

Prototype QTL Strategy: Phenotype bp in Cross hyper

Brian S. Yandell, W. Whipple Neely, Tapan Mehta, Daniel Schriener, Samprit Banerjee, and Nengjun Yi

October 18, 2006

1 Overview

This document analyzes trait bp for dataset hyper using the 1-D and 2-D Bayesian genome scan routines that build on Markov chain Monte Carlo (MCMC) samples from a posterior for the genetic architecture of a trait. Below the generic `cross` is actually the cross passed by the user in a call to `qb.sweave`. This entire document was created automatically by a function call in R. The function is not yet part of R/`qtlbim`. The actual call was

```
> library(qtlbim)

> qb.sweave(hyper, pheno.col = 1,
+ n.iter = 3000, n.draws = 64,
+ scan.type = "2logBF", hpd.level = 0.5,
+ threshold = c(upper = 2),
+ SweaveFile = "/tmp/Rinst1107343038/qtlbim/doc/hyperslide.Rnw",
+ SweaveExtra = "/tmp/Rinst1107343038/qtlbim/external/hyperpaperextra.Rnw",
+ PDFDir = "bpPDF",
+ remove.qb = TRUE)
```

At present, the `threshold` values are somewhat arbitrary, chosen for the hyper dataset to pick up apparently real QTL and previously detected epistasis.

This document automates a search for main and epistatic QTL. The main QTL positions are reliably estimated using `qb.scanone`. This pass also seems to capture the major chromosomes possibly involved in epistasis, although it does not provide very good estimates of positions of purely epistatic QTL within those chromosomes. Next we use `qb.scantwo` and to identify which pairs of QTL are epistatic, and to get initial estimates of their positions. We refine there positions with `qb.slice`. Along the way, we use generic `summary` and `plot` routines to view results.

Once we have reasonable estimates of QTL postions and effects, we use confirmatory ANOVA tools to refine the model. That is, we use R/qtl's simulation-based `fitqtl` followed by a stepwise backward fitting approach, using a new `step.fitqtl`, to confirm key QTL. That completes this automated analysis. It would be possible to add other, user-supplied refined analysis at the end of this document if desired.

2 Generating Samples

Here is a summary of the `cross` copy of the hyper object, followed by creation of 3000 MCMC samples.

```
> summary(cross)

Backcross

No. individuals:    250

No. phenotypes:    1
```

```

Percent phenotyped: 100

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

Total markers:        170
No. markers:          22 8 6 20 14 11 7 6 5 5 14 5 5 5 11 6 12 4 4
Percent genotyped:    47.9
Genotypes (%):        AA:50.1  AB:49.9

```

```

> cross <- qb.genoprob(cross,step=2)
> cross.qb <- qb.mcmc(cross, pheno.col = pheno.col
+   genoupdate=TRUE, n.iter = 3000, mydir = "bpPDF")

```

3 1-D Scans

Now a 1-D scan picks out the major effects. We could use `qb.scanone` directly. Instead, we use `qb.hpdone`, which gives us a profile scan as well as a scan of genotypic means.

```

> hpd.level

[1] 0.5

> cross.hpd <- qb.hpdone(cross.qb, hpd.level)
> sum.one <- summary(cross.hpd)
> sum.one

```

	chr	n.qtl	pos	lo.50%	hi.50%	2logBF	A	H
<NA>	1	0.694	64.5	64.5	69.9	6.796	103.604	99.073
<NA>	4	3.460	29.5	25.1	31.7	11.347	104.561	98.026
<NA>	6	1.107	59.0	56.8	66.7	6.179	99.606	102.924
<NA>	15	0.341	17.5	17.5	17.5	6.032	101.940	100.692

