

GeneFinder 1.0

Software Manual

*Wei-Min Chen**

Kung-Yee Liang†

Yen-Feng Chiu

Johns Hopkins University
Department of Biostatistics
Bloomberg School of Public Health
Baltimore, Maryland 21205
U.S.A.

November 26, 2002

*Email: wechen@jhsph.edu

†Email: kyliang@jhsph.edu

Contents

1	Introduction	2
2	Installation	3
3	Running GeneFinder	4
4	Commands	5
4.1	General Commands	5
4.1.1	help	5
4.1.2	quit	5
4.2	Input Commands	6
4.2.1	in	6
4.2.2	marker	6
4.2.3	pedigree	7
4.2.4	ibd	7
4.3	Executable Commands	8
4.3.1	gee	8
4.3.2	3cgee	9
4.3.3	plot	10
4.4	Analysis Commands	11
4.4.1	condition	11
4.4.2	set	12
4.4.3	map	13
4.5	Auxiliary Commands	14
4.5.1	run	14
4.5.2	system	14
4.5.3	time	15
4.5.4	summary	15
5	An Example Session	16
6	Acknowledgment	23

1 Introduction

Most of the previous non-parametric linkage analysis (including GeneHunter, Allegro, SAGE, MERLIN, etc.) focus on hypothesis testing, i.e., they can only answer within a certain region whether or not susceptibility genes exist. None of them could provide a 95% confidence interval for the gene location (e.g., for the purpose of further fine mapping). GeneFinder uses the GEE method to estimate the location of the susceptibility gene or QTL based on the IBD (identical by descent) sharing of (affected) sib pairs for multiple markers. GeneFinder can further the analysis of GeneHunter by providing a more accurate gene location estimation and the corresponding confidence interval (Liang, Chiu and Beaty 2001). Furthermore, GeneFinder can estimate the gene location in a variety of scenarios, which can be summarized as follows:

1. locating the susceptibility gene (Liang, Chiu and Beaty 2001);
2. locating the susceptibility gene incorporating the information of another unlinked genomic region which is known to contain one of the susceptibility genes (Liang, Chiu, Beaty and Wjst 2001);
3. locating the susceptibility gene incorporating the covariate information on sib pairs which is dichotomous, or which can be dichotomized, such as onset age (Glidden, Liang, Chiu and Pulver 2002);
4. locating the susceptibility gene incorporating the covariate information, without any restriction on the covariate (this function is coming soon. Glidden, Liang, Chiu and Pulver 2002);
5. locating the QTL based on the quantitative trait which can be dichotomized by a certain threshold (Liang, Huang and Beaty 2000);
6. locating the QTL based on the quantitative trait, without any restriction on the trait (this function is coming soon).

2 Installation

GeneFinder can be freely downloaded from website

<http://www.biostat.jhsph.edu/~wmchen/gf.html>

After downloading GeneFinder 1.0 for Solaris, You may need to type 'chmod a+x gf' in UNIX to make the program executable.

WINDOWS users can download and run gf.exe file in WINDOWS or DOS shell. Since GeneFinder relies on GeneHunter's IBD calculation, which can only be obtained in UNIX, WINDOWS users need to do file transferring from UNIX.

For other operating system users, you are welcome to contact the authors at wechen@jhsph.edu for further assistant. This program is written in ANSI C++ and it is easy to be compiled and run without modification in other operating systems.

