

R/qtl & R/qtlbim Tutorials

- R statistical graphics & language system
- R/qtl tutorial
 - R/qtl web site: www.rqtl.org
 - Tutorial: www.rqtl.org/tutorials/rqtltour.pdf
 - R code: www.stat.wisc.edu/~yandell/qtlbim/rqtltour.R
 - `url.show("http://www.stat.wisc.edu/~yandell/qtlbim/rqtltour.R")`
- R/qtlbim tutorial
 - R/qtlbim web site: www.qtlbim.org
 - Tutorial and R code:
 - www.stat.wisc.edu/~yandell/qtlbim/rqtlbimtour.pdf
 - www.stat.wisc.edu/~yandell/qtlbim/rqtlbimtour.R

R/qtl tutorial (www.rqtl.org)

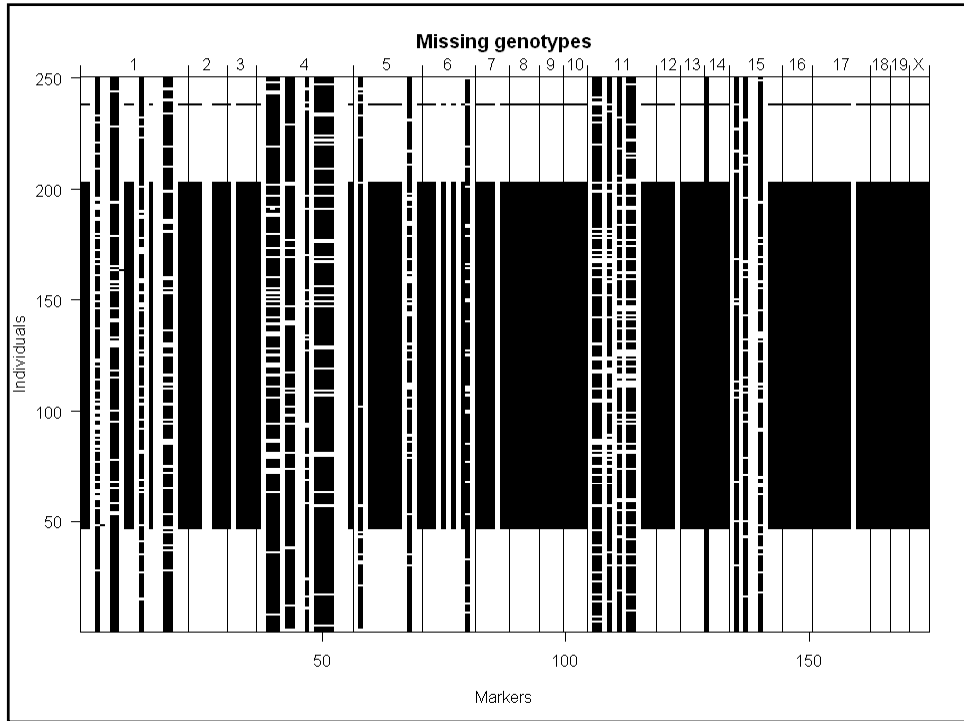
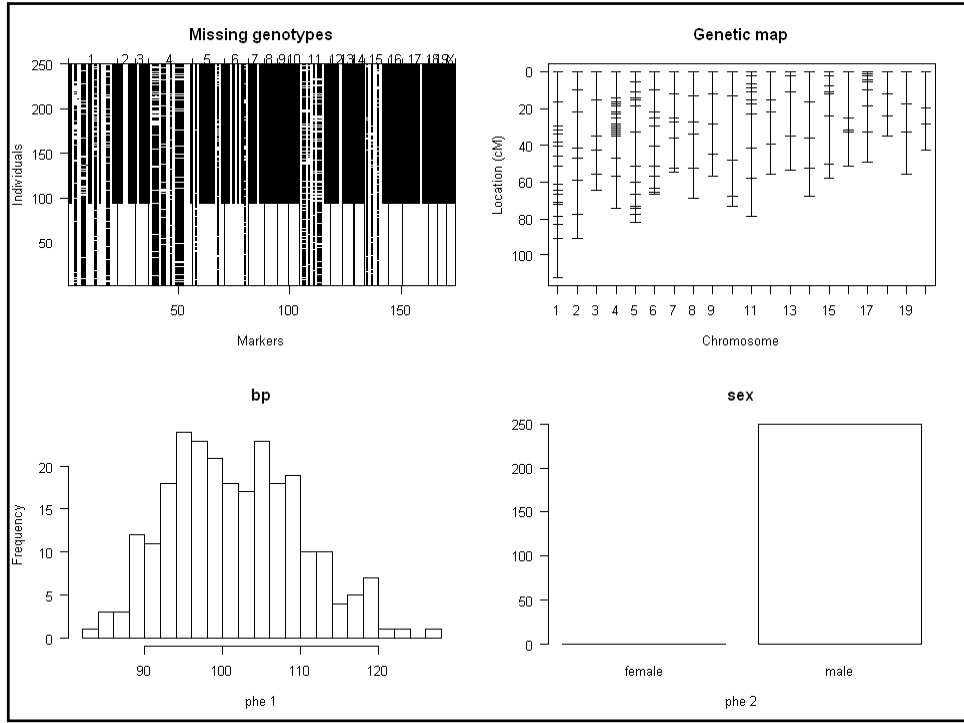
```
> library(qtl)
> data(hyper)
> summary(hyper)
  Backcross

  No. individuals: 250

  No. phenotypes: 2
  Percent phenotyped: 100 100

  No. chromosomes: 20
  Autosomes: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19
  X chr: X

  Total markers: 174
  No. markers: 22 8 6 20 14 11 7 6 5 5 14 5 5 5 11 6 12 4 4 4
  Percent genotyped: 47.7
  Genotypes (%): AA:50.2 AB:49.8
> plot(hyper)
> plot.missing(hyper, reorder = TRUE)
```