The `qb.hpdone` routine mostly automates the selection of peaks. We are still working on how to set reasonable thresholds, but for now a HPD level of 0.5 for the overall (`sum`) picks up the key features nicely for many phenotypes. The variable `hpd.level` was set in the call to `qb.sweave` that created this document. The new variables `chrs` and `pos` capture the chromosome numbers and main QTL positions, respectively.

```

> chrs <- as.vector(sum.one[, "chr"])
> pos <- sum.one[, "pos"]
> pos

<NA> <NA> <NA> <NA>
64.5 29.5 59.0 17.5

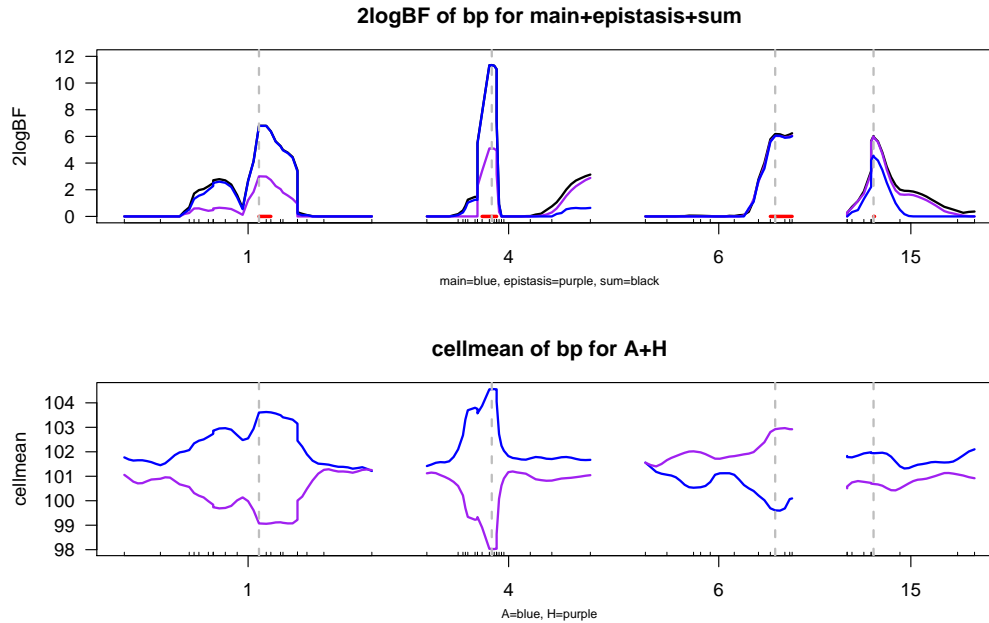
```

The following two figures highlight the selected chromosomes. In the top panel, the blue curve represents the `2logBF` (twice log of Bayes factor) score for the `main` effect of a QTL at each locus, while the purple curve shows the score for any `epistasis` between a locus and any other loci. The black curve shows the combination of main and epistatic effects. The second panel shows the marginal means, which are roughly symmetric about the phenotype mean of 101.

```

> plot(cross.hpd)

```



4 2-D Scan

Now a look at a 2-D scan reveals the strength of epistasis. The summary suggests some other epistases, but some of this may be spurious [i.e. we will want to look further!]. We consider a subset of this summary above the upper threshold of 2.

```
> two <- qb.scantwo(cross.qb, chr = chrs, type = scan.type)
> sum.two <- summary(two, threshold = threshold, refine = TRUE)
> sum.two
```

	chr1	chr2	n.qtl	l.pos1	l.pos2	lower	u.pos1	u.pos2	upper
1:1	1	1	0.362	43.7	77.6	7.583	43.7	74.300	4.697
1:4	1	4	1.352	67.8	29.5	15.705	72.1	29.500	7.303
1:6	1	6	1.831	67.8	59.0	12.611	104.9	0.000	4.756
1:15	1	15	0.298	67.8	17.5	12.012	77.6	17.500	5.794
4:4	4	4	1.145	29.5	74.3	11.820	0.0	30.600	4.756
4:6	4	6	1.561	29.5	66.7	14.884	74.3	59.000	7.728
4:15	4	15	0.446	29.5	17.5	14.539	74.3	35.546	7.350
6:6	6	6	1.214	61.2	65.6	7.442	27.3	66.700	4.756
6:15	6	15	1.080	59.0	17.5	12.779	59.0	17.500	12.751
15:15	15	15	0.105	17.5	27.5	8.125	17.5	25.500	7.234

