

Bayesian QTL mapping for allo-polyploids and diploids with BayesQTLBIC.

Rod Ball (Scion)

Gail Timmerman-Vaughan (Plant and Food) and Phillip Wilcox (Scion)

VISG workshop,
Dunedin, 23–26 Oct 2012



Part I: Peeling for allo-polyploids (and diploids)

2

polyploids_talk_VISG.tex RCS:1.0

19/10/2012

VISG Workshop: Polyploids

VISG Workshop: Polyploids

Peeling

Elston & Stewart 1971.

- method for evaluating probability distribution on a pedigree
- Bayesian graphical model
- utilises conditional independence in graphical model, avoids calculating all possible combinations (often impossibly many)

Strategy: peeling → Bayesian QTL mapping

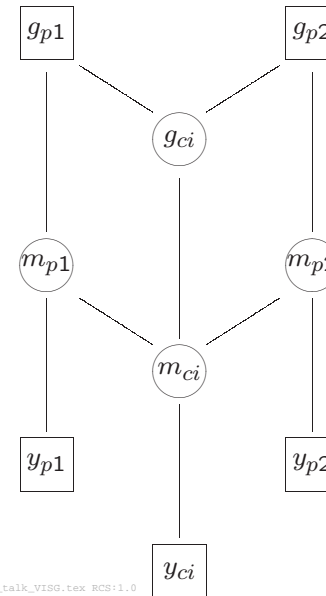
Software: R packages polyploids and BayesQTLBIC

3

4

Polyploids peeling graphical model

5



Peeling—a rough idea

1. sum over possibilities for a node; 'peel' it away \rightsquigarrow structure on reduced graph (' $R(\cdot)$ ' functions on parent nodes).
2. repeat to peel down to single mode (or nothing) \longrightarrow marginal probability (or likelihood)
3. for sampling, reverse the process

6

Peeling equations

Note the $R_{g_{ci}}(\cdot)$ etc functions, don't worry too much about the details.

The joint distribution is:

$$f(m_{p1}, m_{p2}, \dots) = [m_{p1}][m_{p2}][y_{p1} | m_{p1}][y_{p2} | m_{p2}] \times \prod_{i=1}^n [g_{ci} | g_{p1}, g_{p2}][m_{ci} | m_{p1}, m_{p2}, g_{ci}][y_{ci} | m_{ci}] \quad (1)$$

Let:

$$R_{g_{ci}}(g_{ci}, m_{p1}, m_{p2}) = \sum_{m_{ci}} [m_{ci} | m_{p1}, m_{p2}, g_{ci}][y_{ci} | m_{ci}] = I(f_2(f_1(m_{p1}, m_{p2}, g_{ci}) = y_{ci})) \quad (2)$$

7

The joint distribution after ‘peeling’ progeny marker genotypes m_{ci} becomes:

$$f(m_{p1}, m_{p2}, g_{c1}, g_{c2}, \dots) = [m_{p1}][m_{p2}][y_{p1} | m_{p1}][y_{p2} | m_{p2}] \times \prod_{i=1}^n [g_{ci} | g_{p1}, g_{p2}] R_{gci}(g_{ci}, m_{p1}, m_{p2}) \quad (3)$$

Next peel g_{ci} :

$$R_{gp1p2i}(g_{p1}, g_{p2}, m_{p1}, m_{p2}) = \sum_{g_{ci}} [g_{ci} | g_{p1}, g_{p2}] R_{gci}(g_{ci}, m_{p1}, m_{p2}) \quad (4)$$

For allo-tetraploids we obtain:

$$R_{gp1p2i}(g_{p1}, g_{p2}, m_{p1}, m_{p2}) = \sum_{g_{ci}} \frac{1}{2^4} I(g_{ci} \in c(1|2, 5|6, 3|4, 7|8)) \times R_{gci}(g_{ci}, m_{p1}, m_{p2}) \quad (5)$$

where the notation $1|2, 5|6, \dots$ means 1 or 2 and 5 or 6 etc. For allo-polyploids with ploidy $2k$ we have a similar expression with the factor $\frac{1}{2^4}$ is replaced by $\frac{1}{2^{2k}}$ in (5).

$$f(m_{p1}, m_{p2}) = [m_{p1}][m_{p2}][y_{p1} | m_{p1}][y_{p2} | m_{p2}] \times \prod_{i=1}^n R_{gp1p2i}(g_{p1}, g_{p2}, m_{p1}, m_{p2}) \quad (6)$$

Finally peel m_{p2} :

$$R_{mp1}(m_{p1}) = \sum_{m_{p2}} [m_{p2}] I(f_2(m_{p2}) = y_{p2}) R_{gp1p2i}(g_{p1}, g_{p2}, m_{p1}, m_{p2}) \quad (7)$$

$$f(m_{p1}) = [m_{p1}][y_{p1} | m_{p1}] R_{mp1}(m_{p1}) \quad (8)$$

Peeling—small example

Blood example (fictitious).

Diploid: 12×34

- 6 markers, alleles A,B,O.