R/qtl: find genotyping errors

```
> hyper <- calc.errorlod(hyper, error.prob=0.01)
> top.errorlod(hyper)

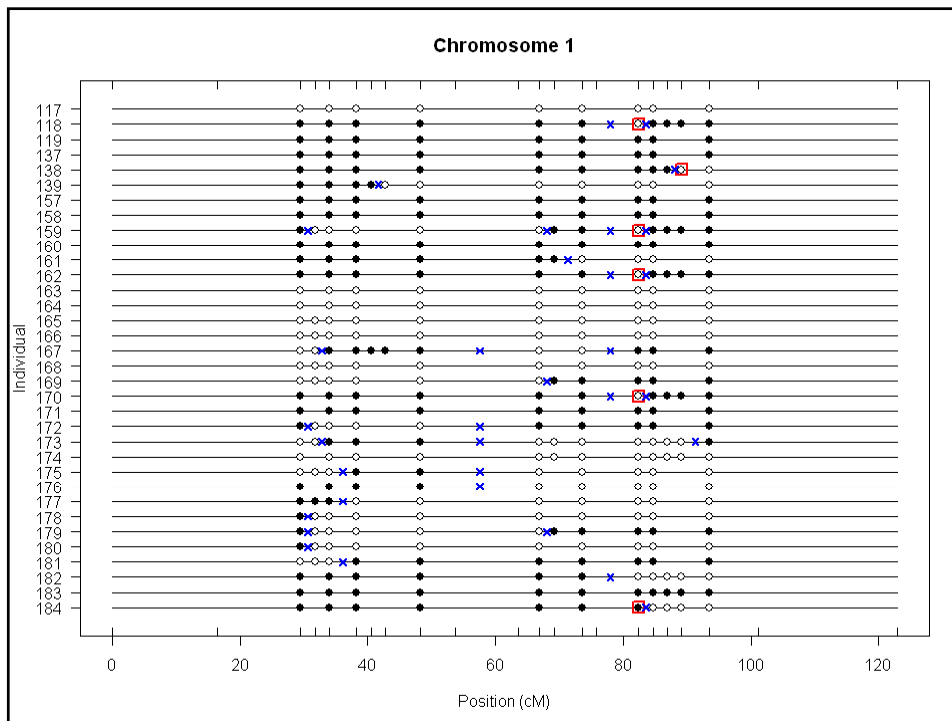
  chr id   marker errorlod
1    1 118   D1Mit14 8.372794
2    1 162   D1Mit14 8.372794
3    1 170   D1Mit14 8.372794
4    1 159   D1Mit14 8.350341
5    1  73   D1Mit14 6.165395
6    1  65   D1Mit14 6.165395
7    1  88   D1Mit14 6.165395
8    1 184   D1Mit14 6.151606
9    1 241   D1Mit14 6.151606
...
16   1 215   D1Mit267 5.822192
17   1 108   D1Mit267 5.822192
18   1 138   D1Mit267 5.822192
19   1 226   D1Mit267 5.822192
20   1 199   D1Mit267 5.819250
21   1  84   D1Mit267 5.808400

> plot.geno(hyper, chr=1, ind=c(117:119,137:139,157:184))
```

R/qtlbim Tutorial

NSF StatGen: Yandell © 2009

5



R/qtl: 1 QTL interval mapping

```
> hyper <- calc.genoprob(hyper, step=1,
  error.prob=0.01)
> out.em <- scanone(hyper)
> out.hk <- scanone(hyper, method="hk")
> summary(out.em, threshold=3)
      chr pos lod
c1.loc45  1 48.3 3.52
D4Mit164  4 29.5 8.02

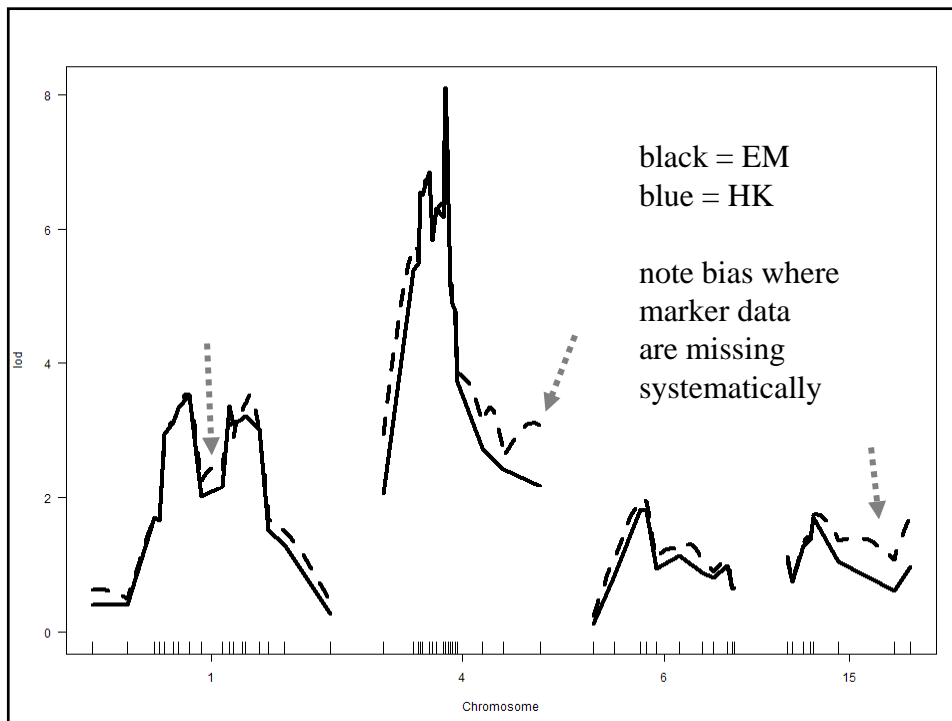
> summary(out.hk, threshold=3)
      chr pos lod
c1.loc45  1 48.3 3.55
D4Mit164  4 29.5 8.09

> plot(out.em, chr = c(1,4,6,15))
> plot(out.hk, chr = c(1,4,6,15), add = TRUE, lty = 2)
```

R/qtlbim Tutorial

NSF StatGen: Yandell © 2009

7



R/qtl: permutation threshold

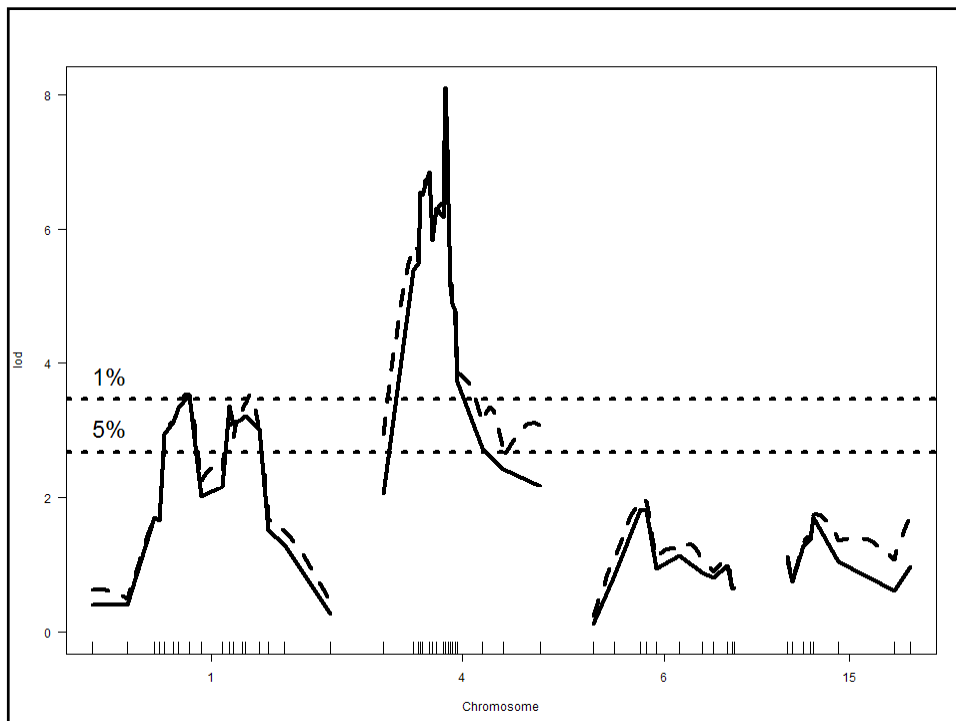
```
> operm.hk <- scanone(hyper, method="hk",
  n.perm=1000)
Doing permutation in batch mode ...
> summary(operm.hk, alpha=c(0.01,0.05))
LOD thresholds (1000 permutations)
      lod
1% 3.79
5% 2.78

> summary(out.hk, perms=operm.hk, alpha=0.05,
  pvalues=TRUE)
  chr pos lod pval
1   1 48.3 3.55 0.015
2   4 29.5 8.09 0.000
```