Now let's extract the genetic architecture implied by this evidence for epistasis. The loci pairs that show epistasis are indexed to the vector of main QTL. In addition we see the pairs of chromosomes involved in epistasis. Some effort is made to merge nearby QTL estimated positions in `qb.arch`.

```
> cross.arch <- qb.arch(sum.two, chrs, pos)
> cross.arch
```

```
main QTL loci:
      1      2      3 4      5      6 7      8      9      10      11
chr 1.0  1.00  1.0 4  4.00  4.0 6  6.0  6.00 15.0 15.00
pos 43.7 72.12 104.9 0 29.87 74.3 0 27.3 60.92 19.1 35.55
```

```
Epistatic pairs by qtl, chr, pos:
  qtlb qtlc chra chrb  posa  posb
1    1    2    1    1  43.70 72.12
2    2    5    1    4  72.12 29.87
3    3    7    1    6 104.90  0.00
4    2   10    1   15  72.12 19.10
5    4    5    4    4   0.00 29.87
6    6    9    4    6  74.30 60.92
7    6   11    4   15  74.30 35.55
8    8    9    6    6  27.30 60.92
9    9   10    6   15  60.92 19.10
Epistatic chromosomes by connected sets:
1,4,6,15
```

5 ANOVA Model Fit

Here we want to merge the 1-D `chrs` and `pos` with the 2-D epistatic pairs to determine the chromosomes and positions to include in an ANOVA fit. We equate QTL that are within, say 10cM of each other. After fitting a (very) full model, we use `step.fitqtl`, a newly written routine, to step-by-step reduce the model to key main effects and interactions, preserving hierarchy.

The following uses R/qtl tools `calc.genoprob`, `sim.geno` and `makeqtl`, plus the new `step.fitqtl`, which calls `fitqtl` multiple times.

```
> cross.sub <- subset(cross, chr = cross.arch$qtl$chr)
> n.draws

[1] 64

> cross.sub <- sim.geno(cross.sub, n.draws = n.draws, step = 2,
+   error = 0.01)
> qtl <- makeqtl(cross.sub, cross.arch$qtl$chr, cross.arch$qtl$pos)
```

Now we run stepwise backward elimination, preserving hierarchy.

```
> cross.step <- step.fitqtl(cross.sub, qtl, pheno.col, cross.arch)
```

drop	LOD	p
1 Chr4@0:Chr4@29.87	0.24500	0.3100
2 Chr4@0	-0.00603	1.0000
3 Chr6@27.3:Chr6@60.92	0.30700	0.2540
4 Chr1@72.12:Chr15@19.1	0.33500	0.2320
5 Chr6@27.3	0.56600	0.1190
6 Chr1@43.7:Chr1@72.12	0.64000	0.0972
7 Chr1@72.12:Chr4@29.87	0.72400	0.0769

```
> summary(cross.step$fit)
```

Full model result

```
-----
```

	df	SS	MS	LOD	%var	Pvalue(Chi2)	Pvalue(F)
Model	13	7512.709	577.90070	30.05975	42.51931	0	0
Error	236	10156.227	43.03486				
Total	249	17668.936					

Drop one QTL at a time ANOVA table:

	df	Type III SS	LOD	%var	F value	Pvalue(F)	
Chr1@43.7	1	273.673	1.443	1.549	6.359	0.01234	*
Chr1@72.12	1	377.222	1.980	2.135	8.765	0.00338	**
Chr1@104.9	2	282.224	1.488	1.597	3.279	0.03939	*
Chr4@29.87	1	2334.723	11.233	13.214	54.252	2.94e-12	***
Chr4@74.3	3	962.446	4.915	5.447	7.455	8.64e-05	***
Chr6@0	2	284.658	1.501	1.611	3.307	0.03832	*
Chr6@60.92	3	2042.300	9.947	11.559	15.819	2.09e-09	***
Chr15@19.1	2	1377.019	6.902	7.793	15.999	3.05e-07	***
Chr15@35.55	2	444.245	2.324	2.514	5.161	0.00640	**
Chr1@104.9:Chr6@0	1	271.467	1.432	1.536	6.308	0.01269	*
Chr4@74.3:Chr6@60.92	1	275.863	1.455	1.561	6.410	0.01200	*
Chr4@74.3:Chr15@35.55	1	430.595	2.254	2.437	10.006	0.00177	**
Chr6@60.92:Chr15@19.1	1	1283.787	6.462	7.266	29.831	1.19e-07	***

The final model may be more complicated than a model found 'by hand' using standard R/qtl tools. Hopefully that model is a subset of the automatically found model.

