Software

for genetic analyses

David Duffy

Queensland Institute of Medical Research
Brisbane, Australia

Introductory thoughts

- There are lots of interesting *statistical* problems in genetics and genetic epidemiology
- Some methods can be brought in from other domains, but often problems are unique
- Statistical geneticist like to roll their own
- And make them freely available over the Internet (some unimportant exceptions)
- There are enough data for everyone
- I’m not talking about bioinformatics

Fun things about genetic datasets

- Elaborately correlated data
- Lots of latent variables
- Strongly specified descriptions of how the latent variables work:
  - Population genetics
  - Mendelism
- Now, lots and lots of observed variables

A survey of available software

First stop: http://www.nslij-genetics.org/soft/

(also mirrored at http://linkage.rockefeller.edu/soft)

An Alphabetic List
of
Genetic Analysis Software

PAGE 1 (A-F)

URL
master: http://www.nslij-genetics.org/soft/
mirror: http://linkage.rockefeller.edu/soft/
searchable database: http:///www.animalgenome.org/soft/ (NEW!)

400,978

Last Update: September 01, 2006

Computer software on the following topics are included here: genetic linkage analysis for human pedigree data, QTL analysis for animal/plant breeding data, genetic marker ordering, genetic association analysis, haplotype construction, pedigree drawing, and population genetics. This list is offered here as a service to the gene mapping community. The inclusion of a program should not be interpreted as an endorsement to that program from us.

This page was created by Dr. Wentian Li, when he was at Columbia University (1995-1996). It was later moved to Rockefeller University (1996-2002), and now takes its new home at North Shore LIJ Research Institute (2002-now). More than 240 programs have been listed by December 2004, and more than 350 programs by August 2005.

Many software can be downloaded from EBI: ftp://ftp.ebi.ac.uk/pub/software/linkage_and_mapping/ (Linkage and Mapping Software Repository).

If you have new programs to add or any updated information, please send a message to webadm@nslij-genetics.org

what's new | link to other sources | obsolete programs
page 1 (A-F) | page 2 (G-L) | page 3 (M-P) | page 4 (Q-Z)
a b c d e f | g h i j k l | m n o p | q r s t u v w x [y] [z]

A

An Alphabetic List
of
Genetic Analysis Software

WHAT'S NEW

(2006-now)

what's new 2005 | what's new 2004 | what's new 2003 | what's new 2002 |
what's new 2001 | what's new 2000 | what's new 1999 | what's new 1998 |
what's new 1997 | what's new 1996 | what's new 1995 |

what's new | link to other sources | obsolete programs
page 1 (A-F) | page 2 (G-L) | page 3 (M-P) | page 4 (Q-Z)
a b c d e f | g h i j k l | m n o p | q r s t u v w x [y] [z]

- 09/01/06
  CFC : Contribution, Inbreeding (F) and Coancestry
- 08/24/06
  POWQ : power analysis (via simulation) of variance component multipoint linkage analysis of QTL
- 08/23/06
  GAIA : genetic association interaction analysis
- 08/21/06
  PSEUDO : fast evaluation of empirical p-values for linkage scans
- 08/11/06
  BIOLAD-DB : system/database for handling clinical and genetic data for addictive diseases
- 07/18/06
  RTDT : robust TDT
- 07/17/06
  HAPLOCLUSTERS :
  LAMP : Linkage and Association Modeling in Pedigrees

CRAN
Mirrors
What's new?
Task Views
Search

About R
R Homepage

Software
R Sources
R Binaries
Packages
Other

Documentation
Manuals
FAQs
Contributed
Newsletter

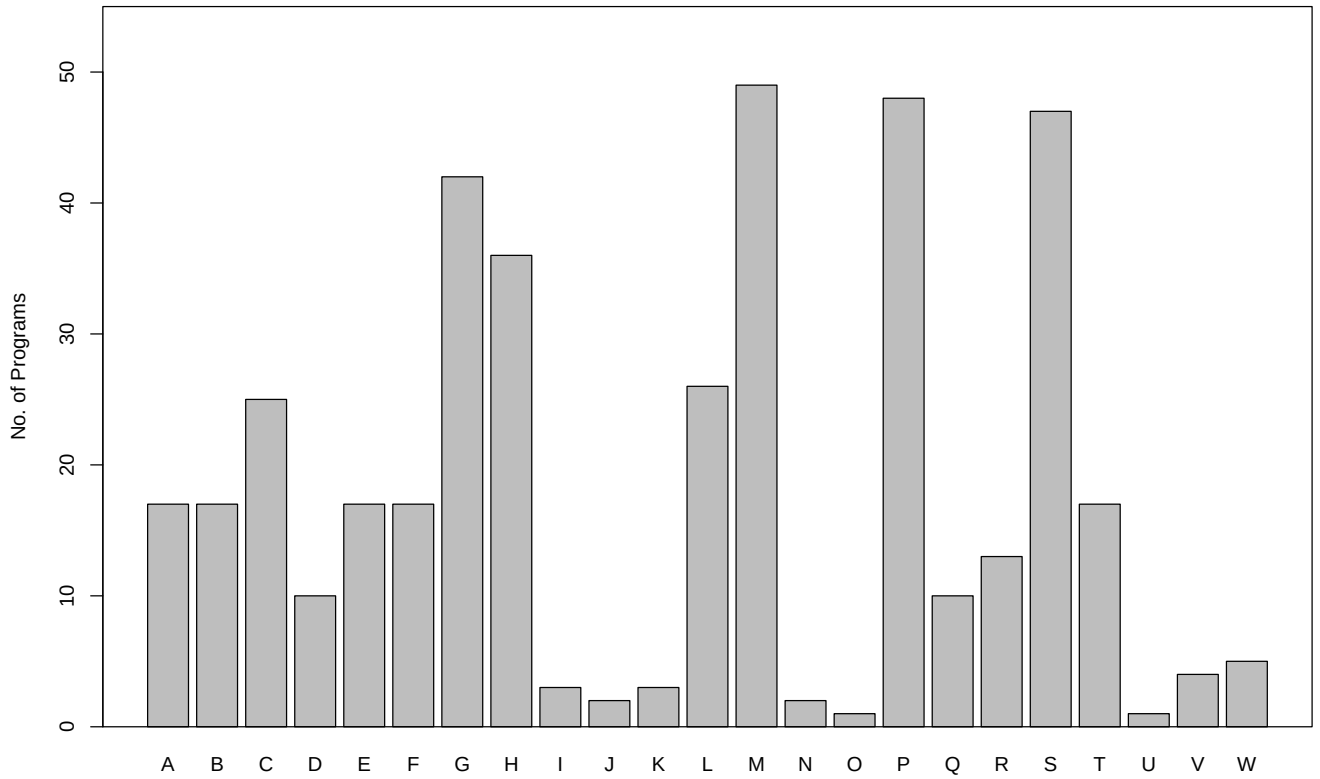
CRAN packages:

- ape
- apTreeshape
- Biodem
- bqtl
- gap (core)
- Geneland
- genetics (core)
- hapassoc
- haplo.ccs
- haplo.stats (core)
- hapsim
- hierfstat
- hwde
- kinship
- ldDesign
- LDheatmap
- Malmig
- mapLD
- multtest
- ouch
- PHYLOGR
- popgen
- powerpkg
- qtl
- qtlDesign
- qvalue
- rmetasim
- stepwise
- tdthap