R/qtlbim Tutorial

NSF StatGen: Yandell © 2009

9



R/qtl: 2 QTL scan

```

> hyper <- calc.genoprob(hyper, step=5, error.prob=0.01)
>
> out2.hk <- scantwo(hyper, method="hk")
--Running scanone
--Running scantwo
(1,1)
(1,2)
...
(19,19)
(19,X)
(X,X)
> summary(out2.hk, thresholds=c(6.0, 4.7, 4.4, 4.7, 2.6))

      pos1f pos2f lod.full lod.fv1 lod.int      pos1a pos2a lod.add lod.av1
c1 :c4   68.3  30.0   14.13   6.51  0.225    68.3  30.0   13.90  6.288
c2 :c19  47.7   0.0    6.71   5.01  3.458    52.7   0.0    3.25  1.552
c3 :c3   37.2  42.2    6.10   5.08  0.226    37.2  42.2    5.87  4.853
c6 :c15  60.0  20.5    7.17   5.22  3.237    25.0  20.5    3.93  1.984
c9 :c18  67.0  37.2    6.31   4.79  4.083    67.0  12.2    2.23  0.708
c12:c19  1.1  40.0    6.48   4.79  4.090     1.1   0.0    2.39  0.697

> plot(out2.hk, chr=c(1,4,6,15))

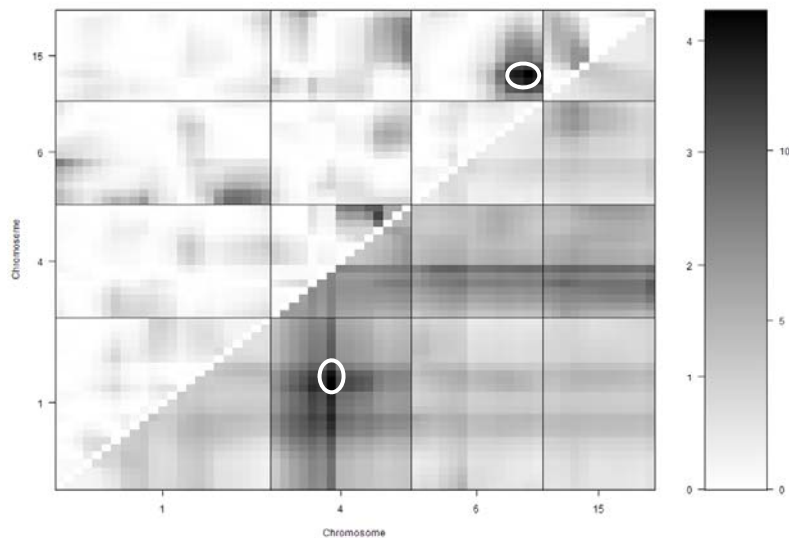
```