6 2-D Epistasis Plots

Should there be any evidence for epistasis that is confirmed by ANOVA, it can be useful to view 2-D plots similar to `scantwo`, but now using the marginal 2logBF. Here is the reduced, final genetic architecture:

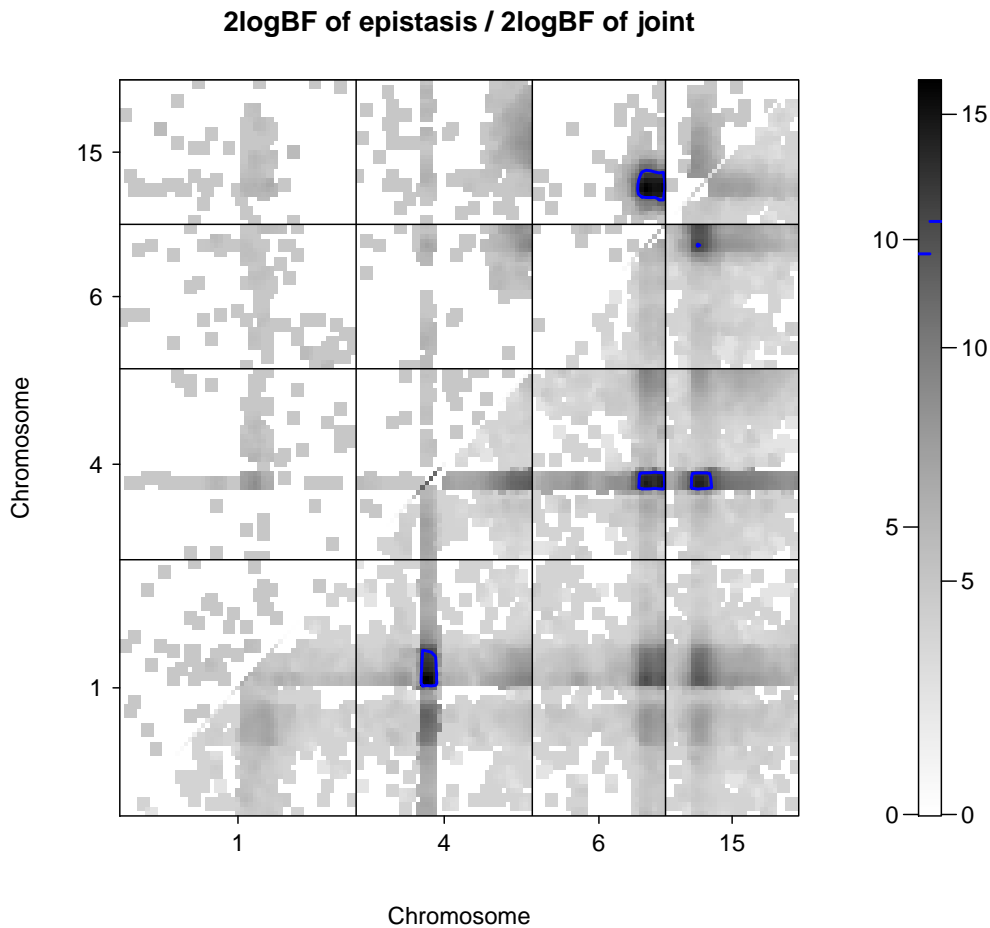
```
> cross.arch <- cross.step$arch
> cross.arch

main QTL loci:
      1      2      3      5      6 7      9      10      11
chr  1.0  1.00  1.0  4.00  4.0 6  6.00 15.0 15.00
pos 43.7 72.12 104.9 29.87 74.3 0 60.92 19.1 35.55
```

```
Epistatic pairs by qtl, chr, pos:
  q1 q2 chra chrb  posa  posb
1  3  7   1    6 104.90  0.00
2  6  9   4    6  74.30 60.92
3  6 11   4   15  74.30 35.55
4  9 10   6   15  60.92 19.10
Epistatic chromosomes by connected sets:
1,4,6,15
```

Here are the plots by clique (if any).