Related links:

- R Software by David Clayton
- BayesMendel, R software for predicting who may carry a cancer susceptibility gene
- R genetics project
- R code for estimating haplotype frequencies from pooled DNA samples
- Computer programs by Jing Hua Zhao (some are included in the gap package)
- An alphabetic list of genetic analysis software
- Hclust.R, software for Choosing Tag SNPs
- Bayesian Adaptive Regression Splines (BARS)
- Direct Simulation Approach (DAS), for simulation of p-values
- Vector Generalized Linear and Additive Models (VGLMs/VGAMs)

A survey of available software: results



QIMR

What we use a lot

- Basic manipulation of genetic data
- Genetic data error checking
- Pictures of genetic marker data
- Drawing pedigrees
- Descriptives of genetic marker data
- Descriptives of familial phenotypic data
- Genetic linkage analysis
- Genetic association analysis

Mucking around with genetic data

- Standard statistical packages (R, SPSS, SAS, Stata)
- Standard databases (Oracle)
- Text oriented computer languages (awk, perl, python …)

For genotypes, nicest way is probably as **text**: “allele1 separator allele2”.

Have a missing **allele** token.

So, one way to store 500000 genotypes for an individual is as a string “ACAACG..ATTACCGG..TTACGGCC” [David Hinds]. Your 1000 cases and 1000 controls occupy 4 GB uncompressed.

Genetic data error checking

- Haploview
- Merlin
- MENDEL
- PEDCHECK
- RELPAIR
- SIMWALK2
- Sib-pair

Measurement error rates for individual genotypes vary from 1/10000 to 1/10.
If you have families, cross check family member genotypes versus each other. Test pedigree misspecifications/sample mixups with RELPAIR. Check sex-linked markers.
With unrelated individuals, can only test if the particular marker (assay) is poor: compare against population data, and test population genetic equilibria (Hardy-Weinberg, linkage etc) for your sample.

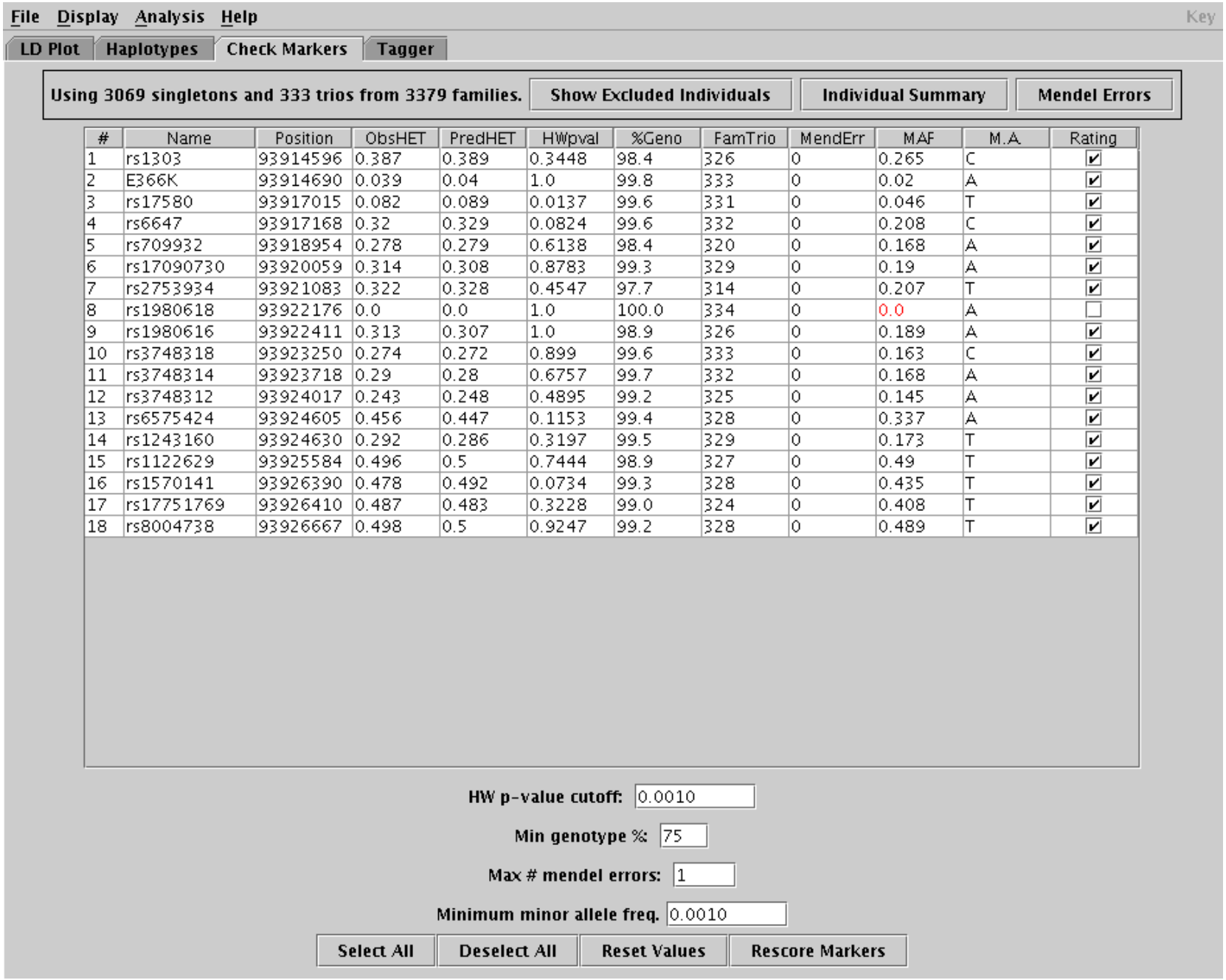

QIMR

Pictures of genetic marker data: Haploview

- For analysis of many contiguous SNPs
- Small families or unrelated cases and controls