The map file

```
> system("cat blood-data/blood_m6_map.dat")
nloci=6
comment="#"
field_separator=":"
# fictitious blood group data
read=6
0:m0:A:0
1:m1:A:10
2:m2:A:15
3:m3:A:20
4:m4:A:25
5:m5:A:30
```

9

Peeling run (stand-alone program)

See file blood_n4m6.out.

Usage: peeling map-file marker-file > output-file

```
> cd visg-course2012/blood-data
> ../software/bin/peeling blood_m6_map.dat blood_n4m6.dat \
  > blood_n4m6.out
```

11

The marker data file

```
> system("cat blood-data/blood_n4m6.dat")
ploidy=2
nloci=6
nparents=2
nprogeny=4
null_allele="0"
na_phenotype="*"
na_allele="*"
dominant_allele=NULL
allele_separator=","
comment="#"
# blood fictitious marker data
# 6 individuals (2 parents+4progeny)

read=6
# locus 0 @0cM
m0;A;A,B;A;B;A;B
m1;A;A;0;A;A;0
m2;A;B;0;A;B;A,B
m3;*;*;0;A;B;A,B
m4;A;B;0;A;B;A
m5;*;*;0;A;B;A
```

10

Peeling, R/polyploids package

```
> library(polyploids) # may be included in BayesQTLBIC
> # read map file
> blood.map <- polyploids.read.map.file("blood-data/blood_m6_map.dat")
> blood.map
  index marker chrom pos.cM
1     0     m0     A     0
2     1     m1     A    10
3     2     m2     A    15
4     3     m3     A    20
5     4     m4     A    25
6     5     m5     A    30
```

12

Peeling, R/polyploids package

```
> # read marker file
> blood.markers <- polyploids.read.marker.file("blood-data/blood_n4m6.dat")
> # do peeling
> blood.pl <- polyploids.peeling(blood.map, blood.markers, nsamples=10,
+   min.prob=5e-4, seed=12345, verbose=1)
> # align samples
> blood.pl <- align.peeling(blood.pl)
```

13

polyploids_talk_VISG.tex RCS:1.0

19/10/2012

VISG Workshop: Polyploids

```
1: A
2: A,B
number of progeny = 4
lclp->n_unique_phenotypes = 4
table of progeny phenotypes
i:      yp,      count
0:      A,        2
1:      A,B,       0
2:      B,        2
3:      O,        0
lclp->n_genotypes = 9
marginal probabilities for parental genotypes
i:      mg,      yp,      p1,      p2
0:      A,A,      A, 0.0000, 0.0000
1:      A,B,      A,B, 0.0000, 0.5000
2:      A,O,      A, 0.5000, 0.0000
3:      B,A,      A,B, 0.0000, 0.5000
. . .
4:      B,B,      B, 0.0000, 0.0000
5:      B,O,      B, 0.0000, 0.0000
. . .
6:      O,A,      A, 0.5000, 0.0000
7:      O,B,      B, 0.0000, 0.0000
8:      O,O,      O, 0.0000, 0.0000
marginal probabilities for parental genotype pairs
```

polyploids_talk_VISG.tex RCS:1.0

19/10/2012

Blood example, single marker peeling

```
allopolyplods peeling: (C) R.D. Ball, Scion/VISG 2011, 2012
6 loci, 2 parental phenotypes, 4 progeny, ploidy = 2, seed = 12345
6 loci single marker peeling, 6 loci HMM peeling
*** single marker peeling for loci:0,1,2,3,4,5
0: 1 3
1: 2 3
2: 1 4
3: 2 4

single marker peeling:
*** single marker peeling for locus[0] m0 ***

*
*** summary data for locus after single marker peeling ***
*

alleles:
0: A
1: B
2: O
parental phenotypes:
```

14

polyploids_talk_VISG.tex RCS:1.0

19/10/2012

VISG Workshop: Polyploids

```
( 2, 1):      A,O,      A,B,      A, 0.25000000
( 2, 3):      A,O,      B,A,      A, 0.25000000
( 6, 1):      O,A,      A,B,      A, 0.25000000
( 6, 3):      O,A,      B,A,      A, 0.25000000
locus probs (logfmpimp2)
:n_genotypes=9, max_log = 0.000000, min_log = -18.420681
(2,1): -8.553
(2,3): -8.553
(6,1): -8.553
(6,3): -8.553
*** single marker peeling for locus[1] m1 ***
.
.
.
```

polyploids_talk_VISG.tex RCS:1.0

19/10/2012

Peeling strategy

Brute force peeling can handle ~ 5 markers in diploids.

Our strategy:

1. single marker peeling
2. identify errors
3. HMM peeling (2 loci to start with)

15

polyploids_talk_VISG.tex RCS:1.0

19/10/2012

VISG W

Single marker peeling

Peel individual markers separately:

1. For $t = 1, \dots, T$
2. Peel marker t
3. Fix genotype errors, repeat step 2 if necessary
4. Eliminate low probability parental combinations
5. Store `parental_combinations(t)` and probabilities



HMM peeling

Iterative strategy, $O(N^2T^2)$ per progeny:

1. Let $t \leftarrow 2, \text{min_prob} > 0$, e.g. $\text{min_prob} = 5 \times 10^{-4}$.
2. For $p \in \text{parental_combinations}(1, \dots, t-1) \times \text{parental_combinations}(t)$
3. do HMM for loci $1 \dots t$ for each progeny conditional on p
4. calculate $\text{Pr}(p) = \text{marginal probability from HMM}$
5. eliminate low probability combinations ($\text{Pr}(p) < \text{min_prob}$).
6. Store `parental_combinations(1, \dots, t)` and probabilities
7. Increment t and repeat steps 2-6 until $t = T$
8. Sample parental combinations
9. Repeat HMM to sample progeny virtual marker genotypes (g_{ci} in Fig. 1)
10. Align marker phases on samples, group into alignment classes C .

