

# Linkage mapping R/QTL

Leo Schalkwyk

June 15, 2011

## qtl package by Karl Broman

- ▶ mapping quantitative trait loci in experimental crosses
- ▶ goal is to make complex QTL mapping methods widely accessible
- ▶ allow users to focus on modeling rather than computing
- ▶ hidden Markov model (HMM) for dealing with missing genotype data
- ▶ map estimation, identifying errors
- ▶ single and pairwise QTL scans

# QTL mapping methods

- ▶ interval mapping (EM algorithm)
- ▶ Hayley-Knott regression
- ▶ multiple imputation
- ▶ MQM (R. Jansen)

# association models

- ▶ 'normal' analogous to ANOVA
  - ▶ Lander and Botstein 1989
- ▶ 'binary' analogous to logistic regression
  - ▶ Xu and Atchley 1996
- ▶ 'two-part' analogous to survival
  - ▶ Boyartchuk et al. 2001
- ▶ 'nonparametric' modified Kruskal-Wallace
  - ▶ resembles Kruglyak and Lander 1995

# applications

- ▶ F2 crosses
- ▶ backcrosses
- ▶ recombinant inbreds
- ▶ advanced intercrosses

# data input formats

- ▶ `read.cross()`
- ▶ csv file with a row for each individual
  - ▶ one or more phenotype columns at beginning
  - ▶ genotype columns following
  - ▶ header row with phenotype/marker names
  - ▶ optional row of positions for markers
- ▶ multiple variants are possible including
  - ▶ rotated csv
  - ▶ separate csv for genotypes and phenotypes
  - ▶ mapmaker, map manager, qtl cartographer

## cross object

- ▶ `data(hyper)`
- ▶ `str(hyper)`
- ▶ list of 2
  - ▶ `geno`
  - ▶ `pheno`
- ▶ `pheno` is a data frame of phenotypes
  - ▶ one or more phenotype columns
  - ▶ `sex`
  - ▶ `pgm` (direction of cross, phase)
- ▶ `geno` is a list of chromosomes
  - ▶ `data` is a matrix of integers
  - ▶ `map` is a vector of positions
- ▶ de novo maps can be generated

## viewing and sanity checking data: genotypes

```
summary(hyper)
plot(hyper)
plot.missing(hyper)
plot.missing(hyper,1)
geno.image(hyper)
plot.rf(est.rf(hyper))
plot.geno(hyper)
```



## viewing and sanity checking data: phenotypes

```
summary(hyper$pheno)
plot(density(hyper$pheno$bp))
qqnorm(hyper$pheno$bp)
```

## calculating and displaying QTL maps

```
hyper <- calc.genoprob(hyper)
out.em <- scanone(hyper)
plot(out.em)
summary(out.em)
str(out.em)
```

## other mapping methods

```
out.hk <- scanone(hyper, method='hk')  
plot(out.hk, col='red', add=T)  
out.np <- scanone(hyper, model='np')  
plot(out.np, col='blue', add=T)
```

## permutation testing

- ▶ generate an empirical null distribution
- ▶ have to be careful about structure of data

```
out.em.perm <- scanone(hyper, n.perm=1000)
summary( out.em.perm )
summary( out.em, perms=out.em.perm , alpha=.05,pvalues=T)
```

## other interesting functions

- ▶ scantwo
  - ▶ two-QTL model
- ▶ cim
  - ▶ composite interval mapping
- ▶ mqmscan
  - ▶ multiple QTL model

## summary

- ▶ qtl package is built on a human-readable data structure
- ▶ it implements a variety of regression methods
- ▶ ... and statistical models
  - ▶ all based on  $AA \neq AB \neq BB$
  - ▶ allowing for covariates
- ▶ several multiple QTL modelling approaches