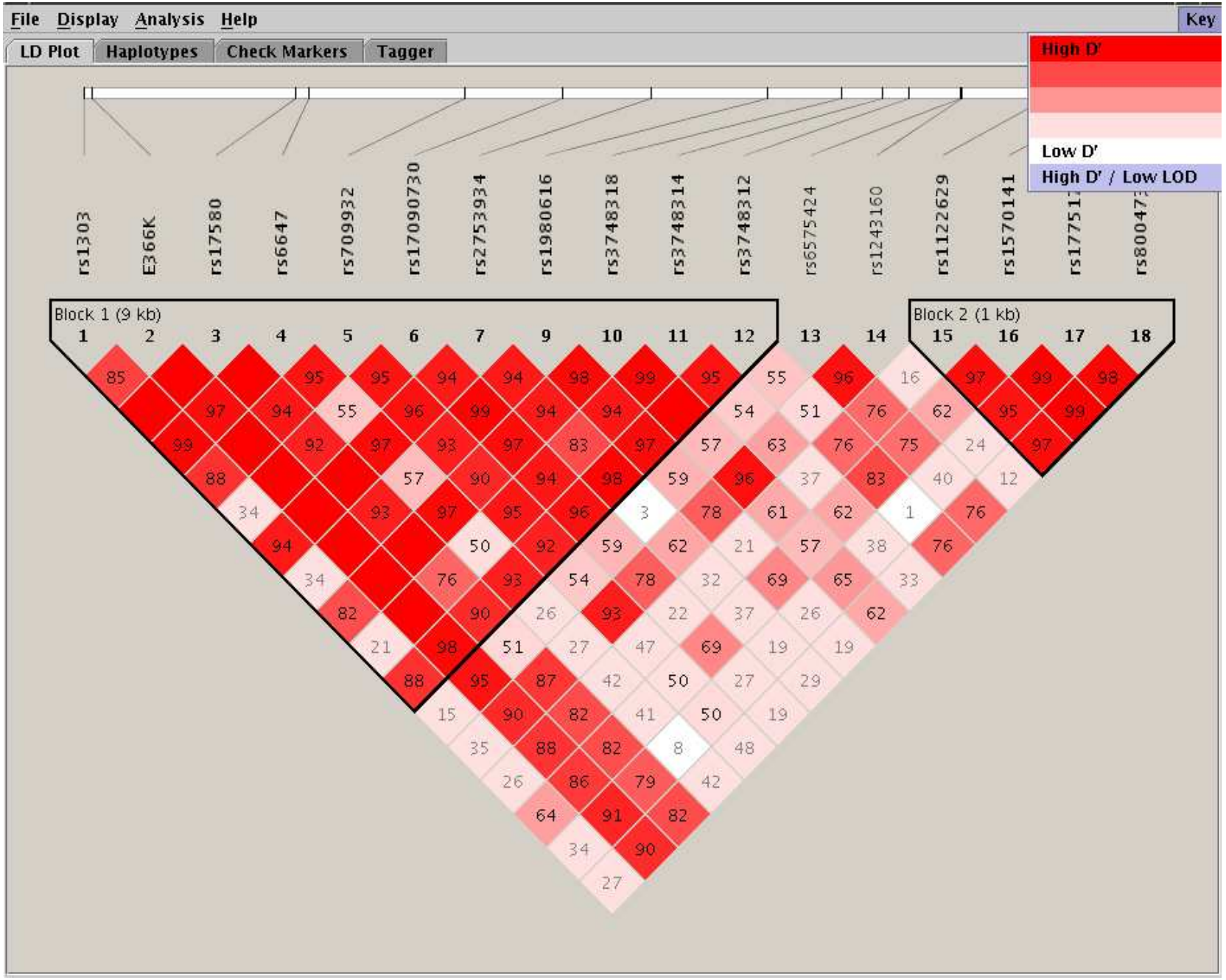
Performs:

- Error checking
- Presentation of intermarker associations (linkage disequilibrium/haplotypes)
- Selection of “tagging” SNP markers
- Global permutation testing of individual SNPs and haplotypes versus a binary trait

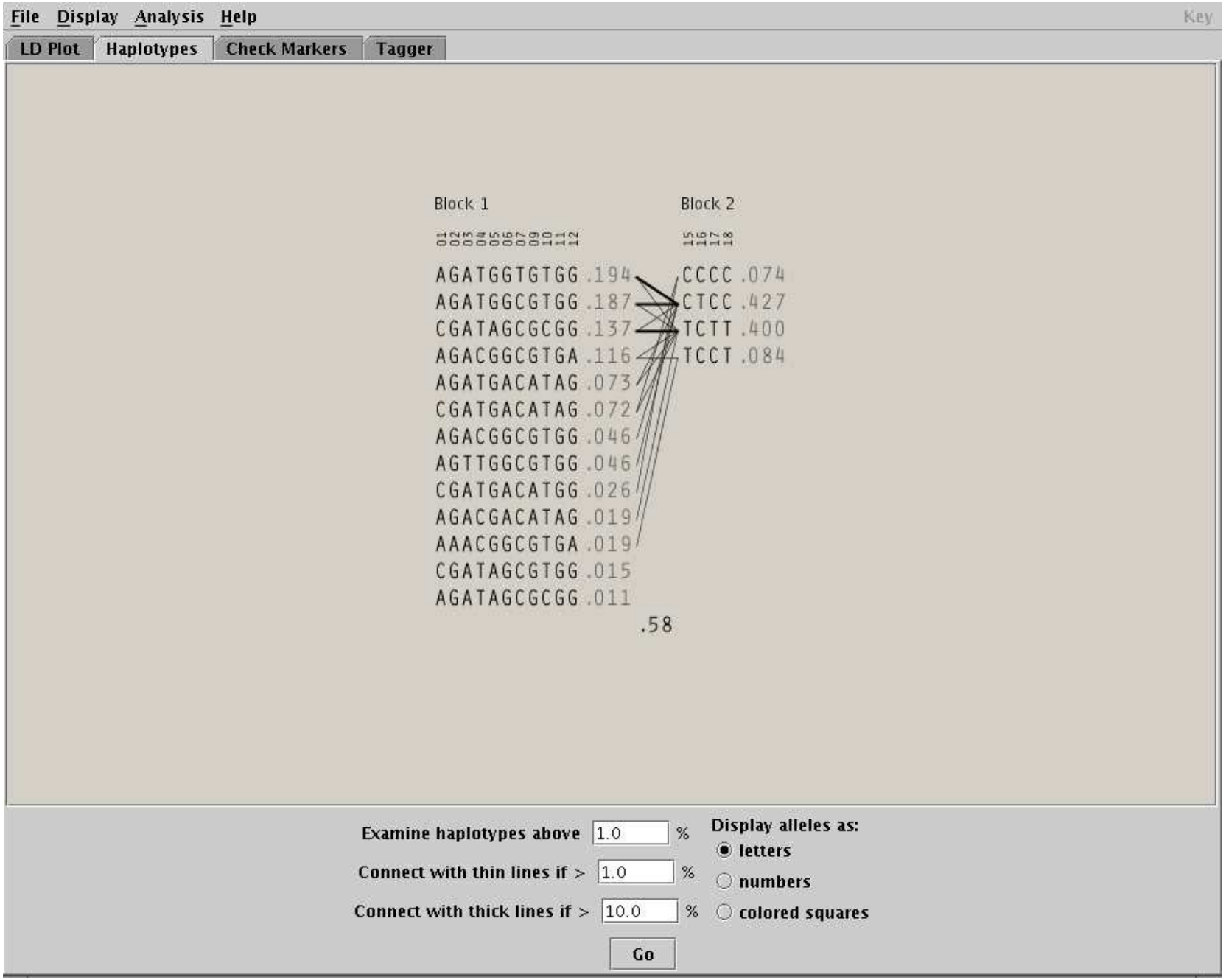


#	Name	Position	ObsHET	PredHET	HWpval	%Geno	FamTrio	MendErr	MAF	M.A.	Rating
1	rs1303	93914596	0.387	0.389	0.3448	98.4	326	0	0.265	C	☑
2	E366K	93914690	0.039	0.04	1.0	99.8	333	0	0.02	A	☑
3	rs17580	93917015	0.082	0.089	0.0137	99.6	331	0	0.046	T	☑
4	rs6647	93917168	0.32	0.329	0.0824	99.6	332	0	0.208	C	☑
5	rs709932	93918954	0.278	0.279	0.6138	98.4	320	0	0.168	A	☑
6	rs17090730	93920059	0.314	0.308	0.8783	99.3	329	0	0.19	A	☑
7	rs2753934	93921083	0.322	0.328	0.4547	97.7	314	0	0.207	T	☑
8	rs1980618	93922176	0.0	0.0	1.0	100.0	334	0	0.0	A	☐
9	rs1980616	93922411	0.313	0.307	1.0	98.9	326	0	0.189	A	☑
10	rs3748318	93923250	0.274	0.272	0.899	99.6	333	0	0.163	C	☑
11	rs3748314	93923718	0.29	0.28	0.6757	99.7	332	0	0.168	A	☑
12	rs3748312	93924017	0.243	0.248	0.4895	99.2	325	0	0.145	A	☑
13	rs6575424	93924605	0.456	0.447	0.1153	99.4	328	0	0.337	A	☑
14	rs1243160	93924630	0.292	0.286	0.3197	99.5	329	0	0.173	T	☑
15	rs1122629	93925584	0.496	0.5	0.7444	98.9	327	0	0.49	T	☑
16	rs1570141	93926390	0.478	0.492	0.0734	99.3	328	0	0.435	T	☑
17	rs17751769	93926410	0.487	0.483	0.3228	99.0	324	0	0.408	T	☑
18	rs8004738	93926667	0.498	0.5	0.9247	99.2	328	0	0.489	T	☑