The first version of GeneFinder was written by Dr. Yen-Feng Chiu, and it is available at website:

<http://www.biostat.jhsph.edu/biostat/research/genefinder.shtml>

3 Running GeneFinder

The only requirement for running GeneFinder is that one must run GeneHunter first to obtain the IBD file. Once this step is done, it is simple to run GeneFinder. Type 'gf' in command line to run GeneFinder. In the GeneFinder running environment, you can type 'help' to get help, and 'quit' to quit. The most frequently used commands are 'in' and 'gee'. A typical session is like:

```
GeneFinder> in gh.in
GeneFinder> plot
GeneFinder> gee
GeneFinder> gee 41
GeneFinder> con p gh2.in 71.1
GeneFinder> cgee
```

'in gh.in' imports the necessary input for GeneFinder, from the marker-locus file, pedigree file, and IBD file specified by file 'gh.in', which is the batch job file of GeneHunter. Command 'plot' plots the allele-sharing scores against locations. Command 'plot' is very helpful for users to choose the correct estimation for gene location. By command 'gee', a series of GEE analysis with various initials are performed. The proper initials should be chosen by users, by the criteria of maximum allele-sharing, if the estimates are not consistent. For the above example, 41 is picked as the proper initial for the location τ and by rerunning 'gee' with parameter, a detailed analysis report is printed. 'con' is the alias of command 'condition'. It prepares the necessary input (as conditions) for conditional analysis. Like 'gee', 'cgee' without parameters runs a series of condition GEE analysis.

4 Commands

4.1 General Commands

4.1.1 help

Purpose: GeneFinder help facility

Usage:

help

help [command]

Example:

help

help gee

Alias: ?

Notes: 'help' without parameters prints general help information. 'help' with a command name as its parameter prints help information specific for this command.

4.1.2 quit

Purpose: exit session

Usage: quit

Example: quit

Alias: exit

4.2 Input Commands

4.2.1 in

Purpose: read marker-locus data, pedigree data, and IBD data from the batch job file of GeneHunter

Usage: in [batch job file]

Example: in gh.in

Notes: Command 'in' prepares necessary inputs for GEE analysis. The [batch job file] is a file that includes a series of commands for GeneHunter. It is also the parameter of 'RUN' command in GeneHunter. Command 'in' can replace the combination of commands 'marker', 'pedigree', 'ibd', and 'map', and is thus highly recommended. By this command, GeneFinder is able to identify keywords 'load marker', 'scan', 'dump ibd', and 'map' in the batch job file. See 'marker', 'pedigree', 'ibd', and 'map' for more details.

4.2.2 marker

Purpose: load marker-locus data

Usage: marker [filename]

Example: marker gf.loc

Alias: mar

Notes: The format of this marker-locus data file must be identical to LINKAGE and GENEHUNTER parameter file. Also see command 'in'.

4.2.3 pedigree

Purpose: load pedigree data

Usage: pedigree [filename]

Example: pedigree gf.ped

Alias: ped

Notes: The format of this pedigree file must be identical to LINKAGE and GENEHUNTER parameter file. Also see command 'pedigree'.

4.2.4 ibd

Purpose: load IBD data generated by GeneHunter

Usage: ibd [filename]

Example: ibd gh.ibd

Notes: Currently only IBD data generated by GeneHunter can be recognized. The estimated probabilities of alleles IBD at each marker locus are provided by this IBD file. Sometimes, besides the marker loci, probabilities at equally spaced positions between adjacent markers are calculated (but these extra positions have no effects on the result of the GEE analysis). This can be accomplished by command 'increment step [n]' for GeneHunter. GeneFinder can identify the step size automatically. IBD data generated from GeneHunter by using 'increment distance 1' is not allowed, since it is not guaranteed to hit every marker position, while the GEE analysis demands IBD information at each marker locus.

4.3 Executable Commands

4.3.1 gee

Purpose: estimate the location of the susceptibility the gene using GEE method

Usage:

```
gee  
gee [location  $\tau$ ] [C]
```

Example:

```
gee  
gee 71  
gee 71 0.3
```

Notes: Command 'gee' is the analysis part of GeneFinder. 'gee' without parameters tries a series of initials automatically and the GEE analysis will repeat a number of times using these different initials. Users can choose the appropriate estimate among these analysis. 'gee' with parameters performs GEE analysis for one time, taking the parameters as initials. It prints a report in more detail. The first parameter for command 'gee' is the initial value for location τ , and the second parameter is the initial value for C (another key parameter for the GEE model which ranges from -1 to 1). In some circumstances, one might need to use 'set' command to change parameters needed in numerical calculation for GEE method. One can use bootstrap confidence interval instead of the asymptotic confidence interval by command "set b n". One can try more random initials by command "set r n". See command 'set' for more details.