R/qtlbim Tutorial

NSF StatGen: Yandell © 2009

11

upper triangle/left scale: epistasis LOD
lower triangle/right scale: 2-QTL LOD



R/qtlbim Tutorial

NSF StatGen: Yandell © 2009

12

Effect & Interaction Plots

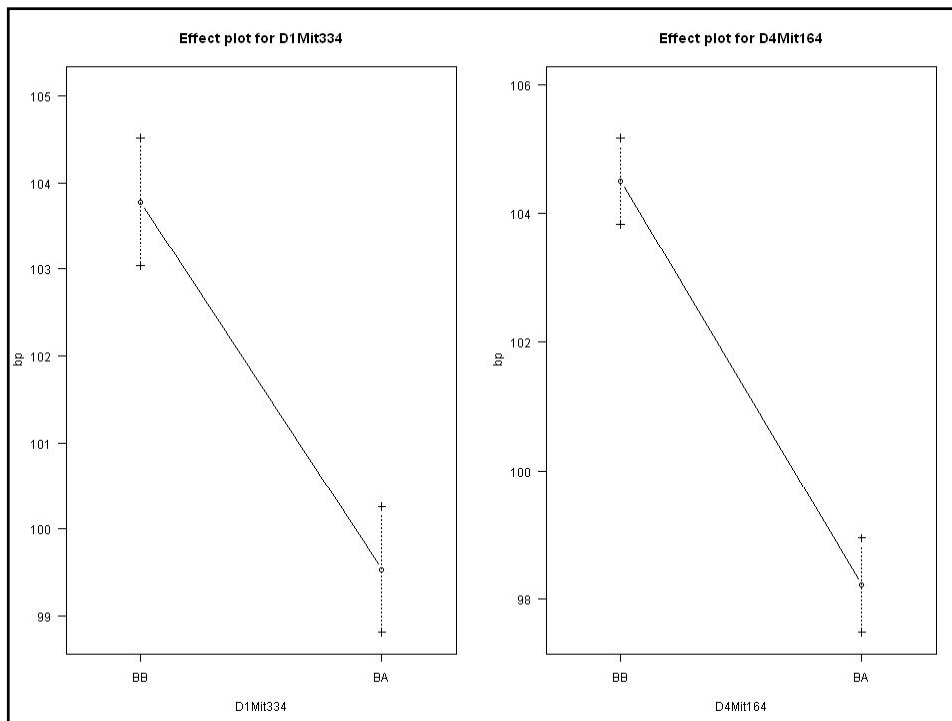
```
## Effect plots and interaction plot.
hyper <- sim.geno(hyper, step=2, n.draws=16, error.prob=0.01)
effectplot(hyper, pheno.col = 1, mname1 = "D1Mit334")
effectplot(hyper, pheno.col = 1, mname1 = "D4Mit164")
markers <- find.marker(hyper, chr = c(6,15), pos = c(70,20))
effectplot(hyper, pheno.col = 1,
           mname1 = markers[1], mname2 = markers[2])
effectplot(hyper, pheno.col = 1,
           mname1 = markers[2], mname2 = markers[1])

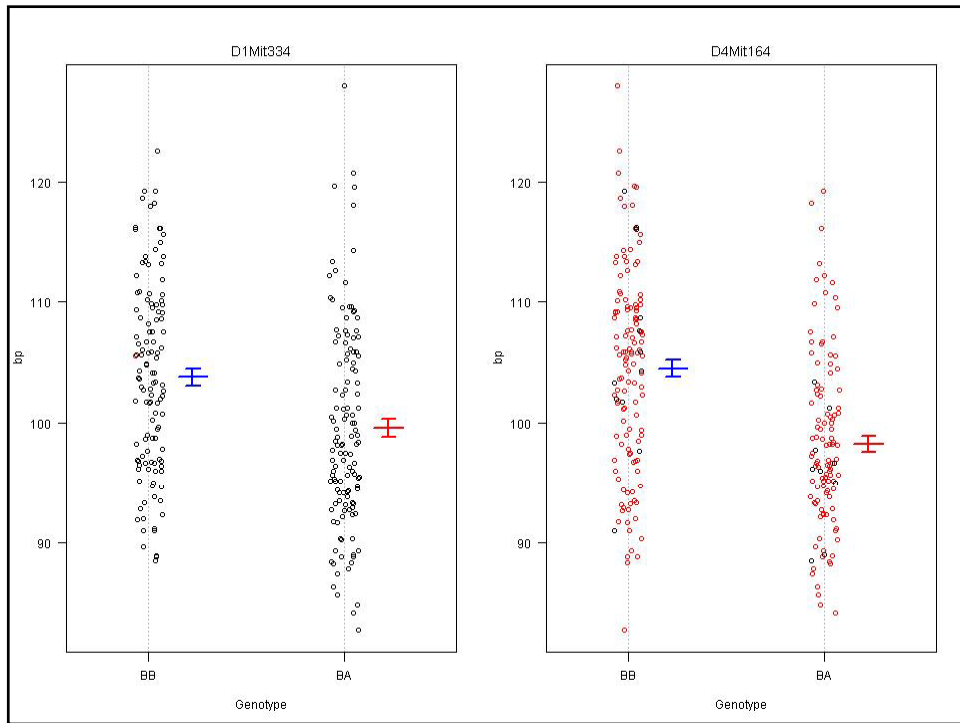
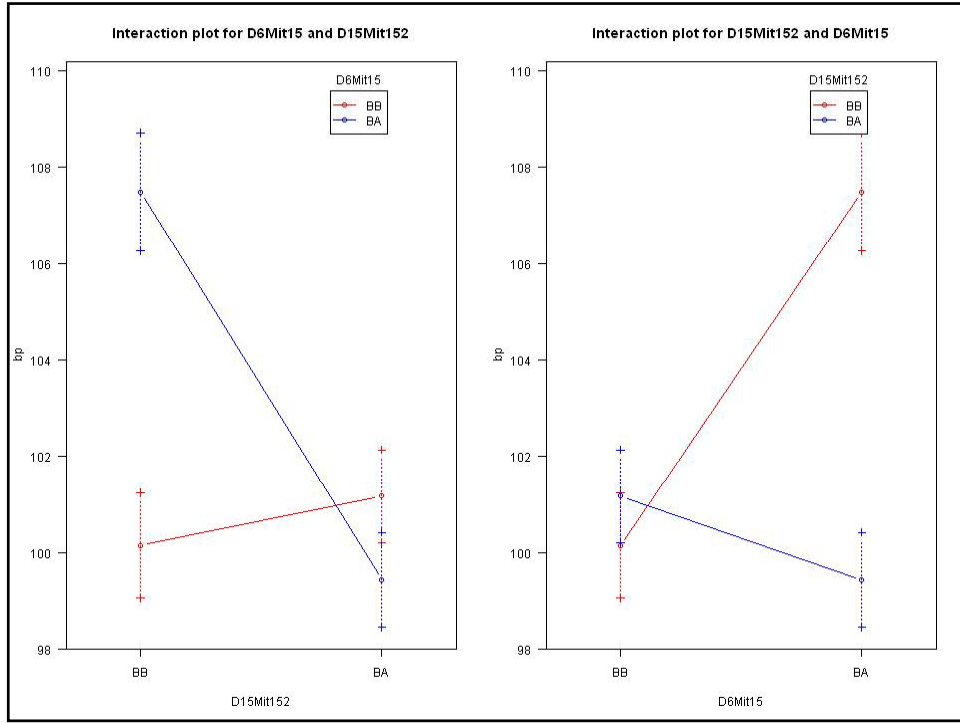
## Strip plot of data (phenotype by genotype).
plot.pwg(hyper, "D1Mit334")
plot.pwg(hyper, "D4Mit164")
plot.pwg(hyper, markers)
```

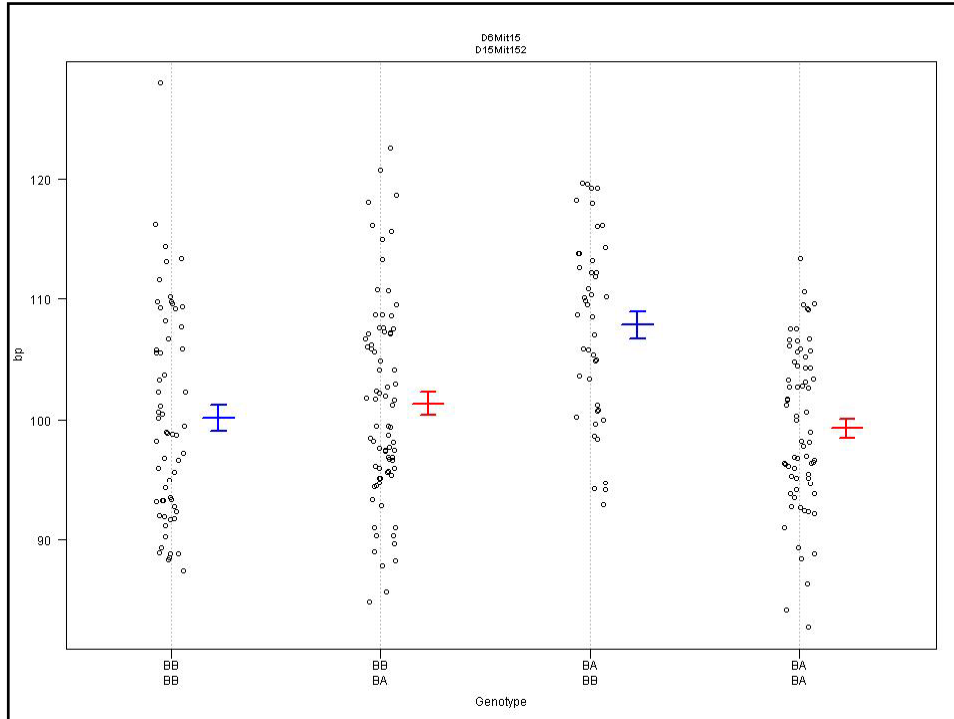
R/qtlibim Tutorial

NSF StatGen: Yandell © 2009

13







R/qtl: ANOVA imputation at QTL

```
> hyper <- sim.geno(hyper, step=2, n.draws=16, error.prob=0.01)
> qtl <- makeqtl(hyper, chr = c(1, 1, 4, 6, 15), pos = c(50, 76, 30, 70, 20))

> my.formula <- y ~ Q1 + Q2 + Q3 + Q4 + Q5 + Q4:Q5
> out.fitqtl <- fitqtl(hyper, pheno.col = 1, qtl, formula = my.formula)
> summary(out.fitqtl)
```

