VISG Virtual Institute of Statistical Genetics

Hidden Markov Model (HMM) peeling and the alignment problem for QTL mapping in allo-polyploids

Rod Ball Scion

Gail Timmerman-Vaughan

Plant and Food

NZ Molecular Mapping meeting,
Havelock North, 27–28 Oct 2011

ABSTRACT

We update developments of statistical methods and software for peeling in allo-polyploids, including extension to HMM peeling and incorporating alignment permutations into the peeling algorithms.

Allo-polyploids are polyploids with several distinct sub-genomes, (e.g. of the form AABBCC which could have arisen from fusion of several ancestral species) where the identity of sub-genomes is preserved in meiosis: chromosomes assort only with their homologues in the same sub-genome.

Previously we were using 'peeling' and 'conditional peeling' to sample from the space of possible genotypes of sets of fully informative 'virtual markers', for progeny in an allo-polyploid QTL mapping family. The virtual marker genotypes are intended to serve as input data to our Bayesian QTL mapping method. This would be a general approach to QTL mapping in polyploids and the first multi-locus Bayesian QTL mapping method for polyploids that takes into account model uncertainty and gives posterior probabilities for QTL to be segregating in a region.

However there was an 'alignment problem' with conditional peeling that was apparent in the clover tetraploid datasets — different samples could not be aligned meaning that QTL in different replicate samples may be in different phases and QTL effects would cancel.

To address this problem we have developed HMM peeling which uses the fact that each progeny's genotypes conditional on the parental genotypes are a Hidden Markov model. HMM peeling combined with an approximate probability calculation strategy for parental genotype combinations is designed to keep computing times from growing exponentially in the number of markers in a linkage group.

Talk outline

- Introduction
- Hidden Markov models
  - forwards recursion
  - backwards sampling
- HMM for polyploids

Talk outline

- Alignment
  - alignment permutations
  - examples
  - the alignment problem
- Implementation progress
- Results
- Next steps

Introduction

VISG polyploids project developing peeling for allo-polyploids.

- estimate marginal probabilities for parental genotypes
- generate samples of fully informative 'virtual marker genotypes' for progeny
- use in Bayesian QTL mapping method (Ball 2001)

Peeling—previous work

Previously

- single marker peeling
- conditional peeling
  - sampling one marker at a time conditional on previous
- joint peeling
  - sampling a small set of markers jointly

Peeling—this year

- Attempted to use conditional peeling results with tetraploid clover data

- Found limitations of conditional peeling

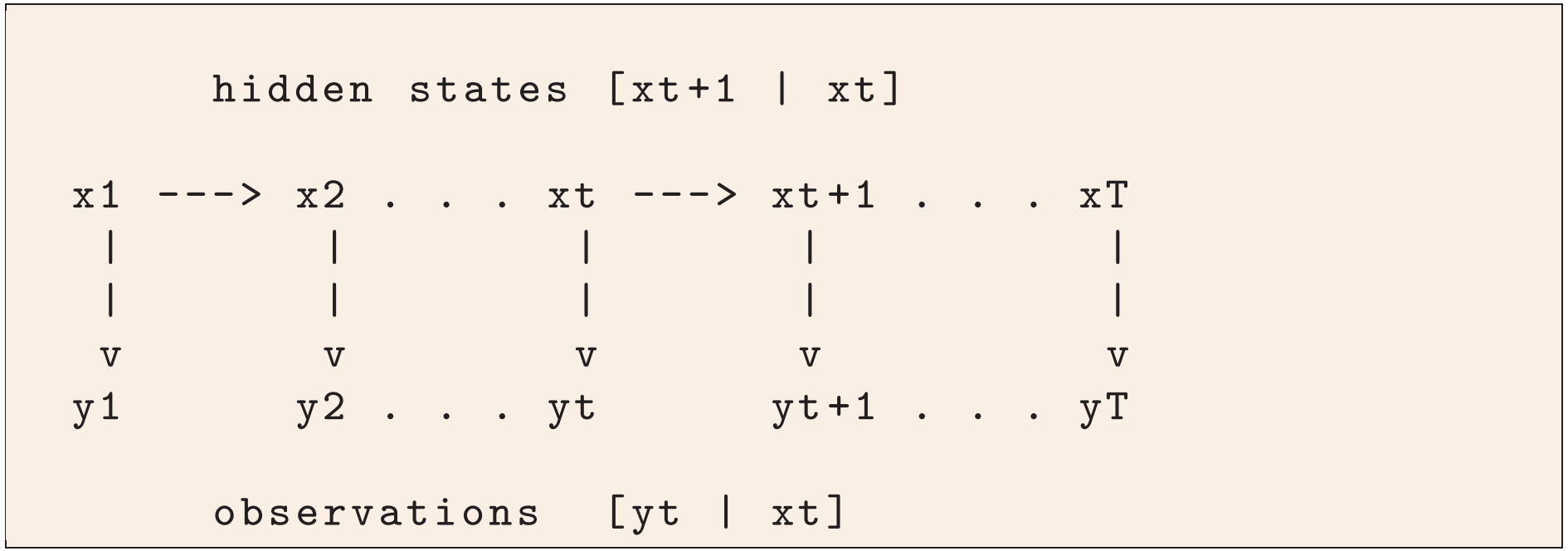
  - alignment problem—different samples could not be aligned so that parental markers had consistent phase and sub-genome assignment