Warning: Small p-value cannot be interpreted as evidence for a 'reasonable' estimate of the gene location, when the location estimation is not consistent (i.e., more than one significant estimation exists). Make sure global maximum of allele-sharing arrives at the estimated position.

4.3.2 3cgee

Purpose: estimate the location of the susceptibility gene using the GEE method, conditioning other information which forms three categories

Usage:

cgee

cgee [location τ] [C_0] [C_1] [C_2]

Example:

cgee

cgee 71

cgee 71 0.2 0.4 0.15

Alias: cgee

Notes: Command '3cgee' performs multipoint linkage analysis using GEE methods, incorporating other information, such as linkage evidence from another unlinked region, or the information of continuous trait values which could be dichotomized. '3cgee' without parameters tries a series of initials automatically and the analysis will repeat a number of times using these different initials. Users can choose the appropriate estimate among these analysis. '3cgee' with parameters performs GEE analysis for one time, taking the parameters as initials. It prints a report in more detail. The first parameter for '3cgee' is the initial value for location, and other parameters are the initial values for C_0 , C_1 , and C_2 , which range from -1 to 1. To run this command, one needs to run command 'condition' first. See command 'condition' for more details. In some circumstances, one might need to use command 'set' to change parameters needed in numerical calculation for GEE method. One can use bootstrap confidence interval instead of the asymptotic confidence interval by command "set b n". One can try more random initials by command "set r n". See command 'set' for more details.

Warning: Small p-value cannot be interpreted as evidence for a 'reasonable' estimate of the gene location, when the location estimation is not consistent (i.e., more than one significant estimation exists). Make sure global maximum of allele-sharing arrives at the estimated position.

4.3.3 plot

Purpose: plot allele-sharing scores along the chromosome

Usage: plot [filename]

Example: plot myplot.ps

alias: draw

notes: Command 'plot' plots the allele-sharing scores, which are two times as large as the estimated proportion of alleles IBD. A postscript file is generated according to the allele-sharing IBD data from GeneHunter. This allele-sharing plot is especially useful when the estimates are not consistent. If one has prepared the 3C GEE analysis by command 'condition', then command 'plot' plots four curves: one is the allele-sharing score, which is exactly the same as in the unconditional analysis, denoted by a thick line; the other three, which are denoted by thin line, dotted line, and even-spaced line, respectively, are allele-sharing scores that are conditional on sharing 0, 1, or 2 alleles IBD of another unlinked region, or LL (low and low), LH (low and high), HH (high and high) quantitative traits or covariates.

4.4 Analysis Commands

4.4.1 condition

Purpose: preparation for the analysis of conditional GEE analysis (cgee)

Usage:

condition marker [additional batch job file] [marker order]

condition locus [additional batch job file] [locus number]

condition position [additional batch job file] [position]

condition trait [threshold] [trait number]

condition covariate [threshold] [covariate number]

Example:

```
condition marker gh2.in 9
```

```
con l gh2.in 9
```

```
con p gh2.in 40.2
```

```
con t 3.2
```

```
con c 18
```

Alias: con

Notes: Several applications of GeneFinder rely on command 'condition'. By this command, the following information can be incorporated into the analysis:

1. information at a known disease susceptibility locus in the other unlinked genomic region (by option 'm' to specify the marker number nearby the known gene, by option 'l' to specify the marker locus nearby the known gene, or by option 'p' to specify the particular position of the unlinked known gene);
2. information of the covariate which is dichotomous or can be dichotomized (by option 'c' followed by a threshold to specify the covariate of interest);
3. information of the quantitative trait which can be dichotomized (by option 't' followed by a threshold to specify the quantitative trait of interest).