16

polyploids_talk

4. compute probabilities for parental combinations, store non-negligible combinations
5. adding one marker at a time, repeat steps 3,4 for stored parental combinations combined with possibilities for the new marker

polyploids_talk_VISG.tex RCS:1.0

19/10/2012

VISG Workshop: Polyploids

Clover data

White clover (*Trifolium repens*).

Allo-tetraploid: 1234×5678 .

- Allo-tetraploid
- $n = 182$ progeny
- Chromosomes A1A2 ... H1H2

17

polyploids_talk_VISG.tex RCS:1.0

19/10/2012

- Up to about 26 markers per chromosome
- diploidised data from Brent
- psuedo-SSR marker calls generated to represent likely raw polyploid data
- traits measured at Lincoln
- seed yield, years 2002–2004 (plus others)

Clover C1C2 peeling (within R)

```
> library(polyploids)
> # read map file
> chromC.map.df <- polyploids.read.map.file(map.file=
+   "clover-data/clover_C1C2_m17_map_v2.dat")
> # read marker file
> chromC.marker.data <- polyploids.read.marker.file(marker.file=
+   "clover-data/clover_C1C2n182m17_v2.dat")
> # do peeling
> date()
> # 0:30
> set.seed(12345)
> chromC.peeling <- polyploids.peeling(chromC.map.df,
+   chromC.marker.data, nsamples=10,min.prob=5.0e-4,
+   seed=12345, verbose=1)
> date()
> save(chromC.peeling, file="clover-data/clover_chromC.peeling_v2b.rda")
```

Peeling run (stand-alone program)

Usage: peeling map-file marker-file > output-file

e.g.

```
> ./peeling clover_A1A2_m25_map.dat clover_A1A2n182m25.dat \
> A1A2_peeling.out
```

Clover C1C2 alignment

Note: 1 alignment class.

```
> pl1C <- align.peeling(chromC.peeling)
chr [1:3128] "0,672" "678,676" "0,672,676" "678,676" "-999" ..
.
.
.
> table(pl1C$alignment$alignment.class)

1
10
```

Single marker peeling for chromosome A1A2

Table 1. Single marker peeling for chromosome A1A2.

locus	phenotype	genotype [†]	p1		p2	
			prob. × nreps [‡]	nreps [‡]	prob. × nreps [‡]	nreps [‡]
0. prs753z	753,755	0,0,753,755	0.1462 × 4	4	0.1462 × 4	4
	753,755	0,753,0,755	0.0497 × 8	8	0.0497 × 8	8
	753,755	0,753,753,755	0.0022 × 8	8	0.0022 × 8	8
1. prs089bx	95	0,0,0,95	0.25 × 4	4	—	—
	91,97	0,91,0,97	—	—	0.125 × 8	8
2. prs685x	685	0,0,0,685	—	—	0.25 × 4	4
	687	0,0,0,687	0.25 × 4	4	—	—
3. ats032x1x2	32,40,42	0,32,40,42	0.125 × 8	8	—	—
	36,44	0,36,0,44	—	—	0.125 × 8	8
4. prs10aex	20	0,0,0,20	0.25 × 4	4	—	—
	10,18,22	0,10,18,22	—	—	0.125 × 8	8
5. prs112x	112,114	0,0,112,114	0.25 × 4	4	—	—
	116,118	0,0,116,118	—	—	0.25 × 4	4
6. prs396d	0	0,0,0,0	—	—	1.0 × 1	1
	402	0,0,0,402	0.25 × 4	4	—	—
9. prs504b	0	0,0,0,0	1.0 × 1	1	—	—
	506	0,0,0,506	—	—	0.25 × 4	4
10. prs497db	0	0,0,0,0	—	—	1.0 × 1	1
	499,503	0,499,0,503	0.125 × 8	8	—	—
23. prs200x1x2	200,202,212,214	200,212,202,214	0.125 × 8	8	—	—
	204,206,208,210	204,210,206,208	—	—	0.125 × 8	8
24. prs380xc	380,390	0,0,380,390	0.25 × 4	4	—	—
	382,384,388	0,382,384,388	—	—	0.1250 × 8	8

21

[†]Ordered genotypes, only one representative per alignment class per genotype shown.

[‡] Number of replicate genotypes equivalent up to alignment permutation, sharing the same probability.