Peeling—this year

- developed Hidden Markov model (HMM) peeling algorithms.
  - more accurate estimation of posterior distribution when alignment uncertain

- incorporating alignment into HMM peeling (in progress).

Hidden Markov models



$y_1, y_2, \ldots, y_T$: observation sequence (e.g. marker phenotypes)

$x_1, x_2, \ldots, x_T$: hidden states (e.g. fully informative virtual marker genotypes)

HMM key properties

- Observations depend only on current state :

$$[y_t \mid x_1, \ldots, x_t, \ldots] = [y_t \mid x_t] \tag{1}$$

- Conditional independence: observations independent given states.

- Markov property (probability of next state depends only on current state):

$$[x_{t+1} \mid x_1, \ldots, x_t] = [x_{t+1} \mid x_t] \tag{2}$$

Hidden Markov models

- Given data at time t the future is independent of the past.

- $\Rightarrow$ fast algorithms
  - forwards recursion
  - backwards recursion
  - backwards sampling

Forwards recursion

Let

$$\alpha_t(x_t) = [y_1, \ldots, y_t \mid x_t][x_t] \tag{3}$$

Then

$$\alpha_1(x_1) = \pi(x_1)[y_1 \mid x_1] \tag{4}$$

$$\alpha_{t+1}(x_{t+1}) = \sum_{x_t=1}^{N} [x_{t+1} \mid x_t][y_{t+1} \mid x_{t+1}]\alpha_t(x_t) \tag{5}$$

Forwards recursion

- Brute force evaluation of all states is $O(N^T)$ (exponential in number of alleles $\times$ number of markers).

- Forwards recursion faster $O(N^2T)$ (linear in T, exponential in max. number of alleles at a *single* marker).

- Calculate marginal distribution:

$$[y_1, \ldots, y_T] = \sum_{x_T=1}^{N} \alpha_T(x_T) \tag{6}$$

Backwards sampling

Sample x_T:

$$[x_T \mid y] \propto \alpha_T(x_T) \tag{7}$$

Recursively sample $[x_t \mid x_{t+1}]$:

$$[x_t \mid x_{t+1}, y] = \frac{\alpha_t(x_t)[x_{t+1} \mid x_t][y_{t+1} \mid x_{t+1}]}{\alpha_{t+1}(x_{t+1})} \tag{8}$$

HMM peeling for polyploids

- Conditional on parental genotypes, marker genotypes (x_t) and phenotypes (y_t) for an individual are HMM where t = position on chromosome.

- Transition probabilities $[x_{t+1} \mid x_t]$ depend on recombination rates.

HMM peeling for polyploids

Still computationally intensive:

- Need HMM for each progeny for each parental genotype combination.

- Can be thousands of parental genotype combinations at any marker; most very low probability.

HMM peeling for polyploids

- First select non-negligible genotypes at each marker by single marker peeling.
  - Detect and fix individual genotype errors at this stage.
- Combinations with joint probabilities also dropped at each step when stepping along the chromosome.
- Massive reduction by dropping low probability combinations.