4.4.2 set

Purpose: set parameters for 'gee' and 'cgee' analysis

Usage:

```
set set bootstrap [times]
set random [times]
set level [a level for CI]
set iterative [times]
set epsilon [a small number]
set delta [criteria for Newton-Raphson algorithm to stop iteratives]
```

Example:

```
set
set b
set b 200
set b 0
set r 50
set l 0.99
set i 5000
set e 0.5
set d 1E-12
```

notes: Command 'set' without any parameters prints the values of all environment variables. Command 'set' with one parameter prints the specific value corresponding to the variable (the parameter). Command 'set' with two parameters assigns the second parameter to the variable represented by the first parameter. Option 'b' with parameter (simulation times) greater than 0 allows GEE can CGEE to use BOOTSTRAP method to calculate the confidence interval. The BOOTSTRAP confidence interval should be more accurate than the confidence interval based on asymptotic theories in some situation. Option 'b' with parameter 0 stops the future BOOTSTRAP calculations. Option 'r' allows GeneFinder to do GEE or 3CGEE analysis (without parameters) with randomly assigned initials, which should be very useful when the performance of GeneFinder is not good (not consistent or with many non-convergence). Option 'l' with parameter (level of confidence interval) allows users to choose a different level of confidence interval, which

is set 0.95 in default. Option i (iteration times), e (to modify the map function) and d (threshold of Newton-Raphson procedure) can fine tune the numerical procedure of the GEE analysis. These environment variable setups apply to both command 'gee' and 'cgee'.

4.4.3 map

Purpose: specify a genetic distance map function

Usage:

map

map [function name]

Example:

map

map Haldane

map Kosambi

notes: The default map function is Kosambi map function. Command 'map' without parameters prints the map function that is currently used in GeneFinder. Geneticists often use Kosambi's (1944) map function, which allows for modest interference: $m = \frac{100}{4} \log\left(\frac{1+2\theta}{1-2\theta}\right)$. Haldane map function assumes crossovers occur randomly and independently over the entire chromosome: $m = -\frac{100 \log(1-2\theta)}{2}$.

4.5 Auxiliary Commands

4.5.1 run

Purpose: perform a series of commands in a batch job file.

Usage: run [batch job file]

Example: run gf.in

notes: Although in most situations GeneFinder is too simple to run and this command might be not necessary, the batch job file is useful when one would like to save the outputs of GeneFinder. To accomplish this, one can either type 'gf <gf.in >gf.out' or 'gf gf.in >gf.out' in command line. Do not be confused by this '.in' file with the '.in' file for GeneHunter.

4.5.2 system

Purpose: execute a command provided by operating system without quitting GeneFinder

Usage: system [command]

alias: sys

notes:

For UNIX user, the following commands provided by UNIX are often useful

sys ls

sys cat [filename]

sys more [filename]

sys pwd

sys emacs [filename]

4.5.3 time

Purpose: show the current time in seconds from the beginning of the program running

Usage: time

4.5.4 summary

Purpose: summarize the necessary information extracted from the data

Usage: sum

alias: sum

5 An Example Session

Here is an example run in DOS shell. The output from UNIX is similar.

```
C:\genefinder>gf
```

```
*****
```

```
GeneFinder version 1.0.0
```

```
by Wei-Min Chen, Kung-Yee Liang, Yen-Feng Chiu, Nov 12, 2002
```

```
Copyright (C) Johns Hopkins University
```

```
Bloomberg School of Public Health Department of Biostatistics
```

```
*****
```

```
Type 'help' for help, 'quit' to quit
```

```
GeneFinder> sys type gh13.in
```

```
skip large off
```

```
haplotype off
```

```
map Haldane
```

```
increment step 1
```

```
load markers chr13.loc
```

```
photo test.pho
```

```
scan chr13.pre
```

```
dump ibd chr13.ibd
```

```
quit
```

```
GeneFinder> in gh13.in
```

```
Load marker-locus data from chr13.loc...loaded.
```

```
Load pedigrees from chr13.pre...loaded.
```

```
Load IBD data from chr13.ibd...loaded.
```