Clover C1C2 parental combinations

Look at parental combinations from peeling (sample 1).

```
> C1C2.Xc <- polyploids.calc.Xc(p11C)
> postpeeling4(p11C, C1C2.Xc, plot=TRUE)
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]
[1,] "0"  "0"  "0"  "672" "678" "0"  "676" "0"
[2,] "619" "617" "625" "0"  "627" "0"  "621" "629"
[3,] "540" "538" "0"  "0"  "538" "540" "0"  "0"
[4,] "0"  "0"  "0"  "0"  "269" "0"  "0"  "0"
[5,] "639" "645" "0"  "643" "0"  "0"  "0"  "0"
[6,] "616" "618" "0"  "0"  "614" "620" "0"  "0"
[7,] "0"  "0"  "369" "367" "0"  "0"  "367" "369"
```

22


```
[8,] "0" "0" "0" "16" "0" "0" "0" "18"
[9,] "0" "0" "396" "0" "0" "0" "398" "0"
[10,] "0" "490" "0" "0" "0" "0" "0" "0"
[11,] "0" "0" "545" "543" "0" "0" "0" "547"
[12,] "0" "0" "0" "0" "711" "0" "0" "0"
[13,] "109" "107" "0" "0" "0" "0" "0" "0"
[14,] "516" "0" "0" "0" "0" "0" "0" "0"
[15,] "0" "296" "0" "0" "0" "298" "0" "0"
[16,] "736" "0" "0" "0" "0" "730" "0" "0"
[17,] "0" "737" "0" "0" "0" "739" "0" "745"
```

Clover C1C2 progeny gci's

Look at 1st 6 markers as 4 column array:

```
> # a look at peeling output (gci's)
> # ploidy 4 x 182 progeny x 17 markers x 10 samples
> dim(pl1C$progeny.gci)
[1] 4 182 17 10
> # 1st 6 markers, progeny 1
> # 4->3 => recombination between m1 and m2 on parent 1, C2
> # 6->5 => recombination between m3 and m4 on parent 2, C1
> matrix(aperm(pl1C$progeny.gci,c(1,3,2,4)), ncol=4,byrow=TRUE)[1:6,]
[,1] [,2] [,3] [,4]
[1,] 1 6 4 7
[2,] 1 6 3 7
[3,] 1 6 3 7
[4,] 1 5 3 7
[5,] 1 5 3 7
[6,] 1 5 3 7
```

Clover C1C2 progeny gci's

Cross is 1234 x 5678.

Allo-tetraploids progeny inherit 1 or 2 and 3 or 4 and 5 or 6 and 7 or 8, e.g. 1537, 2548, 2647, 1458.

1 vs 2 is parent 1 subgenome 1
 3 vs 4 is parent 1 subgenome 2
 5 vs 6 is parent 2 subgenome 1
 7 vs 7 is parent 2 subgenome 2

```
> # 1st 6 markers, progeny 2
> # no recombinations
> matrix(aperm(pl1C$progeny.gci,c(1,3,2,4)), ncol=4,byrow=TRUE)[17+1:6,]
[,1] [,2] [,3] [,4]
[1,] 1 5 3 7
[2,] 1 5 3 7
[3,] 1 5 3 7
[4,] 1 5 3 7
[5,] 1 5 3 7
[6,] 1 5 3 7
> # 1st 6 markers, progeny 3
> # 5->6 => recombination between m3 and m4 on parent 2, C1
> matrix(aperm(pl1C$progeny.gci,c(1,3,2,4)), ncol=4,byrow=TRUE)[2*17+1:6,]
[,1] [,2] [,3] [,4]
[1,] 1 5 4 7
[2,] 1 5 4 7
[3,] 1 5 4 7
[4,] 1 6 4 7
[5,] 1 6 4 7
[6,] 1 6 4 7
```

Clover E1E2 alignment

Note: 2 alignment classes.

```
> # Use saved peeling rda file
> attach("clover-data/clover_chromC.peeling_v2.rda")
> ls(pos=2)
[1] "chromE.peeling"
> pl1E <- align.peeling(chromE.peeling)
chr [1:3128] "37" "0" "37" "37" "0" "37" "37" "-999" ...
Read 184 items
.
.
.
```

26

VISG Workshop: Polyploids

```
[1,] "0" "0" "1" "0" "0" "0" "0" "0"
[2,] "0" "1" "0" "0" "0" "0" "0" "0"
.
.
.
.unaligned parental genotypes:
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]
[1,] "0" "0" "0" "1" "0" "0" "0" "2"
[2,] "0" "0" "0" "1" "0" "0" "2" "0"
permutations aligning loci 1, ... , 6
< nil >
sample ** 2 ** could not be aligned at locus: 6
samples aligned at loci 1 2 3 4 5
[1] 1 2
```

```
*** initial phase alignment for sample 2 ***
locus: 1
.unaligned parental genotypes:
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]
[1,] "0" "0" "0" "1" "0" "0" "0" "0"
[2,] "1" "0" "0" "0" "0" "0" "0" "0"
.
.
.
permutations aligning loci 1, ... , 16
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]
[1,] 3 4 2 1 8 7 6 5
locus: 17
.unaligned parental genotypes:
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]
```

VISG Workshop: Polyploids

```
alignment.class
1 2
9 1
.
.
.
> table(pl1E$alignment$alignment.class)

1 2
9 1
```

Theory

Ball Genetics 2001, 2007. R package BayesQTLBIC.