Full model result

Model formula is: $y \sim Q1 + Q2 + Q3 + Q4 + Q5 + Q4:Q5$

	df	SS	MS	LOD	%var	Pvalue(Chi2)	Pvalue(F)
Model	6	5789.089	964.84822	21.54994	32.76422	0	0
Error	243	11879.847	48.88826				
Total	249	17668.936					

Drop one QTL at a time ANOVA table:

	df	Type III SS	LOD	%var	F value	Pvalue(F)
Chr1@50	1	297.149	1.341	1.682	6.078	0.01438 *
Chr1@76	1	520.664	2.329	2.947	10.650	0.00126 **
Chr4@30	1	2842.089	11.644	16.085	58.134	5.50e-13 ***
Chr6@70	2	1435.721	6.194	8.126	14.684	9.55e-07 ***
Chr15@20	2	1083.842	4.740	6.134	11.085	2.47e-05 ***
Chr6@70:Chr15@20	1	955.268	4.199	5.406	19.540	1.49e-05 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

selected R/qtl publications

www.stat.wisc.edu/~yandell/statgen

- www.rqtl.org
- tutorials and code at web site
 - www.rqtl.org/tutorials
- Broman et al. (2003 *Bioinformatics*)
 - R/qtl introduction
- Broman (2001 *Lab Animal*)
 - nice overview of QTL issues
- Broman & Sen 2009 book (*Springer*)

19

R/qtlbim (www.qtlbim.org)

- cross-compatible with R/qtl
- model selection for genetic architecture
 - epistasis, fixed & random covariates, GxE
 - samples multiple genetic architectures
 - examines summaries over nested models
- extensive graphics

```
> url.show("http://www.stat.wisc.edu/~yandell/qtlbim/rqtlbimtour.R")
```

R/qtlbim: tutorial

(www.stat.wisc.edu/~yandell/qtlbim)

```
> data(hyper)
## Drop X chromosome (for now).
> hyper <- subset(hyper, chr=1:19)
> hyper <- qb.genoprob(hyper, step=2)
## This is the time-consuming step:
> qbHyper <- qb.mcmc(hyper, pheno.col = 1)
## Here we get stored samples.
> data(qbHyper)
> summary(qbHyper)
```

R/qtlbim: initial summaries

```
> summary(qbHyper)

Bayesian model selection QTL mapping object qbHyper on cross object hyper
had 3000 iterations recorded at each 40 steps with 1200 burn-in steps.

Diagnostic summaries:
      nqtl  mean envvar  varadd  varaa  var
Min.   2.000  97.42  28.07  5.112  0.000  5.112
1st Qu. 5.000 101.00  44.33 17.010  1.639 20.180
Median  7.000 101.30  48.57 20.060  4.580 25.160
Mean    6.543 101.30  48.80 20.310  5.321 25.630
3rd Qu. 8.000 101.70  53.11 23.480  7.862 30.370
Max.   13.000 103.90  74.03 51.730 34.940 65.220

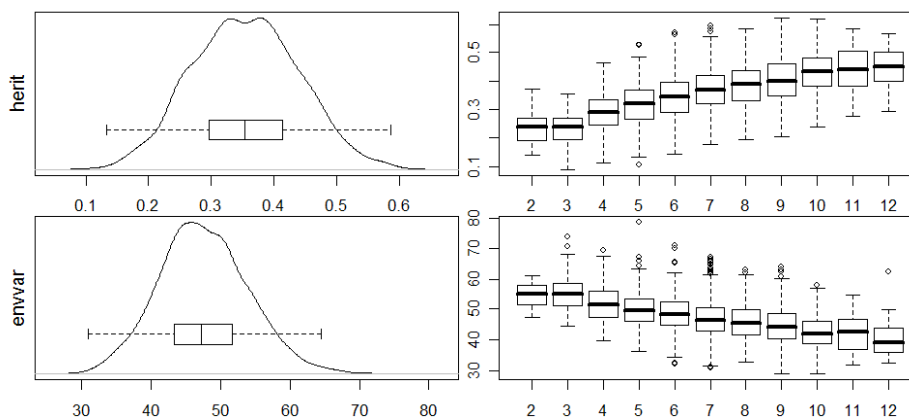
Percentages for number of QTL detected:
 2  3  4  5  6  7  8  9 10 11 12 13
 2  3  9 14 21 19 17 10  4  1  0  0

Percentages for number of epistatic pairs detected:
Pairs
 1  2  3  4  5  6
29 31 23 11  5  1

Percentages for common epistatic pairs:
 6.15  4.15  4.6  1.7 15.15  1.4  1.6  4.9  1.15  1.17  1.5  5.11  1.2  7.15  1.1
  63  18  10  6  6  5  4  4  3  3  3  2  2  2  2

> plot(qb.diag(qbHyper, items = c("herit", "envvar")))
```

diagnostic summaries



R/qtlbim Tutorial

NSF StatGen: Yandell © 2009

23

R/qtlbim: 1-D (*not* 1-QTL!) scan

```
> one <- qb.scanone(qbHyper, chr = c(1,4,6,15), type =
"LPD")
> summary(one)
```

LPD of bp for main,epistasis,sum

	n.qtl	pos	m.pos	e.pos	main	epistasis	sum
c1	1.331	64.5	64.5	67.8	6.10	0.442	6.27
c4	1.377	29.5	29.5	29.5	11.49	0.375	11.61
c6	0.838	59.0	59.0	59.0	3.99	6.265	9.60
c15	0.961	17.5	17.5	17.5	1.30	6.325	7.28