QIMR


File Display Analysis Help
Key
LD Plot
Haplotypes
Check Markers
Tagger
High D'
Low D'
High D' / Low LOD
rs1303
E366K
rs17580
rs6647
rs709932
rs17090730
rs2753934
rs1980616
rs3748318
rs3748314
rs3748312
rs6575424
rs1243160
rs1122629
rs1570141
Block 1 (9 kb)
Block 2 (1 kb)
1 2 3 4 5 6 7 9 10 11 12 13 14 15 16 17 18

QIMR


File Display Analysis Help
Key
LD Plot
Haplotypes
Check Markers
Tagger
Block 1
Block 2
01 02 03 04 05 06 07 09 10 11 12
15 16 17 18
AGATGGTGTGG .194
AGATGGCGTGG .187
CGATAGCGCGG .137
AGACGGCGTGA .116
AGATGACATAG .073
CGATGACATAG .072
AGACGGCGTGG .046
AGTTGGCGTGG .046
CGATGACATGG .026
AGACGACATAG .019
AAACGGCGTGA .019
CGATAGCGTGG .015
AGATAGCGCGG .011
.58
CCCC .074
CTCC .427
TCTT .400
TCCT .084
Examine haplotypes above 1.0 %
Connect with thin lines if > 1.0 %
Connect with thick lines if > 10.0 %
Display alleles as:
letters
numbers
colored squares
Go

QIMR

Pictures of genetic marker data: GRR

- For analysis of many markers of any type
- Families

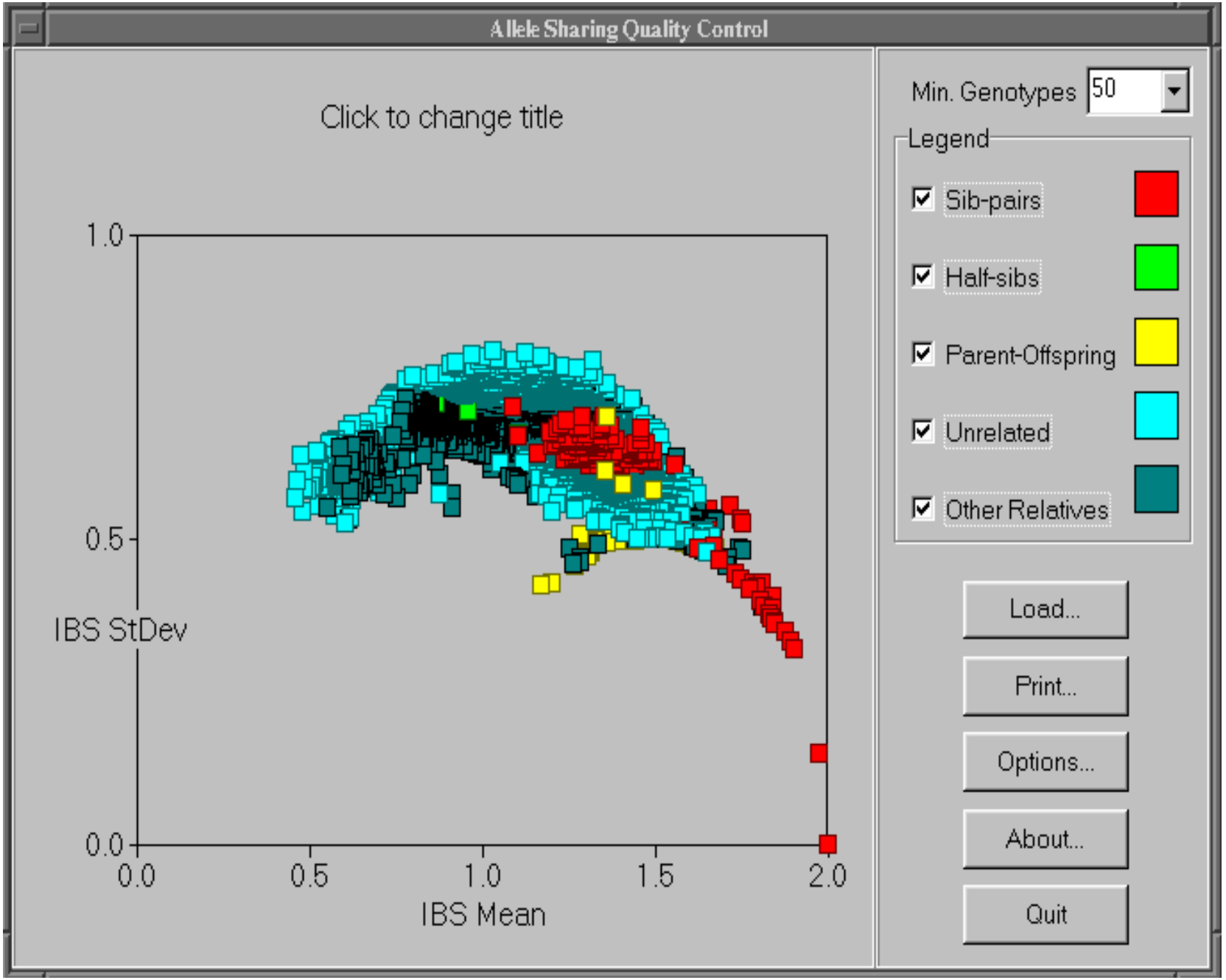
Performs:

- Graphical diagnostics for pedigree misspecification

IBS = Identity by state of genotypes at one locus in two individuals

IBS score takes value 0, 1, 2 alleles shared

Parent and offspring always share at least one allele IBS per locus


Allele Sharing Quality Control
Click to change title
1.0
0.5
0.0
IBS StDev
0.0
0.5
1.0
1.5
2.0
IBS Mean
Min. Genotypes
50
Legend
Sib-pairs
Half-sibs
Parent-Offspring
Unrelated
Other Relatives
Load...
Print...
Options...
About...
Quit