- X matrix of marker contrasts
- multiple models \leftrightarrow selected sets of columns of X
- prior probabilities proportional to marker spacing, expected number of QTL
- approximate probabilities for models from BIC criterion

27

28

polyploids_talk_VISG.tex RCS:1.0

19/10/2012

polyploids_talk_VISG.tex RCS:1.0

19/10/2012

Theory

Interpretation:

- consider models \mathcal{M}_γ according to posterior probabilities

$$\mathcal{M}_\gamma : y = X_\gamma b_\gamma + e \quad (9)$$

$$\Pr(\mathcal{M}_\gamma | y) \approx \exp(-\text{BIC}/2) \times \text{prior} \quad (10)$$

- marker m_i selected in $\mathcal{M}_\gamma \leftrightarrow$ QTL in vicinity of m_i .
- $b_{\gamma,i}$ is the causal effect of m_i assuming \mathcal{M}_γ is the true model.
- compute model averaged quantities
- $\Pr(\text{QTL} \in \text{region})$

29

polyploids_talk_VISG.tex RCS:1.0

19/10/2012

polyploids_talk_VISG.tex RCS:1.0

19/10/2012

- probabilities for number of QTL

BayesQTLBIC R package

Key functions:

- `bicreg.qtl()`: analysis for a set of loci, e.g. 1 chromosome
- `bicreg.models()`: combine analyses from `bicreg.qtl()`, e.g. from separate chromosomes
- `summary()`: summary method

VIS

BIC method for polyploids QTL mapping

- R package BayesQTLBIC (Ball 2001, 2009).
- Poisson prior, $n_{QTL} \sim \text{Poisson}(\lambda_Q)$, e.g. $\lambda_Q = 10$
- Prior probability per marker is proportional to marker spacing, λ_Q
- Populate model matrix $X^{(s)}$ with contrasts from virtual marker genotypes from peeling samples $s = 1, \dots, n_s$
- Use multiple imputation for each alignment class \mathcal{C}
 - + combine X^s , for $s \in \mathcal{C}$ by row, to form X .
 - + choose search strategy e.g.
 - exhaustive by chromosome/subgenome/parent, or,
 - exhaustive by genomic window/subgenome/parent, and/or
 - limit model size, e.g. ≤ 3 QTL per chromosome
 - + search through space of models $\mathcal{M}_\gamma : y = X_\gamma \beta_\gamma + \epsilon$,
 $X_\gamma \leftrightarrow$ selected columns of $X \leftrightarrow$ putative QTL
 - + compute: $\Pr(\mathcal{M}_\gamma) \propto \exp(-\text{BIC}/2) \times \text{prior}$, for models
 - + compute summaries of interest e.g.

$$\Pr(\text{QTL in region}) = \sum \{\Pr(\mathcal{M}_\gamma) \text{ with QTL in region}\} \quad (1)$$

- + sample from models for each chromosome and recalculate probabilities for a 'genome-wide' analysis (optional).

- Combine information across alignment classes

VISG Workshop: Polyploids

Clover QTL mapping data

```
> clover.traits.n182.df[1:5,c(1,4:6)]
      Entry LR02_SEEDWGT LR03_SEEDWGT LR04_SEEDWGT
1 clover001          7.17          5.93          2.40
2 clover002          11.60          6.07          2.42
3 clover003          16.17          19.63         11.70
4 clover004          15.67          10.67          4.67
5 clover005          23.07          13.70          6.17
> # seed weight traits
> Y <- clover.traits.n182.df[,c("LR02_SEEDWGT",
+                               "LR03_SEEDWGT", "LR04_SEEDWGT")]
> seedwt <- apply(Y,1,mean)
> seedwt02 <- clover.traits.n182.df[, "LR02_SEEDWGT"]
> seedwt04 <- clover.traits.n182.df[, "LR04_SEEDWGT"]
```

QTL analysis for clover C1C2— Separate analysis by subgenome/parent.

- Key function bicreg.qtl.
- Uses pl1C from peeling.

```
> library(BayesQTLBIC)
> nloci <- length(pl1C$locus.names)
> nsamples <- 10
> # calculate prior
> C1C2.prior <- polyploids.calc.prior(pos.cM=pl1C$locus.pos.cM,
+   prior.rate=10, genome.length.cM=591)
```

33

polyploids_talk_VISG.tex RCS:1.0

19/10/2012

```
> # calculate X-matrix, contrasts from progeny
> # virtual marker genotypes (gci)
> # separate analyses by chromosome/parent/subgenome
> # i.e. virtual marker genotypes
> C1C2.Xc <- polyploids.calc.Xc(pl1C)
> # 1 vs. 2 (p1, C1)
> C1C2.seedwt02.res1 <- bicreg.qtl(C1C2.Xc[,1:nloci],
+   y=rep(seedwt02,nsamples), num.imputations=nsamples,
+   prior=C1C2.prior)
> # 5 vs. 6 (p2, C1)
> C1C2.seedwt02.res2 <- bicreg.qtl(C1C2.Xc[,nloci+1:nloci],
+   y=rep(seedwt02,nsamples), num.imputations=nsamples,
+   prior=C1C2.prior)
```