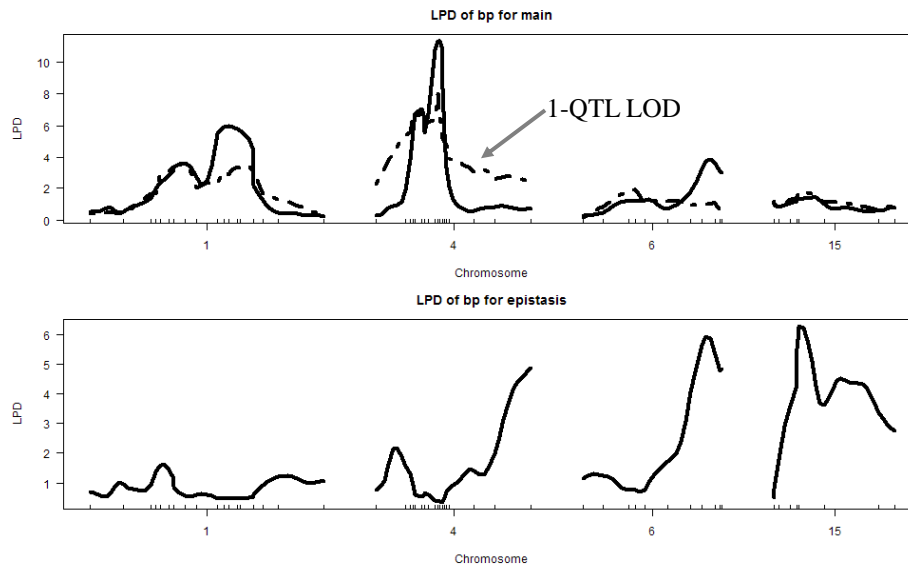
```
> plot(one, scan = "main")
> plot(out.em, chr=c(1,4,6,15), add = TRUE, lty = 2)
> plot(one, scan = "epistasis")
```

R/qtlbim Tutorial

NSF StatGen: Yandell © 2009

24

1-QTL LOD vs. marginal LPD



R/qtlbim Tutorial

NSF StatGen: Yandell © 2009

25

most probable patterns

```
> summary(qb.BayesFactor(qbHyper, item = "pattern"))
```

	nqtl	posterior	prior	bf	bfse
1,4,6,15,6:15	5	0.03400	2.71e-05	24.30	2.360
1,4,6,6,15,6:15	6	0.00467	5.22e-06	17.40	4.630
1,1,4,6,15,6:15	6	0.00600	9.05e-06	12.80	3.020
1,1,4,5,6,15,6:15	7	0.00267	4.11e-06	12.60	4.450
1,4,6,15,15,6:15	6	0.00300	4.96e-06	11.70	3.910
1,4,4,6,15,6:15	6	0.00300	5.81e-06	10.00	3.330
1,2,4,6,15,6:15	6	0.00767	1.54e-05	9.66	2.010
1,4,5,6,15,6:15	6	0.00500	1.28e-05	7.56	1.950
1,2,4,5,6,15,6:15	7	0.00267	6.98e-06	7.41	2.620
1,4	2	0.01430	1.51e-04	1.84	0.279
1,1,2,4	4	0.00300	3.66e-05	1.59	0.529
1,2,4	3	0.00733	1.03e-04	1.38	0.294
1,1,4	3	0.00400	6.05e-05	1.28	0.370
1,4,19	3	0.00300	5.82e-05	1.00	0.333

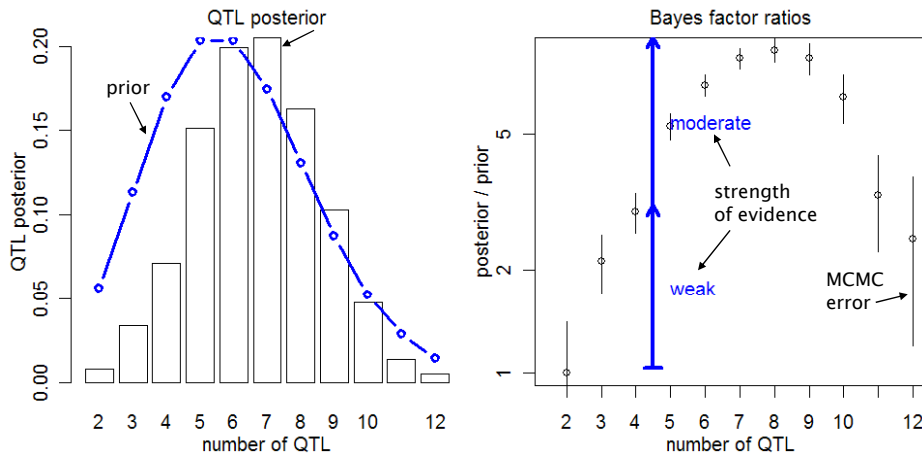
```
> plot(qb.BayesFactor(qbHyper, item = "nqtl"))
```

R/qtlbim Tutorial

NSF StatGen: Yandell © 2009

26

hyper: number of QTL posterior, prior, Bayes factors



R/qtlbim Tutorial

NSF StatGen: Yandell © 2009

27

what is best estimate of QTL?

- find most probable pattern
 - 1,4,6,15,6:15 has posterior of 3.4%
- estimate locus across all nested patterns
 - Exact pattern seen ~100/3000 samples
 - Nested pattern seen ~2000/3000 samples
- estimate 95% confidence interval using quantiles

```
> best <- qb.best(qbHyper)
> summary(best)$best
```

	chrom	locus	locus.LCL	locus.UCL	n.qtl	
	247	1	69.9	24.44875	95.7985	0.8026667
	245	4	29.5	14.20000	74.3000	0.8800000
	248	6	59.0	13.83333	66.7000	0.7096667
	246	15	19.5	13.10000	55.7000	0.8450000

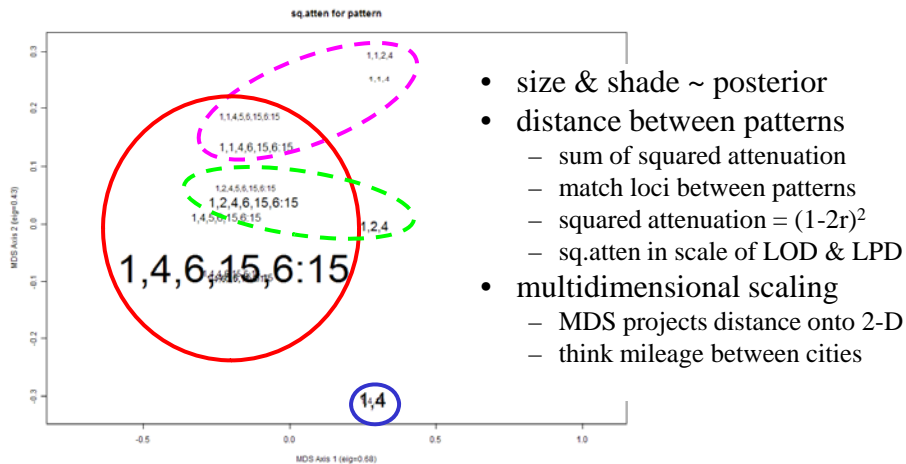
```
> plot(best)
```

R/qtlbim Tutorial

NSF StatGen: Yandell © 2009

28

what patterns are “near” the best?