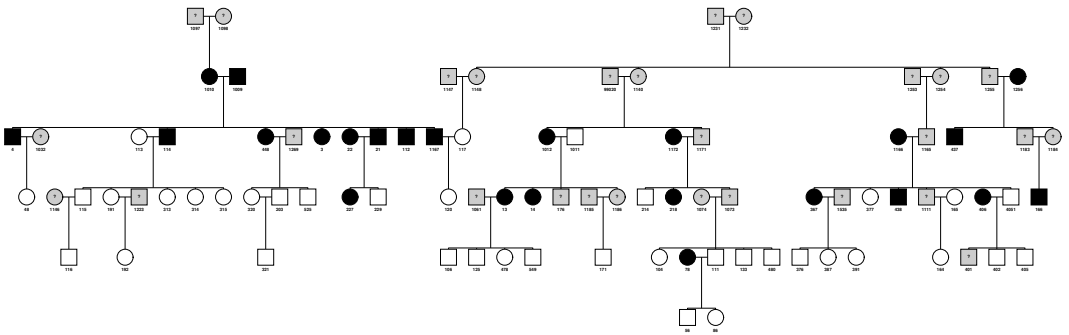
QIMR

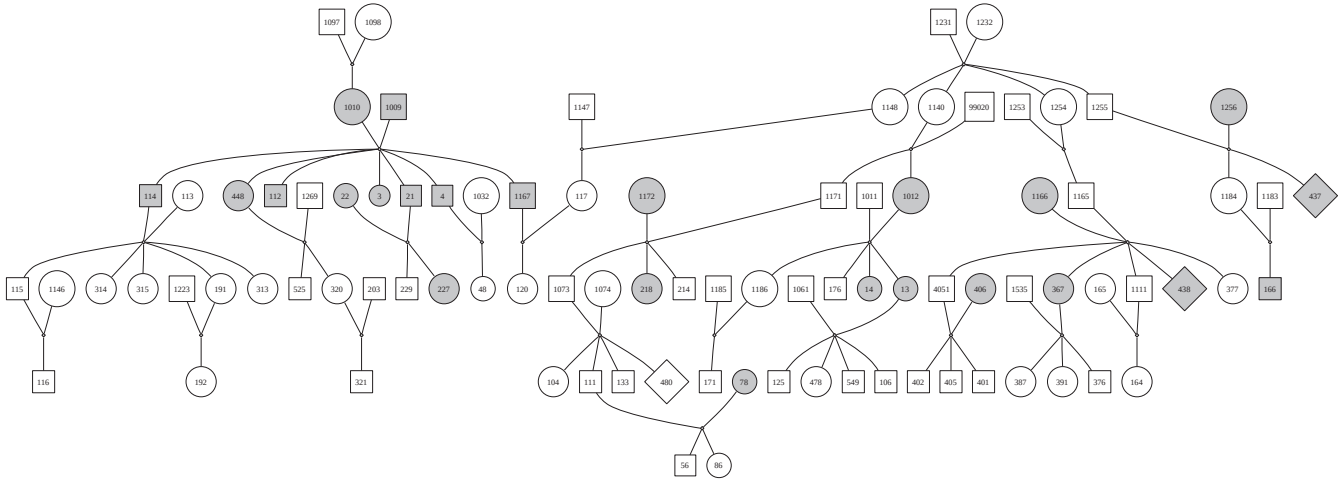
Drawing pedigrees

- Cranefoot
- Cyrillic
- Genehunter
- Graphviz
- Haplopainter
- Madeline
- Pedraw
- Pedfiddler
- Progeny

Drawing large pedigrees automatically from a data file is not done that well.

Family – S_02





Pedigree S_02

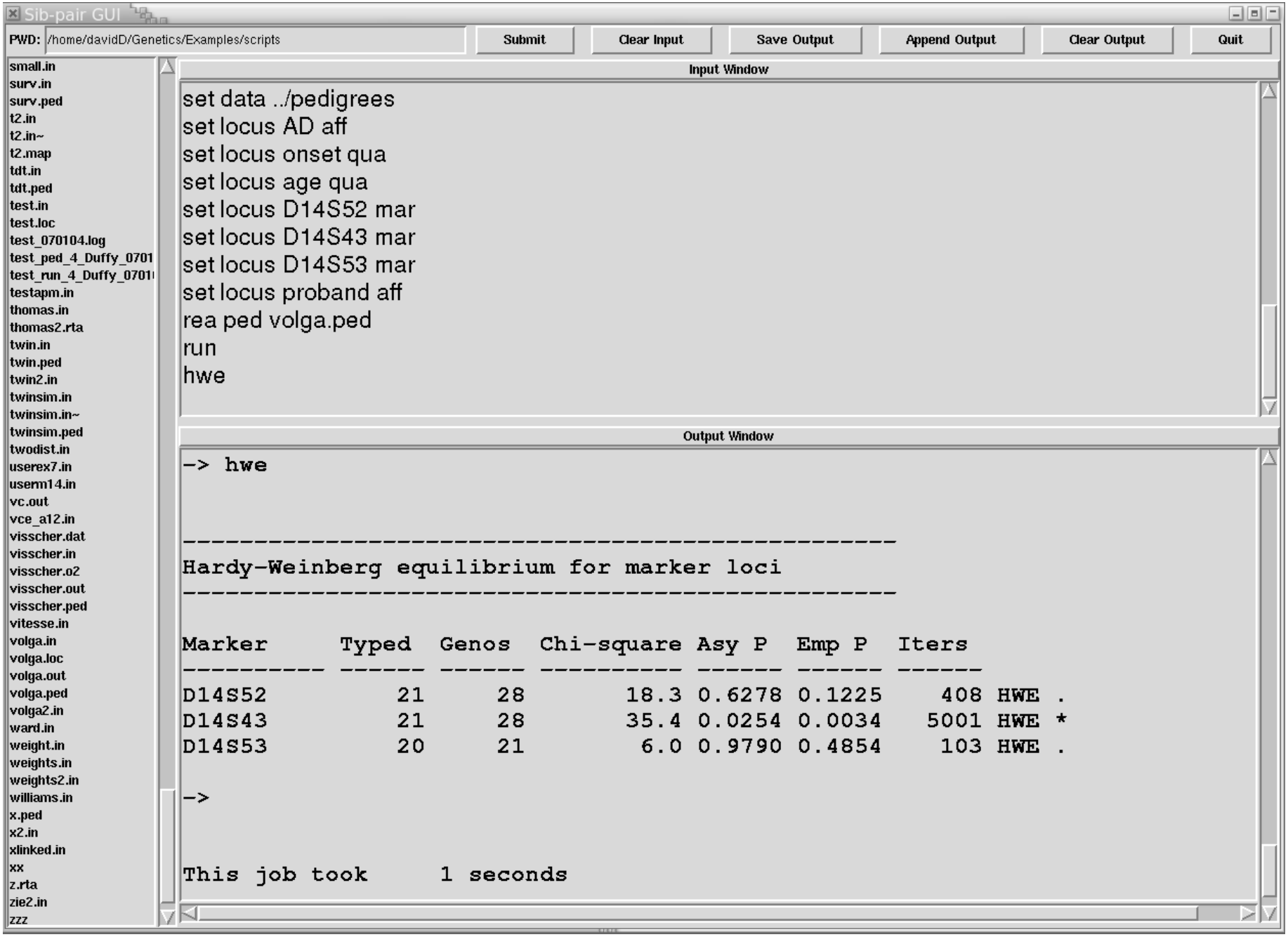
Overview of Sib-pair

First code written in 1995: for multiallelic TDT

Creeping featurism has continued to today (33000 lines of code + 6000 lines comments)

Has been run on many architectures (even a Mac version)