polyploids_talk_VISG.tex RCS:1.0

19/10/2012

```
> # 3 vs. 4 (p1, C2)
> C1C2.seedwt02.res3 <- bicreg.qtl(C1C2.Xc[,2*nloci+1:nloci],
+   y=rep(seedwt02,nsamples), num.imputations=nsamples,
+   prior=C1C2.prior)
> # 7 vs. 8 (p2, C2)
> C1C2.seedwt02.res4 <- bicreg.qtl(C1C2.Xc[,3*nloci+1:nloci],
+   y=rep(seedwt02,nsamples), num.imputations=nsamples,
+   prior=C1C2.prior)
```

polyploids_talk_VISG.tex RCS:1.0

19/10/2012

QTL analysis for clover C1C2—summary output

- Look at top 10 models with summary()
- Note low probability 3.7×10^{-7} for model size 0 \Rightarrow strong evidence for QTL.

```
> summary(C1C2.seedwt02.res4, nbest=10)
R-squared, BIC, and approximate posterior probabilities for individual models
  X4m1 X4m2 X4m3 X4m4 X4m5 X4m6 X4m7 X4m8 X4m9 X4m10 X4m11 X4m12 X4m13 X4m1
1    0    0    0    0    0    0    0    1    0    0    0    0    0    0
2    0    0    0    0    0    0    0    1    0    0    0    0    0    0
3    0    0    0    0    0    0    0    1    0    0    0    0    0    0
4    0    0    0    0    0    0    0    1    0    0    0    0    0    0
5    0    0    0    0    0    0    0    1    0    0    0    0    0    0
6    0    0    0    0    0    0    0    1    0    0    0    0    0    1
7    0    0    0    0    0    0    0    1    0    0    0    1    0    0
```

34

polyploids_talk_VISG.tex RCS:1.0

19/10/2012

```

8 0 0 0 0 0 0 1 0 0 1 0 0 0
9 0 0 0 0 0 0 1 1 0 0 0 0 0
10 0 0 0 0 0 0 1 0 0 0 0 1 0
   X4m15 X4m16 X4m17      R2      BIC  postprob  cumprob
1 0 0 0 18.961 -25.34475 0.50292851 0.5029285
2 0 1 0 22.905 -23.47481 0.19744843 0.7003769
3 0 0 1 22.946 -21.60274 0.07743553 0.7778125
4 1 0 0 21.573 -21.07197 0.05938596 0.8371984
5 0 0 0 21.299 -20.78715 0.05150353 0.8887020
6 0 0 0 20.874 -19.62150 0.02875532 0.9174573
7 0 0 0 20.748 -18.79274 0.01900009 0.9364574
8 0 0 0 20.661 -18.60854 0.01732837 0.9537858
9 0 0 0 19.760 -17.96640 0.01256954 0.9663553
10 0 0 0 20.636 -17.77004 0.01139409 0.9777494

```

polyploids_talk_VISG.tex RCS:1.0

19/10/2012

QTL analysis for clover C1C2—Bayes factors by chromosome

$$B = \frac{\Pr(H_1 | y)}{\Pr(H_0 | y)} \times \frac{\pi(H_0)}{\pi(H_1)} \quad (11)$$

```

> calc.bf.chrom(pH0=C1C2.seedwt02.res1$postprob.size["0"], prior=C1C2.prior)
0
0.4071328
> calc.bf.chrom(pH0=C1C2.seedwt02.res2$postprob.size["0"], prior=C1C2.prior)
0
0.04090569
> calc.bf.chrom(pH0=C1C2.seedwt02.res3$postprob.size["0"], prior=C1C2.prior)
0
0.02443742
> calc.bf.chrom(pH0=C1C2.seedwt02.res4$postprob.size["0"], prior=C1C2.prior)
0
829605.7

```

polyploids_talk_VISG.tex RCS:1.0

19/10/2012

marginal probabilities for model sizes

```

0 1 2 3 4
3.720903e-07 5.031256e-01 4.846549e-01 1.212378e-02 9.534188e-05

```

marginal probabilities for individual variables

```

X4m7 X4m14 X4m15 X4m16 X4m17
0.97774939 0.05150353 0.05938596 0.19744843 0.07743553

```

attr,"prior")

```
[1] 0.07380106 0.11930084 0.14852022 0.12570141 0.07888157 0.07290340
```

```
[7] 0.08218792 0.12038065 0.06696074 0.06324747 0.06279026 0.04369076
```

```
[13] 0.08065201 0.08780076 0.07476120 0.05349705 0.02068211
```

attr,"intercept")

```
[1] TRUE
```

polyploids_talk_VISG.tex RCS:1.0

19/10/2012

Combined analysis with bicreg.models()

- Sample models from each fit.
- Combine samples across fits.
- Re-evaluate probabilities

```

> # combined analysis using bicreg.models()
> C1C2.seedwt02.fits <- list(C1C2.seedwt02.res1,C1C2.seedwt02.res2,
+   C1C2.seedwt02.res3,C1C2.seedwt02.res4)
> nsim <- 1000
> mWhich <- sample.bicreg.qtl.models(C1C2.seedwt02.fits,nsim=nsim)
> date() # 4 mins.
> C1C2.seedwt02.mbic <- bicreg.models(x=C1C2.Xc,y=rep(seedwt02,nsamples),
+   which=mWhich,prior=C1C2.prior,num.imputations=nsamples)
> date()