- size & shade ~ posterior
- distance between patterns
 - sum of squared attenuation
 - match loci between patterns
 - squared attenuation = $(1-2r)^2$
 - sq.atten in scale of LOD & LPD
- multidimensional scaling
 - MDS projects distance onto 2-D
 - think mileage between cities

R/qtlbim Tutorial

NSF StatGen: Yandell © 2009

29

how close are other patterns?

```
> target <- qb.best(qbHyper)$model[[1]]
> summary(qb.close(qbHyper, target))

score by sample number of qtl
  Min. 1st Qu. Median Mean 3rd Qu.  Max.
2  1.437  1.735  1.919 1.834  1.919 2.000
3  1.351  1.735  1.916 1.900  1.919 2.916
4  1.270  1.916  2.437 2.648  3.574 4.000
5  1.295  1.919  2.835 2.798  3.611 4.000
6  1.257  2.254  3.451 3.029  3.648 4.000
...
13 3.694  3.694  3.694 3.694  3.694 3.694

score by sample chromosome pattern
      Percent  Min. 1st Qu. Median Mean 3rd Qu.  Max.
4@1,4,6,15,6:15  3.4 2.946  3.500 3.630 3.613  3.758 4.000
2@1,4            1.4 1.437  1.735 1.919 1.832  1.919 2.000
5@1,2,4,6,15,6:15 0.8 3.137  3.536 3.622 3.611  3.777 3.923
3@1,2,4          0.7 1.351  1.700 1.821 1.808  1.919 2.000
5@1,1,4,6,15,6:15 0.6 3.257  3.484 3.563 3.575  3.698 3.916
5@1,4,5,6,15,6:15 0.5 3.237  3.515 3.595 3.622  3.777 3.923
5@1,4,6,6,15,6:15 0.5 3.203  3.541 3.646 3.631  3.757 3.835
...
```

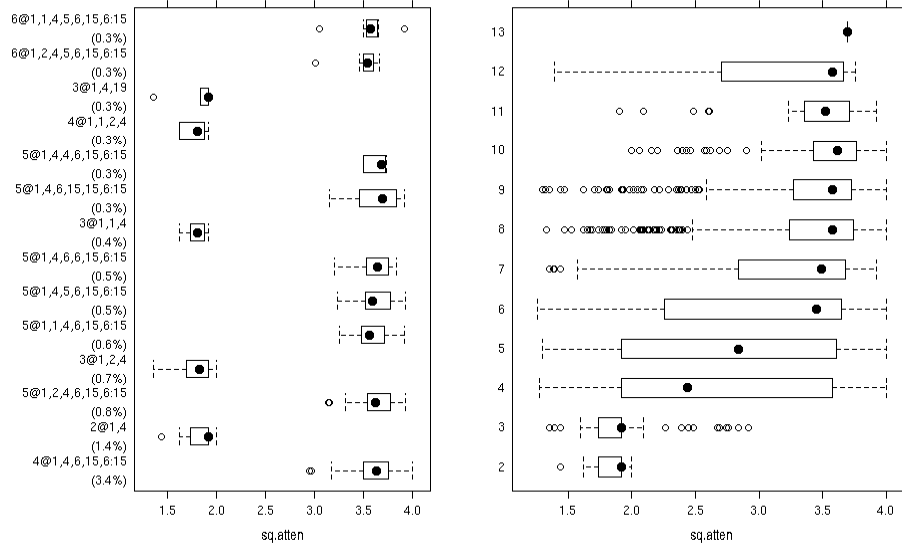
```
> plot(close)
> plot(close, category = "nqtl")
```

R/qtlbim Tutorial

NSF StatGen: Yandell © 2009

30

how close are other patterns?



R/qtlbim Tutorial

NSF StatGen: Yandell © 2009

31

R/qtlbim: automated QTL selection

```
> hpd <- qb.hpdone(qbHyper, profile = "2logBF")
> summary(hpd)
```

chr	n.qtl	pos	lo.50%	hi.50%	2logBF	A	H	
1	1	0.829	64.5	64.5	72.1	6.692	103.611	99.090
4	4	3.228	29.5	25.1	31.7	11.169	104.584	98.020
6	6	1.033	59.0	56.8	66.7	6.054	99.637	102.965
15	15	0.159	17.5	17.5	17.5	5.837	101.972	100.702

