becker and Dan Geiger of Technion. A description of what we did can be found in:

A. Becker, D. Geiger, A. A. Schaffer (1998) Automatic Selection of Loop Breakers for Genetic Linkage Analysis. *Hum Hered* 48:49-60

This paper comes as paper6.ps with the code distribution.
To retrieve for other than DOS:
ftp fastlink.nih.gov
binary
cd pub/fastlink
get fastlink.tar.Z
For DOS replace the last two lines with:
cd pub/fastlink/dos
mget *

Thanks to Jim Tomlin (NIH) and John Powell (NIH) for vigorous beta testing. Thanks to Marilyn Raymond (NIH) for a great challenge data set. Thanks to Don Plugge for assistance with the port to VMS.

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