```

polyploids_talk_VISG.tex RCS:1.0

19/10/2012

VISG Workshop: Polyploids

```
> attach("./clover-data/C1C2.seedwt02.mbic.rda")
> summary(C1C2.seedwt02.mbic,nbest=20,min.marker.prob=0.05)
R-squared, BIC, and approximate posterior probabilities for individual models
  1m1 1m2 1m3 1m4 1m5 1m6 1m7 1m8 1m9 1m10 1m11 1m12 1m13 1m14 1m15 1m16 1m
1    0  0  0  0  0  1  0  0  0  0  0  0  0  0  0  0  0
2    0  0  0  0  0  1  0  0  0  0  0  0  0  0  0  0  0
3    0  0  0  0  0  0  0  0  0  0  0  0  0  1  0  0  0
4    0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
.
.
.
  2m1 2m2 2m3 2m4 2m5 2m6 2m7 2m8 2m9 2m10 2m11 2m12 2m13 2m14 2m15 2m16 2m
1    0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
2    0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
3    0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
4    0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
.
.
.
  3m1 3m2 3m3 3m4 3m5 3m6 3m7 3m8 3m9 3m10 3m11 3m12 3m13 3m14 3m15 3m16 3m
```

37

polyploids_talk_VISG.tex RCS:1.0

19/10/2012

VISG Workshop: Polyploids

```
1    0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
2    0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
3    0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
4    0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
.
.
.
  4m1 4m2 4m3 4m4 4m5 4m6 4m7 4m8 4m9 4m10 4m11 4m12 4m13 4m14 4m15 4m16 4m
1    0  0  0  0  0  0  1  0  0  0  0  0  0  0  0  0  1
2    0  0  0  0  0  0  1  0  0  0  0  0  0  0  0  0  0
3    0  0  0  0  0  0  1  0  0  0  0  0  0  0  0  0  0
4    0  0  0  0  0  0  1  0  0  0  0  0  0  0  0  0  0
.
.
.
      R2      BIC      postprob      cumprob
1  28.24 -17.74027 0.072279168 0.07227917
2  23.64 -17.38259 0.060442644 0.13272181
3  23.51 -17.30671 0.058192434 0.19091425
4  18.96 -16.67581 0.042449194 0.23336344
5  27.79 -16.43196 0.037576699 0.27094014
```

polyploids_talk_VISG.tex RCS:1.0

19/10/2012

VISG Workshop: Polyploids

```
6 23.11 -16.17196 0.032995904 0.30393604
7 23.44 -16.07551 0.031442334 0.33537838
8 23.12 -15.97683 0.029928696 0.36530707
9 28.28 -15.87287 0.028412673 0.39371975
10 27.48 -15.87111 0.028387695 0.42210744
11 27.39 -15.83084 0.027821917 0.44992936
12 22.90 -14.79631 0.016585995 0.46651535
13 27.39 -14.76612 0.016337527 0.48285288
14 26.70 -14.59058 0.014964671 0.49781755
15 27.82 -14.53871 0.014581578 0.51239913
16 22.41 -14.35808 0.013322358 0.52572149
17 23.41 -14.03532 0.011336917 0.53705840
18 27.44 -13.98734 0.011068130 0.54812653
19 27.50 -13.95243 0.010876628 0.55900316
20 26.22 -13.75258 0.009842327 0.56884549
marginal probabilities for model sizes
      0      1      2      3      4
3.144121e-08 4.244919e-02 3.102480e-01 5.287792e-01 1.178055e-01 7.178861e-0
6
2.409075e-07
marginal probabilities for individual variables
```

polyploids_talk_VISG.tex RCS:1.0

19/10/2012

VISG Workshop: Polyploids

```
      1m5      1m6      1m13      1m14      4m7      4m16      4m17
0.18594148 0.08208697 0.07226023 0.09708248 0.56884549 0.19898900 0.06493901
attr,"prior")
[1] 0.07380106 0.11930084 0.14852022 0.12570141 0.07888157 0.07290340
[7] 0.08218792 0.12038065 0.06696074 0.06324747 0.06279026 0.04369076
[13] 0.08065201 0.08780076 0.07476120 0.05349705 0.02068211
attr,"intercept")
[1] TRUE
```

polyploids_talk_VISG.tex RCS:1.0

19/10/2012

Combined analysis with bicreg.pppeeling()

```
> cloverC.seedwt02.bic <- bicreg.pppeeling(y=seedwt02, pl1C,
+     prior.rate=20, genome.length.cM=591, nsim=100)
> summary(cloverC.seedwt02.bic, nbest=38)
R-squared, BIC, and approximate posterior probabilities for individual models
  1m1 1m2 1m3 1m4 1m5 1m6 1m7 1m8 1m9 1m10 1m11 1m12 1m13 1m14 1m15 1m16 1m
1    0  0  0  0  1  0  0  0  0  0  0  0  0  0  0  0
2    0  0  0  0  0  1  0  0  0  0  0  0  0  0  0  0
3    0  0  0  0  1  0  0  0  0  0  0  0  0  0  0  0
.
.
.
  2m1 2m2 2m3 2m4 2m5 2m6 2m7 2m8 2m9 2m10 2m11 2m12 2m13 2m14 2m15 2m16 2m
1    0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
2    0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
3    0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
.
.
.
  3m1 3m2 3m3 3m4 3m5 3m6 3m7 3m8 3m9 3m10 3m11 3m12 3m13 3m14 3m15 3m16 3m
```