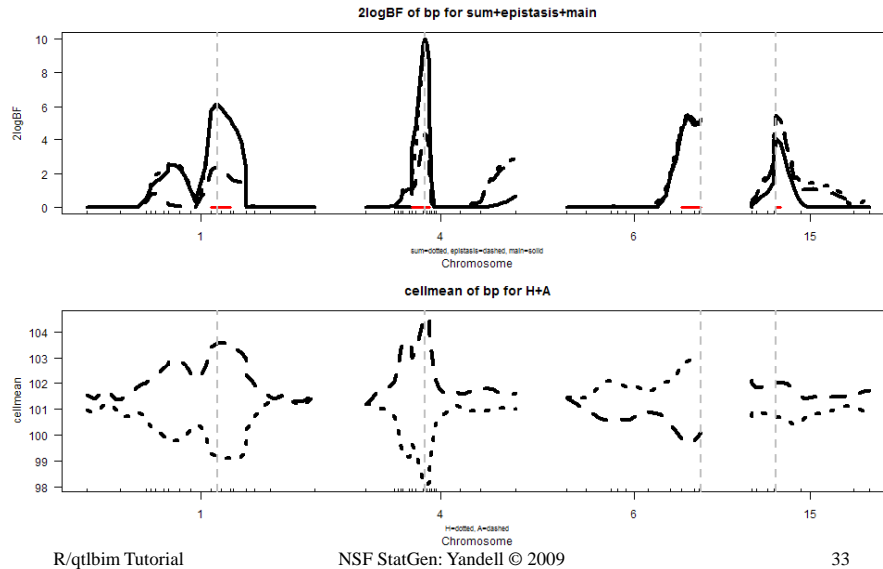
```
> plot(hpd)
```

R/qtlbim Tutorial

NSF StatGen: Yandell © 2009

32

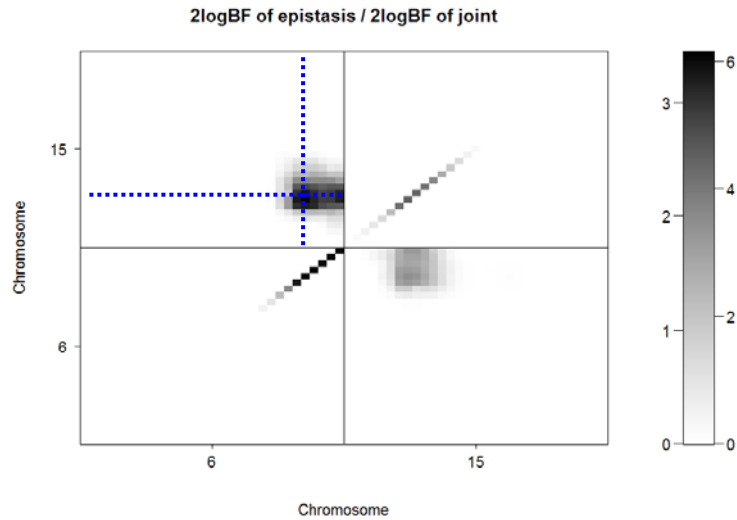
2log(BF) scan with 50% HPD region



R/qtlbim: 2-D (*not* 2-QTL) scans

```
> two <- qb.scantwo(qbHyper, chr = c(6,15),  
  type = "2logBF")  
> plot(two)  
  
> plot(two, chr = 6, slice = 15)  
> plot(two, chr = 15, slice = 6)  
  
> two.lpd <- qb.scantwo(qbHyper, chr = c(6,15),  
  type = "LPD")  
> plot(two.lpd, chr = 6, slice = 15)  
> plot(two.lpd, chr = 15, slice = 6)
```

2-D plot of 2logBF: chr 6 & 15

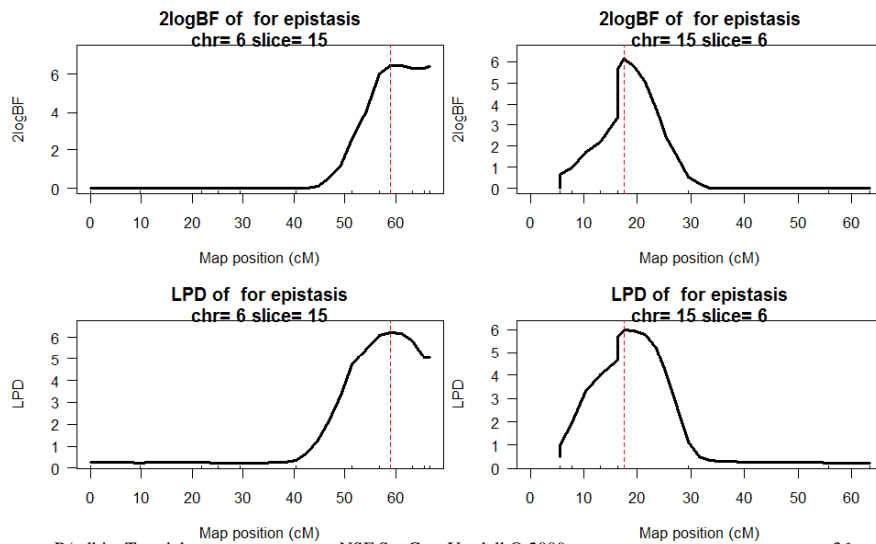


R/qtlbim Tutorial

NSF StatGen: Yandell © 2009

35

1-D Slices of 2-D scans: chr 6 & 15



R/qtlbim Tutorial

NSF StatGen: Yandell © 2009

36

R/qtlbim: slice of epistasis

```

> slice <- qb.slicetwo(qbHyper, c(6,15), c(59,19.5))
> summary(slice)

2logBF of bp for epistasis

      n.qtl  pos m.pos e.pos epistasis slice
c6  0.838 59.0 59.0 66.7    15.8 18.1
c15 0.961 17.5 17.5 17.5    15.5 60.6

cellmean of bp for AA,HA,AH,HH

      n.qtl  pos m.pos  AA  HA  AH  HH slice
c6  0.838 59.0 59.0 97.4 105 102 100.8 18.1
c15 0.961 17.5 17.5 99.8 103 104  98.5 60.6

estimate of bp for epistasis

      n.qtl  pos m.pos e.pos epistasis slice
c6  0.838 59.0 59.0 66.7    -7.86 18.1
c15 0.961 17.5 17.5 17.5    -8.72 60.6

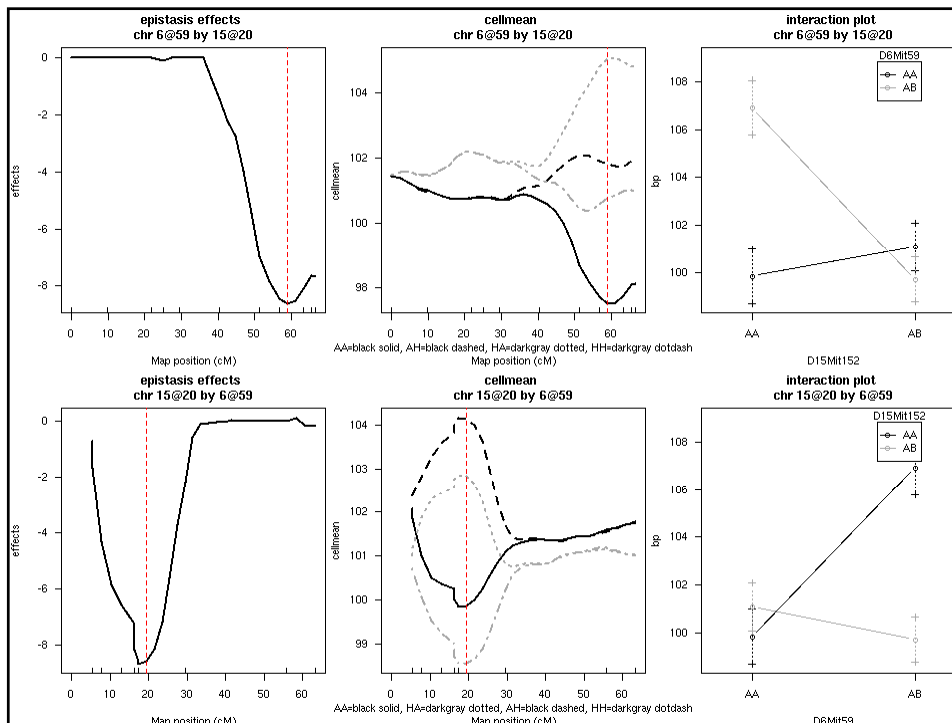
> plot(slice, figs = c("effects", "cellmean", "effectplot"))

```

R/qtlbim Tutorial

NSF StatGen: Yandell © 2009

37



selected publications

www.stat.wisc.edu/~yandell/statgen

- www.qtlbim.org
- vignettes in R/qtlbim package
- Yandell, Bradbury (2007) *Plant Map* book chapter
 - overview/comparison of QTL methods
- Yandell et al. (2007 *Bioinformatics*)
 - R/qtlbim introduction
- Yi et al. (2005 *Genetics*, 2007 *Genetics*)
 - methodology of R/qtlbim