	Marker	location	allele_sharing
1	D13S115	0.00	0.98922
2	D13S787	1.73	0.98911
3	D13S192	16.25	1.00188
4	D13S218	50.11	0.96041
5	D13S325	55.38	0.92271
6	D13S126	59.19	0.92813
7	D13S155	61.13	0.94096
8	D13S133	62.34	0.93191
9	HTR2A	62.34	0.93181
10	D13S788	62.64	0.99465
11	D13S131	71.12	1.07154
12	D13S152	75.91	1.10844
13	D13S162	78.89	1.13697
14	D13S317	85.46	1.21834
15	D13S170	85.46	1.21834
16	D13S795	94.23	1.27260
17	D13S167	95.45	1.27862

18	D13S197	100.16	1.29829
19	D13S71	104.77	1.34904
20	D13S159	109.16	1.41523
21	D13S779	111.20	1.35916
22	D13S1240	111.70	1.41185
23	D13S225	112.30	1.37850
24	D13S174	113.93	1.44117
25	D13S280	114.53	1.43279
26	D13S274	117.20	1.30548
27	D13S797	121.37	1.28672
28	D13S796	126.64	1.28421

GeneFinder> plot chr13.ps

File chr13.ps is created to plot allele-sharing against locations.

GeneFinder> gee

tau0	C0	tau	C
0.86	0.330	Estimated	coeff=1.#Q0 out of range [-1, 1].
8.99	0.330	Estimated	coeff=-1.#I0 out of range [-1, 1].
33.18	0.330	Estimated	coeff=1.#Q0 out of range [-1, 1].
52.74	0.330	Estimated	coeff=1.#Q0 out of range [-1, 1].
57.28	0.330	Estimated	coeff=1.#Q0 out of range [-1, 1].
60.16	0.330	Estimated	coeff=1.#Q0 out of range [-1, 1].
61.73	0.330	Estimated	coeff=1.#Q0 out of range [-1, 1].
62.34	0.330	Estimated	coeff=-1.#I0 out of range [-1, 1].
62.49	0.330	113.23	0.347
66.88	0.330	113.23	0.347
73.52	0.330	113.23	0.347
77.40	0.330	113.23	0.347
82.17	0.330	113.23	0.347
85.46	0.330	113.23	0.347
89.85	0.330	113.23	0.347
94.84	0.330	113.23	0.347
97.81	0.330	113.23	0.347
102.47	0.330	113.23	0.347
106.96	0.330	113.23	0.347
110.18	0.330	113.23	0.347
111.45	0.330	113.23	0.347
112.00	0.330	113.23	0.347
113.12	0.330	113.23	0.347
114.23	0.330	113.23	0.347
115.87	0.330	113.23	0.347
119.29	0.330	113.23	0.347
124.01	0.330	113.23	0.347

Choose the initials (according to allele-sharing values if not consistent) and run 'gee <tau0> <C0>' again to get a report in more details.

GeneFinder> set b 200

GeneFinder> gee 110

GeneFinder 1.0

Number of markers: 28
 Number of families: 41
 Number of ASP: 66

	Location	C
Estimates	113.23	0.347
95% BOOTSTRAP Confidence Interval	101.91 116.84	0.170 0.556
95% right-sided CI bound	105.79	0.202
95% left-sided CI bound	116.33	0.524
95% ASYMPTOTIC Confidence Interval	108.74 117.72	0.155 0.538
95% right-sided CI bound	109.47	0.186
95% left-sided CI bound	117.00	0.507
Covariance Matrix	5.25 -0.01	-0.006 0.010
Initial values	110.00	0.330
Z score for estimation	3.54	
P-value for estimation	0.0004	

```
GeneFinder> in gh22.in
Load marker-locus data from chr22.loc...loaded.
Load pedigrees from chr22.pre...loaded.
Load IBD data from chr22.ibd...loaded.
```