Version in R – lags in terms of features


Sib-pair GUI
PWD: /home/davidD/Genetics/Examples/scripts
Submit
Clear Input
Save Output
Append Output
Clear Output
Quit
small.in
surv.in
surv.ped
t2.in
t2.in~
t2.map
tdt.in
tdt.ped
test.in
test.loc
test_070104.log
test_ped_4_Duffy_0701
test_run_4_Duffy_0701
testapm.in
thomas.in
thomas2.rta
twin.in
twin.ped
twin2.in
twinsim.in
twinsim.in~
twinsim.ped
twodist.in
userex7.in
userm14.in
vc.out
vce_a12.in
visscher.dat
visscher.in
visscher.o2
visscher.out
visscher.ped
vitesse.in
volga.in
volga.loc
volga.out
volga.ped
volga2.in
ward.in
weight.in
weights.in
weights2.in
williams.in
x.ped
x2.in
xlinked.in
xx
z.rta
zie2.in
zzz
Input Window
set data ../pedigrees
set locus AD aff
set locus onset qua
set locus age qua
set locus D14S52 mar
set locus D14S43 mar
set locus D14S53 mar
set locus proband aff
rea ped volga.ped
run
hwe
Output Window
-> hwe
Hardy-Weinberg equilibrium for marker loci
Marker Typed Genos Chi-square Asy P Emp P Iters
D14S52 21 28 18.3 0.6278 0.1225 408 HWE .
D14S43 21 28 35.4 0.0254 0.0034 5001 HWE *
D14S53 20 21 6.0 0.9790 0.4854 103 HWE .
->
This job took 1 seconds

QIMR

Basic use of Sib-pair: data entry

Getting some data in:

```
read locus merlin chrom19.dat
read map chrom19.map
read pedigree chrom19.ped
run
```

```
#          Locus     Type    cM       ——- Comments ———
set locus melanoma   aff       .      Melanoma case (y) or control (n)
set locus molecount  quant     .      Flat mole count
set locus rs719010   marker  63.2766  63276577  G>A bad locus?
set locus rs1549492  marker  63.2771  63277060  T>C
set locus rs260454   marker  63.459   63458980  T>C
read pedigree chrom19.ped
run
```

Basic use of Sib-pair: data manipulation

Some data manipulations:

```
drop where monomorphic
drop where “bad“
flip rs15*
select containing 2 where \
       isnon and melanoma and (numtyp == 3)
molecount = log (molecount+1)
write structure c19.data
write locus structure c19.param c19.data
drop $q $a
write arlequin c19.arl parents
```

Basic genetic analyses using Sib-pair

```
list $m
show pedigrees
describe $q
describe snps
hwe
hwe founders
diseq
table cmm rs260*
table cmm 5
tdt cmm
assoc cmm covariates molecount
assoc cmm genotypes
homoz cmm
varcomp molecount covariates cmm
qtl molecount full
```

Descriptives of genetic marker data: Sib-pair

Many programs (including Haploview) can produce a compact summary describing a set of genotypes at multiple markers. In Sib-pair,

```
describe snps
```

obtains

Marker	NAll	Allele(s)		Freq	Het	Ntyped	HWE-P
E366K	2	A	(G)	0.0197	0.0387	4887	0.7147
rs1122629	2	T	(C)	0.4964	0.5000	4837	0.6872
rs1243160	2	T	(C)	0.1760	0.2901	4871	0.1519
rs1303	2	C	(A)	0.2676	0.3921	4820	0.2546
rs1570141	2	T	(C)	0.4297	0.4902	4865	0.1880
rs17090730	2	A	(G)	0.1963	0.3155	4858	0.3875
rs17580	2	T	(A)	0.0454	0.0867	4880	0.0299
rs17751769	2	T	(C)	0.4153	0.4857	4854	0.9529
rs1980616	2	A	(G)	0.1962	0.3154	4847	0.6488
rs1980618	1	A		1.0000	-	4912	1.0000
rs2753934	2	T	(C)	0.2023	0.3228	4784	0.9286
rs3748312	2	A	(G)	0.1465	0.2502	4862	0.0585
rs3748314	2	A	(G)	0.1795	0.2946	4889	0.7708
rs3748318	2	C	(T)	0.1623	0.2719	4881	1.0000
rs6575424	2	A	(G)	0.3352	0.4457	4872	0.0826
rs6647	2	C	(T)	0.2063	0.3276	4878	0.1370
rs709932	2	A	(G)	0.1686	0.2803	4823	0.2585
rs8004738	2	T	(C)	0.4960	0.5000	4852	0.7522

Descriptives of familial phenotypic data: Sib-pair

——— Familial correlations (pairwise) ———

Rel 1	Rel 2	Std Dev 1	Std Dev 2	Correlation	N Pairs
Husband	Wife	5.2300	4.5710	0.0437	2299
Gparent	Gchild	5.5547	5.2041	0.1016	6594
Halfsib	Hsib	5.2550		0.2324	42226
Parent	Off	5.2429	5.3769	0.1581	5141
Fullsib	Fsib	5.1561		0.4399	211
Father	Son	5.1062	4.9018	-0.0196	97
Father	Dau	5.2690	4.3482	0.2433	491
Mother	Son	4.9589	5.2384	0.1495	95
Mother	Dau	4.8222	4.3857	0.1819	468
Brothers		0.0000		0.0000	0
Sisters		5.4283		0.8507	3
Brother-Sister		0.0000	0.0000	0.0000	1

QIMR

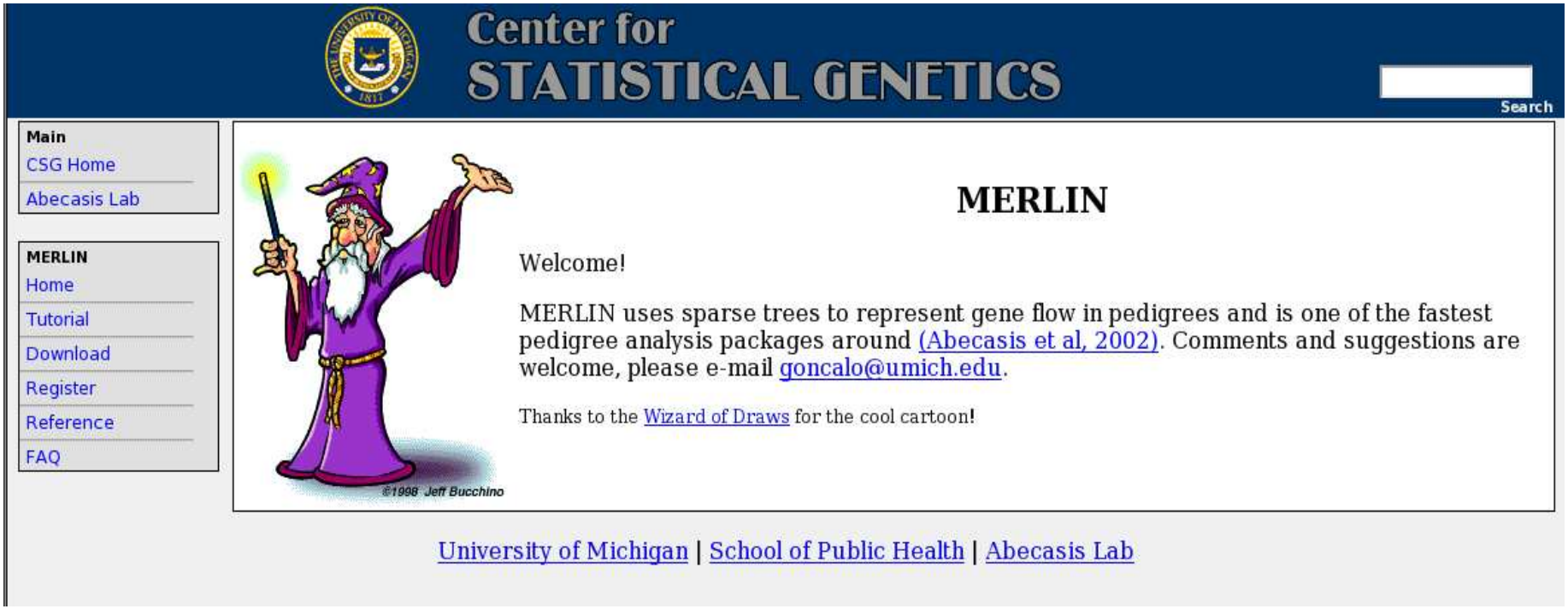
Segregation ratios for trait "sga"