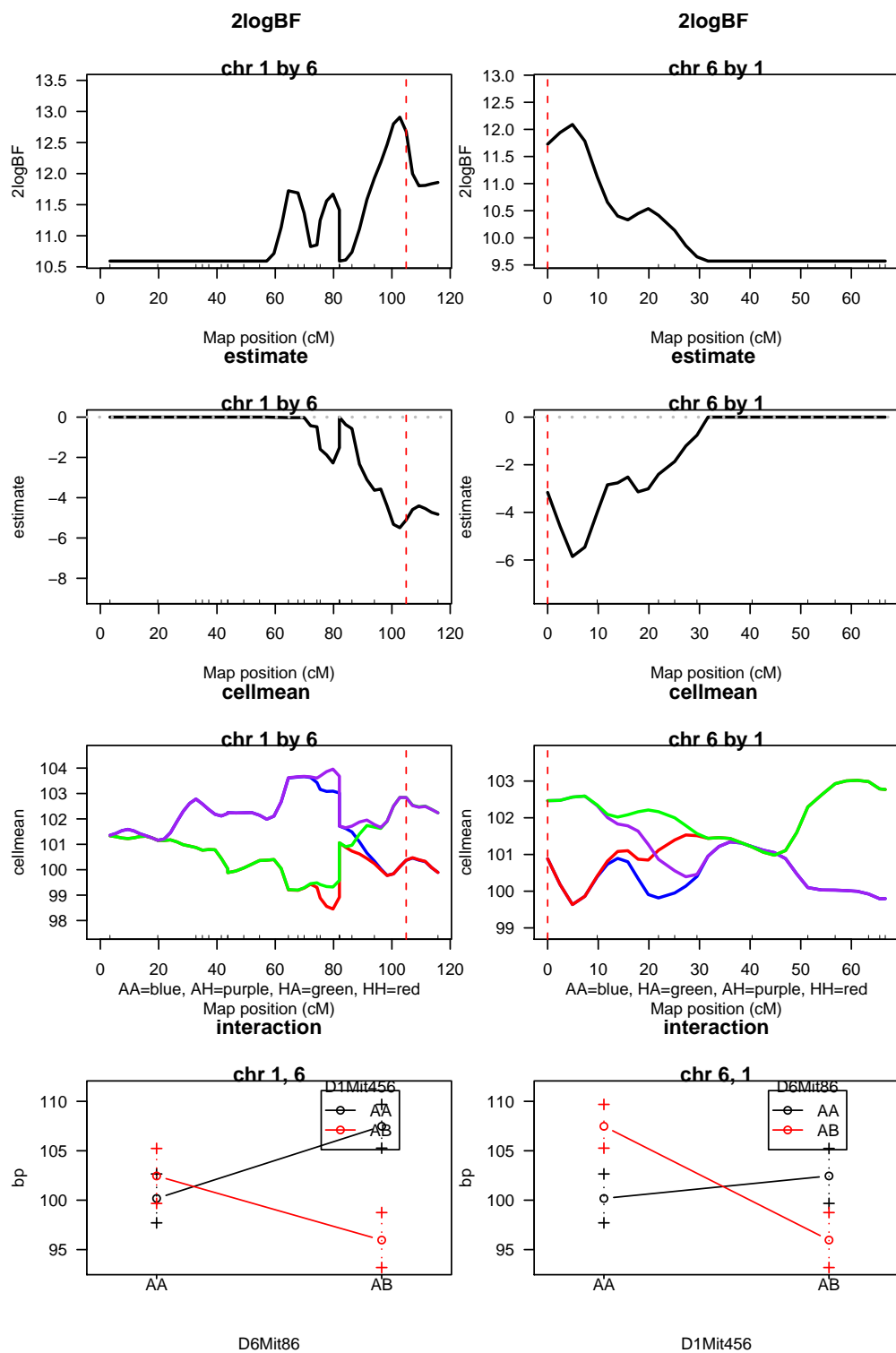
```
> for(i in names(cross.arch$chr.by.set))
+   plot(two, chr = cross.arch$chr.by.set[[i]], smooth = 3,
+        col = "gray", contour = 3)
```