	Marker	location	allele_sharing
1	2-D22S311	0.00	1.01509
2	2-D22S941	5.83	1.02314
3	04-COMT	11.65	0.99926
4	05-D22S264S	15.07	0.99868
5	06-CRYB2S	28.79	1.11182
6	8-D22S258	30.31	1.06661
7	7-D22S315	31.32	1.02762
8	9-D22S693Q	37.71	1.02249
9	10-D22S268S	44.11	1.12316
10	13-D22S280	45.32	1.10349
11	13-D22S1268Q	46.53	1.17626
12	16-ACT2A09Q	47.75	1.16639
13	15-D22S304	49.48	1.16634
14	14-D22S691Q	50.49	1.16430

15	17-D22S278Q	51.50	1.15554
16	11-D22S684Q	52.51	1.14616
17	19-D22S683Q	53.52	1.08748
18	18-D22S283Q	55.66	1.07405
19	20-D22S692	58.55	1.07619
20	22-D22S1045Q	59.86	1.09085
21	21-IL2RB	61.18	1.10298
22	23-D22S445	63.12	1.07019
23	26-PDGFB	65.05	1.01300
24	25-D22S299	65.76	0.98205
25	30-D22S307Q	66.46	0.97069
26	24-CYP2D	66.77	0.95893
27	28-D22S282	67.27	0.94619
28	29-D22S270Q	67.77	0.93034
29	27-D22S302	70.55	0.96276
30	31-D22S274Q	74.39	1.03599
31	2-ARSA	80.22	1.04288
32	33-D22S1267	86.04	1.07009

GeneFinder> gee

tau0	C0	tau	C
2.91	0.330	36.62	0.298
8.74	0.330	36.62	0.298
13.36	0.330	36.62	0.298
21.93	0.330	36.62	0.298
29.55	0.330	36.62	0.298
30.82	0.330	36.62	0.298
34.52	0.330	36.62	0.298
40.91	0.330	40.57	0.282
44.71	0.330	40.57	0.282
45.93	0.330	40.57	0.282
47.14	0.330	40.57	0.282
48.61	0.330	40.57	0.282
49.98	0.330	40.57	0.282
50.99	0.330	40.57	0.282
52.00	0.330	40.57	0.282
53.01	0.330	40.57	0.282
54.59	0.330	40.57	0.282
57.11	0.330	40.57	0.282
59.21	0.330	40.57	0.282
60.52	0.330	40.57	0.282
62.15	0.330	40.57	0.282
64.09	0.330	40.57	0.282
65.41	0.330	40.57	0.282
66.11	0.330	40.57	0.282
66.62	0.330	40.57	0.282
67.02	0.330	40.57	0.282
67.52	0.330	40.57	0.282
69.16	0.330	40.57	0.282
72.47	0.330	40.57	0.282
77.30	0.330	40.57	0.282
83.13	0.330	40.57	0.282

Choose the initials (according to allele-sharing values if not consistent) and run 'gee <tau0> <C0>' again to get a report in more details.

GeneFinder> gee 38

.....

GeneFinder 1.0

Number of markers: 32
 Number of families: 41
 Number of ASP: 66

```
-----
                Location      C
    Estimates      40.57      0.282
    95% BOOTSTRAP  19.76     -0.079
    Confidence Interval 61.10      0.561
    95% right-sided CI bound 23.22      0.098
    95% left-sided CI bound 57.07      0.548
    95% ASYPTOTIC  30.77     -0.009
    Confidence Interval 50.38      0.574
    95% right-sided CI bound 32.34      0.038
    95% left-sided CI bound 48.81      0.527
    Covariance      25.04     -0.068
    Matrix          -0.07      0.022
    Initial values  38.00      0.330
    Z score for estimation 1.90
    P-value for estimation 0.057
```