Total sample	All	Fndrs	Nonfndrs
Aff/Tot	306/3414	0/ 0	306/3414
Prop Aff	0.090	0.000	0.090
Missing	618	607	11

Mating Type	UxU	UxA	AxA
Matings	2093	204	2
Aff/Tot	188/2207	32/ 212	1/ 3
Prop Aff	0.085	0.151	0.333

Relative pair	RecRisk	Aff-Aff	Aff-UnA
Marital	0.019	2	204
Gparent	0.077	31	744
Halfsib	0.183	611	5470
Par-Off	0.100	34	613
Fullsib	0.541	10	17

Genetic linkage analysis: MERLIN



506 citations since 2002.

- For linkage analysis of binary or quantitative traits and many markers
- Small to moderately large families

Performs:

- parametric and non-parametric linkage analysis
- variance components linkage analysis of quantitative traits
- regression-based analysis of quantitative traits
- multimarker-based ibd and kinship estimation
- haplotyping
- error detection
- simulation of marker data under null hypothesis of no linkage

Linkage disequilibrium between markers is allowed in models

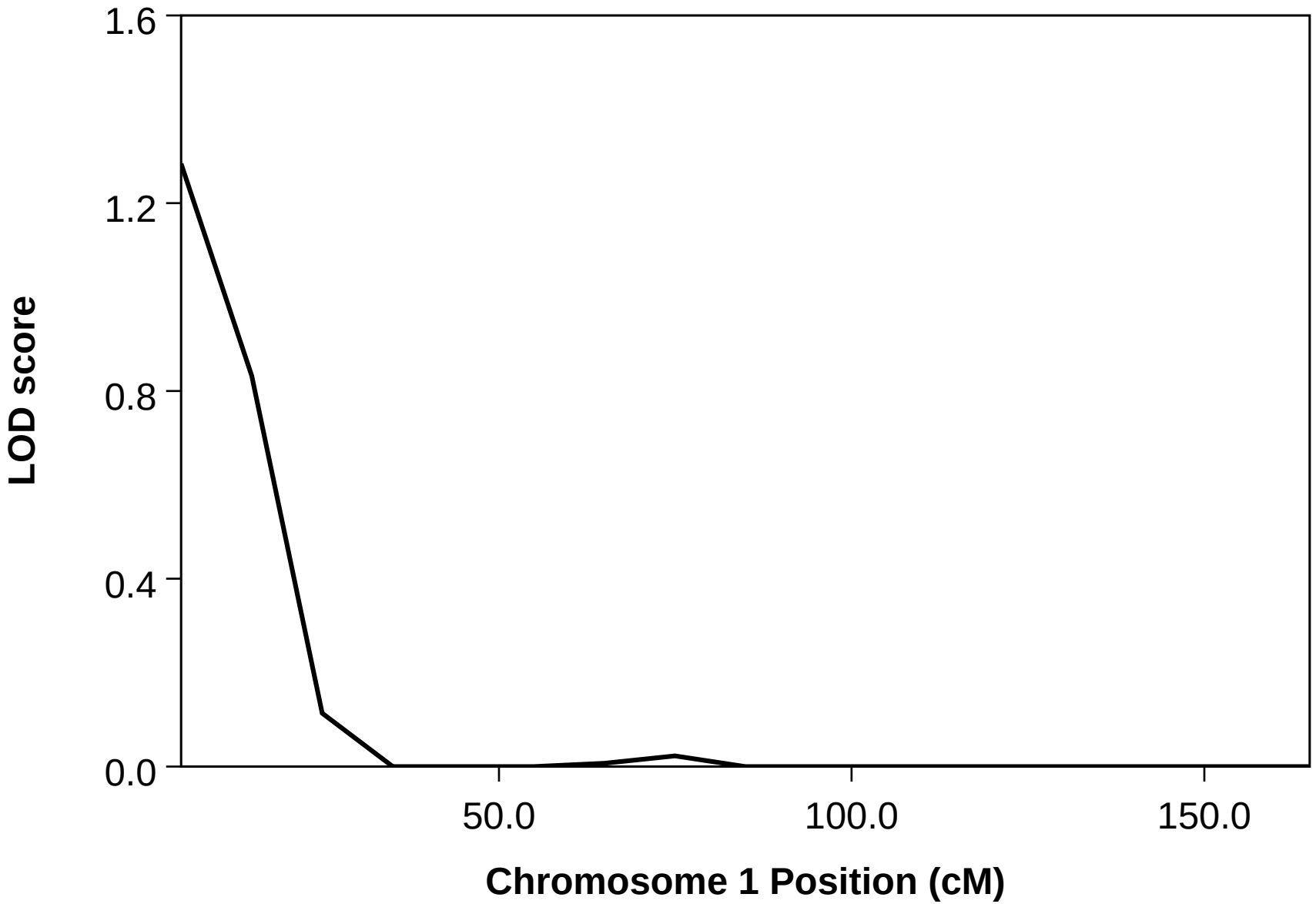
```
> merlin -d chr1.dat -m chr1.map -p chr1.ped –grid 10 –vc –pdf

…
Family: 42258 - Founders: 2  - Descendants: 2  - Bits: 2
  Skipping Marker D11S2008_S [BAD INHERITANCE]
  Skipping Marker ATA27C11_M [BAD INHERITANCE]

Family: 99008 - Founders: 2  - Descendants: 2  - Bits: 2
  Skipping Marker D11S1301_S [BAD INHERITANCE]

Phenotype: ige1 [VC] (773 families, h2 = 48.46%)
============================================================
            Position      H2    ChiSq     LOD  pvalue
               4.939   37.76%    5.91    1.28   0.008
              14.939   31.25%    3.83    0.83    0.03
              24.939   10.59%    0.52    0.11     0.2
              34.939    0.00%    0.00    0.00     0.5
              44.939    0.00%    0.00    0.00     0.5
              54.939    0.00%    0.00    0.00     0.5
              64.939    1.51%    0.03    0.01     0.4
              74.939    3.38%    0.10    0.02     0.4
              84.939    0.00%    0.00    0.00     0.5
```

QIMR


ige1 [VC]
1.6
1.2
0.8
0.4
0.0
LOD score
50.0
100.0
150.0
Chromosome 1 Position (cM)

Linkage and association: MENDEL

36 citations since 2001.