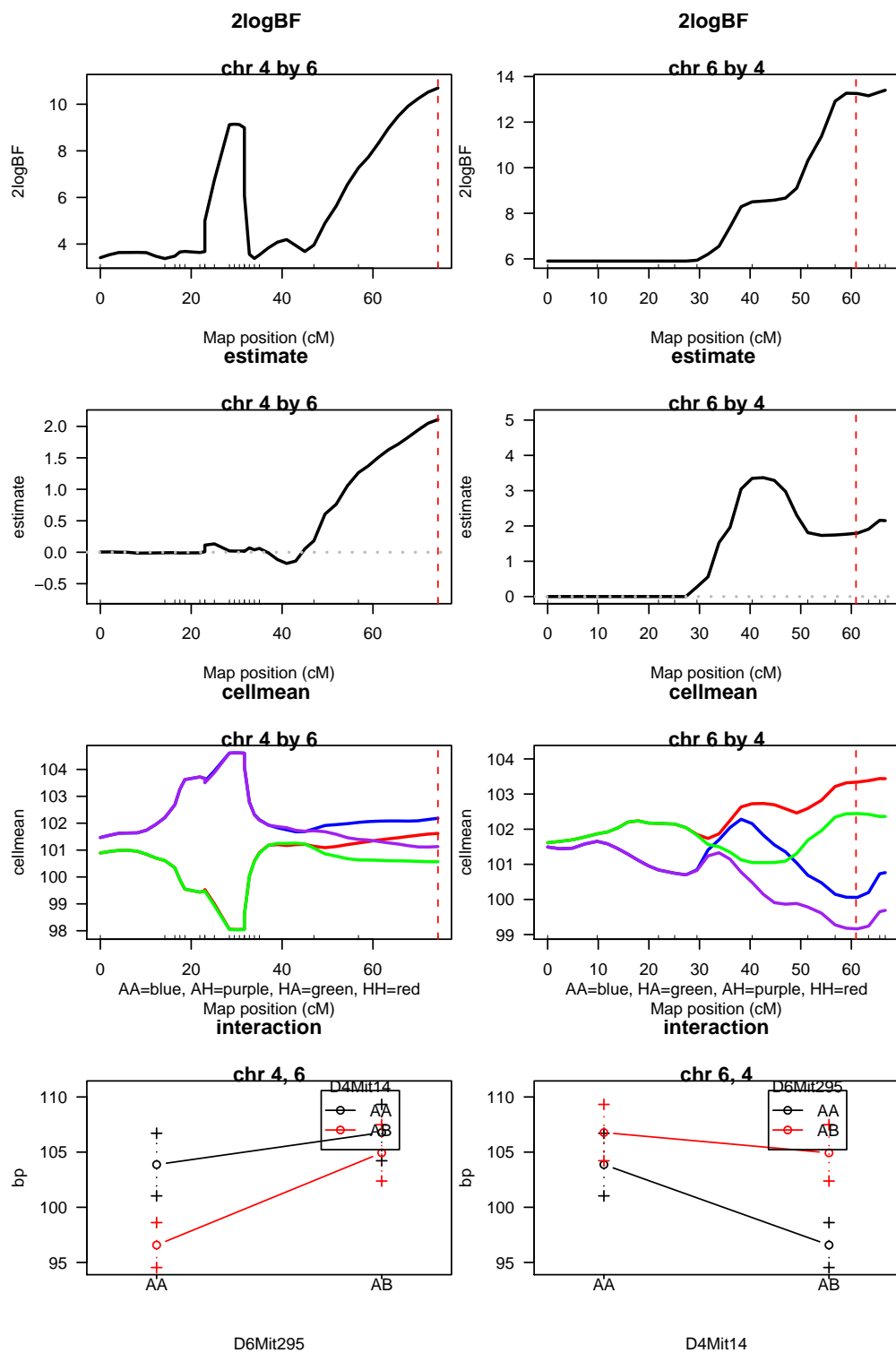


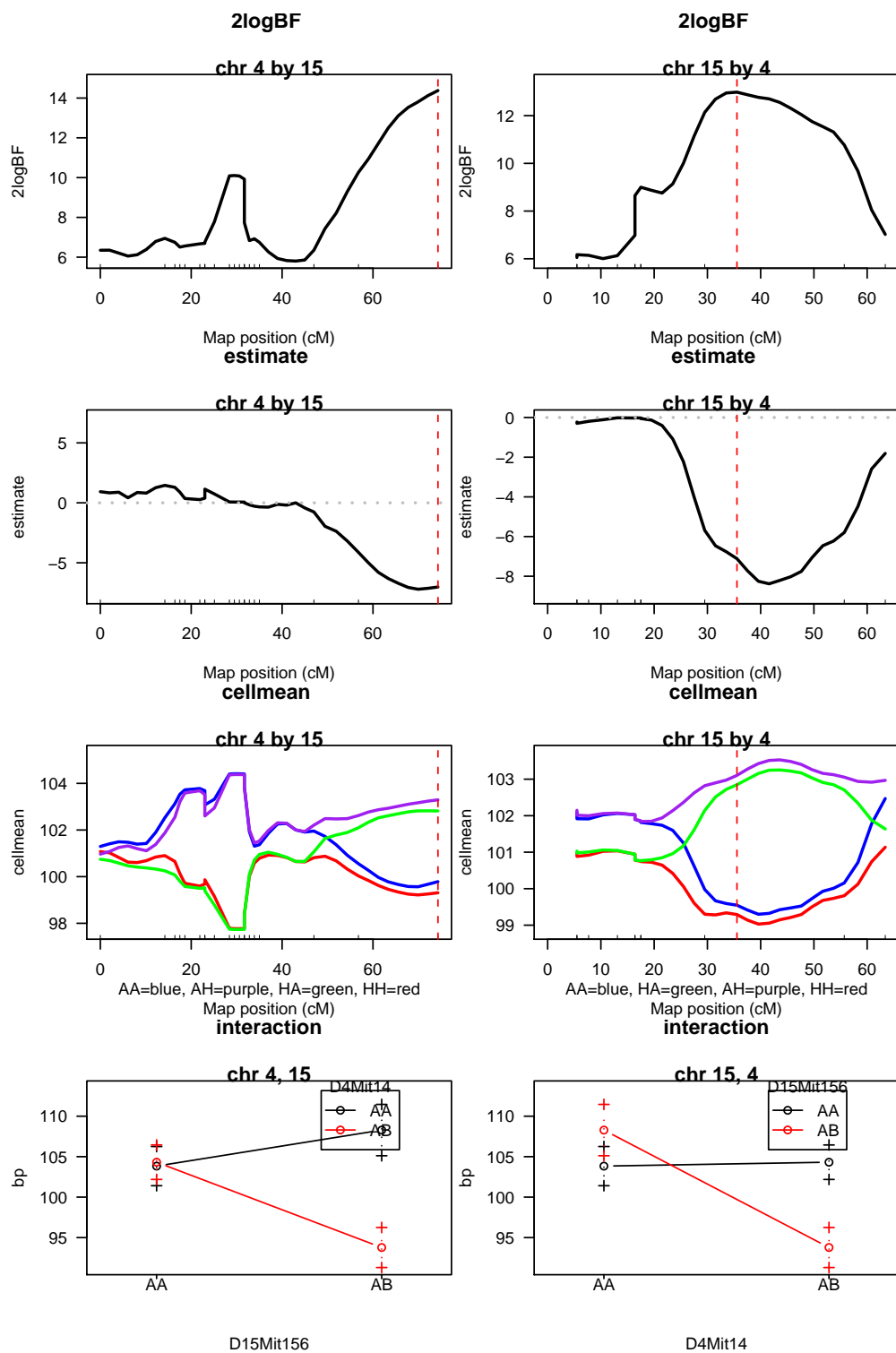
7 1-D Epistasis Slices

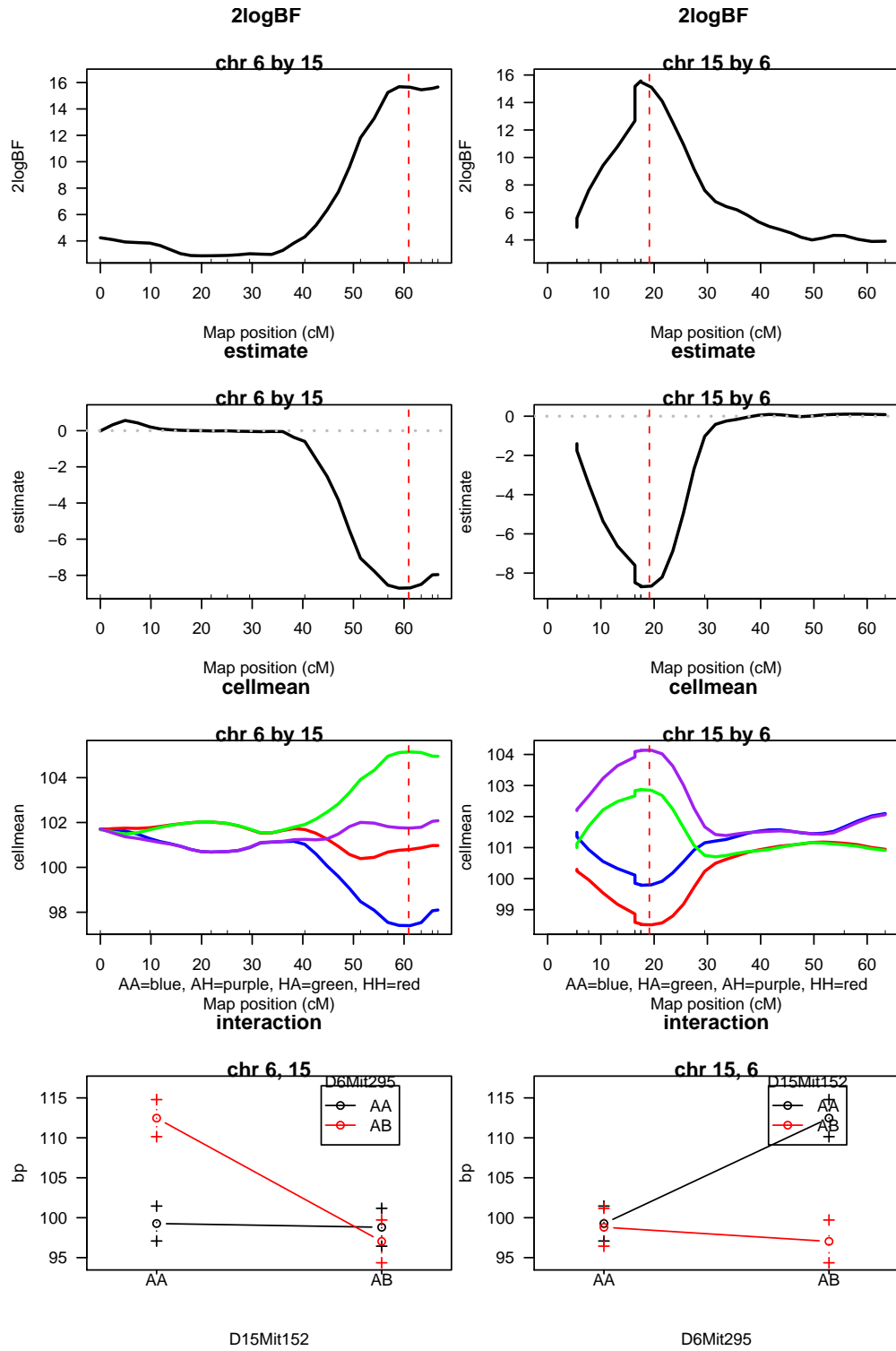
We then examine 1-D slices through the 2-D surface for epistatic pairs in the reduced model, to focus on epistasis for those identified pairs. We show 1-D slices of the LOD and the epistatic effects. In addition, we show interaction plots at the nearest markers.

```
> if(!is.null(cross.arch$pair.by.chr)) {
+   for(i in seq(nrow(cross.arch$pair.by.chr$chr))) {
+     chri <- cross.arch$pair.by.chr$chr[i,]
+     posi <- cross.arch$pair.by.chr$pos[i,]
+     plot(qb.slicetwo(cross.qb, chri, posi, scan.type), byrow = FALSE)