GeneFinder> con m gh13.in 24

	Marker	location	score	score1	score2	score3
1	2-D22S311	0.00	1.0151	1.1673	1.1134	0.9975
2	2-D22S941	5.83	1.0231	1.1519	1.1230	1.0269
3	04-COMT	11.65	0.9993	1.2002	1.1724	0.9561
4	05-D22S264S	15.07	0.9987	1.2039	1.1817	0.9486
5	06-CRYB2S	28.79	1.1118	1.2000	1.1251	1.1841
6	8-D22S258	30.31	1.0666	1.1985	1.1184	1.1202
7	7-D22S315	31.32	1.0276	1.1981	1.1077	1.0678
8	9-D22S693Q	37.71	1.0225	1.1026	1.1375	1.0159
9	10-D22S268S	44.11	1.1232	1.3007	1.2140	1.0837
10	13-D22S280	45.32	1.1035	1.3005	1.1309	1.1107
11	13-D22S1268Q	46.53	1.1763	1.3002	1.1896	1.1852

12	16-ACT2A09Q	47.75	1.1664	1.2990	1.1527	1.1825
13	15-D22S304	49.48	1.1663	1.2980	1.1831	1.1626
14	14-D22S691Q	50.49	1.1643	1.2979	1.2002	1.1482
15	17-D22S278Q	51.50	1.1555	1.2984	1.0888	1.1969
16	11-D22S684Q	52.51	1.1462	1.2500	1.1159	1.1889
17	19-D22S683Q	53.52	1.0875	1.1989	0.9441	1.2186
18	18-D22S283Q	55.66	1.0741	1.1998	0.9781	1.1319
19	20-D22S692	58.55	1.0762	1.2000	1.0645	1.0756
20	22-D22S1045Q	59.86	1.0909	1.1992	1.0089	1.1373
21	21-IL2RB	61.18	1.1030	1.2130	1.0085	1.1549
22	23-D22S445	63.12	1.0702	1.2686	0.9577	1.1362
23	26-PDGFB	65.05	1.0130	1.3980	0.9141	1.0834
24	25-D22S299	65.76	0.9821	1.4454	0.9078	1.0348
25	30-D22S307Q	66.46	0.9707	1.4924	0.9421	0.9894
26	24-CYP2D	66.77	0.9589	1.2948	0.9067	1.0338
27	28-D22S282	67.27	0.9462	1.3162	0.8818	1.0240
28	29-D22S270Q	67.77	0.9303	1.3376	0.8611	1.0079
29	27-D22S302	70.55	0.9628	1.4921	0.8932	0.9992
30	31-D22S274Q	74.39	1.0360	1.5071	0.9470	1.0694
31	2-ARSA	80.22	1.0429	1.3573	1.0115	1.0464
32	33-D22S1267	86.04	1.0701	1.2573	1.0587	1.0713

GeneFinder> cgee

tau0	C0_0	C0_1	C0_2	tau	C_0	C_1	C_2
2.91	0.200	0.400	0.150	37.30	0.421	0.237	0.305
8.74	0.200	0.400	0.150	37.30	0.421	0.237	0.305
13.36	0.200	0.400	0.150	37.30	0.421	0.237	0.305
21.93	0.200	0.400	0.150	37.30	0.421	0.237	0.305
29.55	0.200	0.400	0.150	37.30	0.421	0.237	0.305
30.82	0.200	0.400	0.150	37.30	0.421	0.237	0.305
34.52	0.200	0.400	0.150	37.30	0.421	0.237	0.305
40.91	0.200	0.400	0.150	41.87	0.404	0.206	0.294
44.71	0.200	0.400	0.150	41.87	0.404	0.206	0.294
45.93	0.200	0.400	0.150	41.87	0.404	0.206	0.294
47.14	0.200	0.400	0.150	41.87	0.404	0.206	0.294
48.61	0.200	0.400	0.150	41.87	0.404	0.206	0.294
49.98	0.200	0.400	0.150	41.87	0.404	0.206	0.294
50.99	0.200	0.400	0.150	41.87	0.404	0.206	0.294
52.00	0.200	0.400	0.150	41.87	0.404	0.206	0.294
53.01	0.200	0.400	0.150	45.52	0.373	0.178	0.275
54.59	0.200	0.400	0.150	45.52	0.373	0.178	0.275
57.11	0.200	0.400	0.150	45.52	0.373	0.178	0.275
59.21	0.200	0.400	0.150	45.52	0.373	0.178	0.275
60.52	0.200	0.400	0.150	45.52	0.373	0.178	0.275
62.15	0.200	0.400	0.150	45.52	0.373	0.178	0.275
64.09	0.200	0.400	0.150	45.52	0.373	0.178	0.275
65.41	0.200	0.400	0.150	45.52	0.373	0.178	0.275
66.11	0.200	0.400	0.150	45.52	0.373	0.178	0.275
66.62	0.200	0.400	0.150	45.52	0.373	0.178	0.275
67.02	0.200	0.400	0.150	45.52	0.373	0.178	0.275
67.52	0.200	0.400	0.150	45.52	0.373	0.178	0.275
69.16	0.200	0.400	0.150	45.52	0.373	0.178	0.275

```

72.47 0.200 0.400 0.150 73.04 0.503 0.034 0.255
77.30 0.200 0.400 0.150 73.04 0.503 0.034 0.255
83.13 0.200 0.400 0.150 73.04 0.503 0.034 0.255