- For linkage and association analysis
- Small to large families
- Larger families can be analysed with a companion program SIMWALK2

Performs a wide variety of analyses. I would highlight:

- GLM (including censoring) haplotype association analysis for pedigrees, with LD
- multivariate variance components linkage analysis
- association conditional on linkage (in pedigrees)

Table 0.1 from the MENDEL manual

#	**Analysis Option**	#	**Analysis Option**
1	MAPPING_MARKERS	12	CASES_AND_CONTROLS
2	LOCATION_SCORES	13	TDT
3	HAPLOTYPING	14	PENETRANCES
4	NPL	15	ETHNIC_ADMIXTURE
5	MISTYPING	16	COMBINING_ALLELES
6	ALLELE_FREQUENCIES	17	GENE_DROPPING
7	GENETIC_COUNSELING	18	COMBINING_SNPS
8	GAMETE_COMPETITION	19	POLYGENIC_QTL
9	PEDIGREE_SELECTION	20	QTL_ASSOCIATION
10	KINSHIP_MATRICES	21	TRIM_PEDIGREES
11	GENETIC_EQUILIBRIUM	22	ASSOCIATION_GIVEN_LINKAGE

Some sample output:

```
…
 FIRST LOCUS NAME          : snp17505
 LAST LOCUS NAME           : snp19827
 FISHER P-VALUE            : 0.0010000 +/- 0.0006321
 ZMAX   P-VALUE            : 0.0530000 +/- 0.0044807
 MOST ABERRANT TYPE        : A/A,G/G,A/A,C/C,C/C,C/C,T/T,C/C,C/C,G/G
 AFFECTED SAMPLE SIZE      :    17
 NORMAL   SAMPLE SIZE      :    16

 FIRST LOCUS NAME          : snp17652
 LAST LOCUS NAME           : snp20189
 FISHER P-VALUE            : 0.0006000 +/- 0.0004898
 ZMAX   P-VALUE            : 0.0158000 +/- 0.0024940
 MOST ABERRANT TYPE        : G/G,A/A,C/C,C/C,C/C,T/T,C/C,C/C,G/G,G/G
 AFFECTED SAMPLE SIZE      :    17
 NORMAL   SAMPLE SIZE      :    17
```

Analysis of the same data using Haploview:

- Global permutation P-value
- No covariates possible
- No facility for sliding a window

File Display Analysis Help Key

LD Plot | Haplotypes | Check Markers | Tagger | Association

Single Marker | Haplotypes | Permutation Tests

- ○ Single Markers Only
- ◉ Single Markers and Haplotypes in Blocks Number of Permutations To Perform: 10000
- ○ Haplotypes in Blocks Only

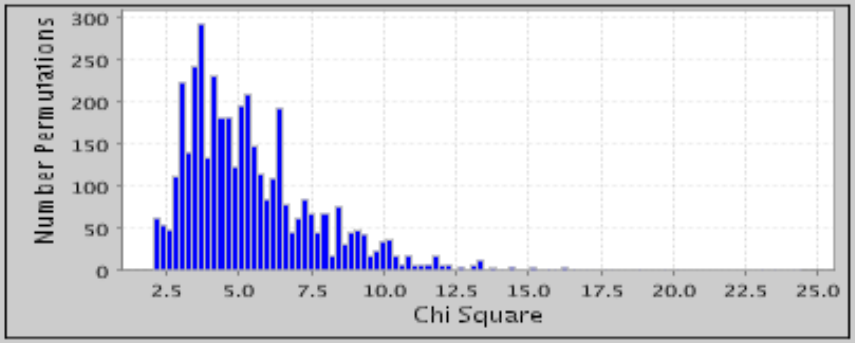
Do Permutations Stop

Best Observed Chi-Square: 13.183 (snp17064)

Best Permutation Chi-Square: 24.395

54 permutations out of 4063 exceed highest observed chi square.

Name	Chi Square	Permutation p-value
snp17064	13.183	0.0133
snp12946	12.807	0.0143
snp5206	10.968	0.0298
snp3586	9.743	0.0603
Block 1: GTCACGCTAA..	9.476	0.0652
snp5859	8.762	0.0982
snp6149	8.571	0.1071
snp19550	8.42	0.1120
Block 1: GTCACGCTAA	8.285	0.1154



QIMR

MCMC GLMMs in Sib-pair: link functions

The trait model can be:

- Gaussian
- Binomial with identity, probit or logit link
- Multifactorial threshold model
- Poisson (log link)
- Weibull

MCMC GLMMs in Sib-pair: latent variables

In the Sib-pair MCMC GLMM, the simulated (unobserved) variables include:

- Diallelic QTL genotypes
- Gaussian breeding values
- Maternal effect (“permanent environment”) values
- Family environmental effect values
- A single QTL allele frequency (shared by all QTLs in the major gene or finite polygenic model)
- Up to three genotypic means (shared by all QTLs in the FPM)
- V_A, V_C, V_M

MCMC GLMMs in Sib-pair: commands

Contingency table analysis	*table*
GLMM polygenic models	*fpm nqtl 0*
GLMM segregation models	*fpm nqtl 1*
Finite polygenic models	*fpm nqtl N*

MCMC GLMMs in Sib-pair: a simple genetic example

Binomial GLMM analysis of rat toxicology dataset of Weil et al (1972) using different approaches. PQL1 is the penalised quasilikelihood approach implemented as glmmPQL() in the MASS package [Venables and Ripley 2002], while PQL2, Laplace are results from lmer() in the lme4 package of Bates and Sarkar [2005] using penalized quasilikelihood, Laplace approximation respectively. The SAS results are from Wang and Louis [2002].

	Parameter Estimate (SE)				
Method	Sib-pair	Laplace	PQL1	PQL2	SAS NLMIXED
SD Litter RE	1.36 (0.34)	1.30	1.27	1.49	1.34 (0.33)
Intercept	2.58 (0.49)	2.63 (0.45)	2.37 (0.41)	2.37 (0.48)	2.62 (0.48)
Treatment	-1.07 (0.65)	-1.09 (0.60)	-0.96 (0.56)	-0.96 (0.66)	-1.07 (0.62)

MCMC GLMMs in Sib-pair: a nongenetic example

Poisson GLMM analysis of European male melanoma death rate dataset of Langford et al (1998) using different approaches. PQL1 is the penalised quasilikelihood approach implemented as glmmPQL() by Ripley and Venables [2002] in the MASS package, while PQL2, AGQ are results from lmer() in the lme4 package of Bates and Sarkar (2005). The STATA result used the xtpois command, and comes from the review article at *http://www.mlwin.com/softrev/revstata.html.*

	Parameter Estimate (SE)				
Method	Sib-pair	AGQ	PQL1	PQL2	STATA
Region variance	0.188 (0.037)	0.170 (-)	0.161	0.125	0.102 (-)
Intercept	-0.151 (0.058)	-0.139 (0.043)	-0.129 (0.049)	-0.129 (0.043)	-0.138 (0.017)
UVB insolation	-0.035 (0.011)	-0.034 (0.009)	-0.038 (0.010)	-0.038 (0.009)	-0.056 (0.004)

Last thoughts

- Lots of different tools
- Use many, and crosscheck results from different types of analysis and program

A few more gems:

- In R, the *kinship* and *haplo.stats* packages
- FBAT, QTDT and Unphased for within-family association analysis
- WOMBAT and ASREML, for *very* fast (AI-REML) variance components (linkage) analysis