38

```
1    0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
2    0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
3    0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
      4m1 4m2 4m3 4m4 4m5 4m6 4m7 4m8 4m9 4m10 4m11 4m12 4m13 4m14 4m15 4m16 4m
1    0  0  0  0  0  0  0  1  0  0  0  0  0  0  0  0  1
2    0  0  0  0  0  0  0  1  0  0  0  0  0  0  0  0  1
3    0  0  0  0  0  0  0  1  0  0  0  0  0  0  0  0  0
.
.
.
      R2      BIC      postprob      cumprob
1  28.24 -10.570756  0.153251940  0.1532519
2  27.79  -9.255728  0.079405630  0.2326576
3  23.64  -8.771042  0.062316521  0.2949741
4  23.51  -8.705317  0.060301906  0.3552760
5  27.39  -8.671485  0.059290438  0.4145664
6  27.39  -7.568496  0.034156517  0.4487230
7  23.11  -7.562423  0.034052963  0.4827759
8  26.70  -7.444544  0.032103893  0.5148798
9  30.71  -7.437330  0.031988307  0.5468681
10 23.44  -7.435846  0.031964576  0.5788327
```

```
11 23.12 -7.358564 0.030752992 0.6095857
12 27.82 -7.327749 0.030282796 0.6398685
13 27.44 -6.793248 0.023180917 0.6630494
14 27.50 -6.750191 0.022687206 0.6857366
15 30.06 -6.647346 0.021550053 0.7072867
16 18.96 -6.594120 0.020984109 0.7282708
17 26.77 -6.371281 0.018771614 0.7470424
18 26.53 -6.163801 0.016921855 0.7639642
19 22.90 -6.156653 0.016861480 0.7808257
20 26.03 -5.838770 0.014383628 0.7952093
21 30.16 -5.821093 0.014257058 0.8094664
22 27.43 -5.665178 0.013187828 0.8226542
23 29.73 -5.590516 0.012704594 0.8353588
24 30.75 -5.538818 0.012380401 0.8477392
25 26.74 -5.510335 0.012205331 0.8599445
26 25.69 -5.124997 0.010066407 0.8700110
27 29.00 -4.950675 0.009226162 0.8792371
28 22.05 -4.527381 0.007466273 0.8867034
29 26.81 -4.467113 0.007244641 0.8939480
30 25.47 -4.415860 0.007061344 0.9010094
31 25.28 -4.367444 0.006892456 0.9079018
```

```
32 22.95 -4.271112 0.006568342 0.9144702
33 26.57 -4.259309 0.006529692 0.9209999
34 25.60 -4.239187 0.006464326 0.9274642
35 22.25 -4.192555 0.006315349 0.9337795
36 28.69 -4.030877 0.005824912 0.9396045
37 21.57 -3.782131 0.005143690 0.9447481
38 25.18 -3.702885 0.004943867 0.9496920
marginal probabilities for model sizes
      0      1      2      3      4
7.438204e-09 2.098411e-02 2.747538e-01 5.816783e-01 1.214279e-01 1.155234e-0
6
7.114546e-07
marginal probabilities for individual variables
      1m5      1m6      1m13      1m14      1m16      4m7      4m15
0.35847453 0.15964234 0.09267385 0.14966572 0.14439901 0.94969201 0.05163121
      4m16      4m17
0.48641589 0.12206182
attr,"prior")
[1] 0.14215551 0.22436898 0.27498219 0.23560198 0.15154083 0.14049189
[7] 0.15762098 0.22626980 0.12943773 0.12249470 0.12163790 0.08547263
[13] 0.15479928 0.16789255 0.14393316 0.10413216 0.04093647 0.14215551
```



```
[19] 0.22436898 0.27498219 0.23560198 0.15154083 0.14049189 0.15762098
[25] 0.22626980 0.12943773 0.12249470 0.12163790 0.08547263 0.15479928
[31] 0.16789255 0.14393316 0.10413216 0.04093647 0.14215551 0.22436898
[37] 0.27498219 0.23560198 0.15154083 0.14049189 0.15762098 0.22626980
[43] 0.12943773 0.12249470 0.12163790 0.08547263 0.15479928 0.16789255
[49] 0.14393316 0.10413216 0.04093647 0.14215551 0.22436898 0.27498219
[55] 0.23560198 0.15154083 0.14049189 0.15762098 0.22626980 0.12943773
[61] 0.12249470 0.12163790 0.08547263 0.15479928 0.16789255 0.14393316
[67] 0.10413216 0.04093647
attr,"intercept")
[1] TRUE
```



Acknowledgements

Thanks to Brent Barrett for providing the clover data, Gail Timmerman-Vaughan for generating the pseudo SSR data and for testing and feedback on the R package.

This project was funded by the New Zealand Foundation for Research, Science, and Technology through a grant to the New Zealand Virtual Institute for Statistical Genetics (VISG).