```

Choose the initials (according to allele-sharing values if not consistent) and run 'cgee <tau0> <C0>' again to get a report in more details.

```
GeneFinder> cgee 41
```

```
.....
...

```

GeneFinder 1.0

```

Number of markers: 32
Number of families: 41
Number of ASP: 8+26+32

```

```
-----
```

	Location	C_0	C_1	C_2
Estimates	41.87	0.404	0.206	0.294
95% BOOTSTRAP	21.20	-0.618	-0.487	-0.178
Confidence Interval	73.71	0.739	0.608	0.608
95% right-sided CI bound	24.77	-0.389	-0.449	-0.040
95% left-sided CI bound	73.09	0.703	0.524	0.572
95% ASYMPTOTIC	33.94	-0.683	-0.239	-0.050
Confidence Interval	49.80	1.491	0.652	0.639
95% right-sided CI bound	35.22	-0.508	-0.168	0.005
95% left-sided CI bound	48.53	1.316	0.580	0.584
Covariance	16.36	-0.965	0.036	0.099
Matrix	-0.96	0.308	0.003	0.004
	0.04	0.003	0.052	-0.001
	0.10	0.004	-0.001	0.031
Initial values	41.00	0.200	0.400	0.150
Chi-square with df 3	4.11			
P-value for estimation	0.25			

6 Acknowledgment

The current GeneFinder software is based on a Fortran program and several script programs written by Dr. Yen-Feng Chiu. This work is supported by National Institutes of Health grant GM49909. May thanks to Dr. Karl Broman for his helpful comments and to Ms. Virginia K. Lasseter, Dr. Fang-Chi Hsu, and Ming-Yao Li for bug reporting.

References

- [1] Liang KY, Zeger SL (1986) Longitudinal data analysis using generalized linear models. *Biometrika* 73: 13-22.
- [2] Liang KY, Chiu YF, Beaty TH (2001) A robust identity by descent procedure using affected sib pairs: a multipoint approach for complex diseases. *Hum Hered* 51: 64-78.
- [3] Liang KY, Chiu YF, Beaty TH, Wjst M (2001) A multipoint analysis using affected sib pairs: incorporation of linkage evidence from unlinked regions. *Genetic Epidemiology* 21: 105-122.
- [4] Liang KY, Huang CY, Beaty TH (2000) A unified sampling approach for multipoint analysis of qualitative and quantitative traits in sib pairs. *Am J Hum Genet* 66: 1631-1641.
- [5] Glidden DV, Liang KY, Chiu YF, Pulver AE (2002) Multipoint linkage methods for localizing susceptibility genes of complex diseases (in press).
- [6] Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES (1996) Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am J Hum Genet* 58: 1347-1363.