+   }
+ }
```









8 User Customized Section

We know from previous work that there are main QTLs on chromosomes 1 and 4, and epistatic pairs involving 6 and 15, and 7 and 15. Here we pick the nested model that contains these QTL.

```

> arch3 <- qb.arch(cross.step, main = c(1, 4), epistasis = data.frame(q1 = c(6,
+ 7), q2 = rep(15, 2)))
> arch3
> cross.step2 <- step.fitqtl(cross.sub, qtl, pheno.col, arch3)

> summary(cross.step2$fit)

```

Now we do a formal comparison of this reduced model with the fuller model we automatically uncovered. It appears that the fuller model is a much better fit.

```

> anova(cross.step, cross.step2)

```

9 Cleaning Up

That completes the template. Now the penultimate task is to remove the objects created by R/qtlbim, if this is desired by the user.

```

> remove.qb

[1] FALSE

> if (remove.qb) {
+   qb.remove(cross.qb)
+   rm(cross, cross.sub, pheno.col, threshold, n.iter, n.draws,
+       remove.qb)
+ }

```

Finally, externally rename file hyperslide.tex to bp.tex and run pdflatex twice on it. Use a free Acrobat reader to view.

```

> file.rename("hyperslide.tex", "bp.tex")
> invisible(system("pdflatex bp.tex",intern=TRUE))
> invisible(system("pdflatex bp.tex",intern=TRUE))

```