HMM peeling algorithm for polyploids

- First estimate marginal probabilities for parental genotype combinations, discarding combinations with low probability.

- Then sample from the parental combinations, and for each sampled value, sample progeny virtual fully informative marker genotypes at each marker.

- Eliminate redudancy by using alignment permutations.

HMM peeling algorithm for polyploids

```
min_prob <- 0.005
For t = 1,...,T
  Do single marker peeling(t) -> P(t)
P(1,2) <- P(1) x P(2)
Do joint peeling(1,2) --> f(p)
P(1,2) <- {p in P(1,2): f(p) > min_prob}
For t = 3,...,T
  P(1,...,t) <- P(1,...t-1) x P(t)
  For each p in P(1,...,t)
    For each progeny i
      Do HMM peeling(1,...,t|p,i):
      alpha_t(x_t) --> [y1,...yt]  --> f(p,i)
  f(p) <- product(f(p,i))
  P(1,...,t) <- {p in P(1,...,t): f(p) > min_prob}
```

HMM peeling algorithm for polyploids

```
Nsamples <- 20
For s = 1,...,Nsamples
  Sample p(s) from P(1,...,T), f(p)
  For each progeny i
    Do HMM peeling(1,...,T|p(s),i):
    --> alpha_t(.)
    Sample x(s,i)[T] from alpha_T(.)
    For t=T-1,...,1
      Backwards sample x(s,i)[t]
        from x(s,i)[t] | x(s,i)[t+1]
Return x
```

Alignment

Alignment $\equiv$ assignment of chromosomes to grandparents and sub-genomes e.g.:

```
    [ 1 ------44----------------20----335-----17----------
A1  |
    [ 2 ----------303----24---22----336-----0-----19----

    [ 3 -----721-----------------20----340-----17---------
A2  |
    [ 4 ------------310----26---22----338-----0------0---
```

Nuclear family $\Rightarrow$ alignment for parents not determined by the data.

Alignment

- Diploids
  - parental phase usually uniquely determined with high probability
  - often worked out by hand
- Polyploids
  - more missing information
  - more permutations
  - relative parental phase or sub-genome assignment may be uncertain

Alignment permutations

- the initial phase or 'alignment' is indeterminate, no information in the data—e.g. which physical chromosome is subgenome 1, which is 2, and within each subgenome which comes from granparent 1, which from grandparent 2

- $\longrightarrow$ a group of permutations generated by switching sub-genomes within a chromosome or switching grandparents within a sub-genome.

- $\Rightarrow$ the posterior probability of a parental combination is unchanged if we apply the same permutation to all markers.

The alignment problem

- For each peeling sample we would like to find an alignment permutation that matches the parental data with the first sample.

- One fully informative marker would uniquely determine the permutation.

- A series of partially informative markers may combine information until the permutation is determined.

Alignment permutations for allo-tetraploid

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 4 & 3 \\ 2 & 1 & 3 & 4 \\ 2 & 1 & 4 & 3 \\ 3 & 4 & 1 & 2 \\ 3 & 4 & 2 & 1 \\ 4 & 3 & 1 & 2 \\ 4 & 3 & 2 & 1 \end{pmatrix} \times \begin{pmatrix} 5 & 6 & 7 & 8 \\ 5 & 6 & 8 & 7 \\ 6 & 5 & 7 & 8 \\ 6 & 5 & 8 & 7 \\ 7 & 8 & 5 & 6 \\ 7 & 8 & 6 & 5 \\ 8 & 7 & 5 & 6 \\ 8 & 7 & 6 & 5 \end{pmatrix} \tag{9}$$

$$8 \times 8 = 64 \tag{10}$$

Number of alignment permutations

- Ploidy $2k \Rightarrow k! \times 2^k$ alignment permutations per parent.

Table 1. Number of alignment permutations.

		no. perms	
ploidy	k	1 parent	2 parents
diploid	1	2	4
tetra	2	8	64
hexa	3	48	2304
octo	4	384	147456
deca	5	3840	1.5×10^7
dodeca	6	46080	2.1×10^9

Alignment example 1.

```
Sample 1:
(330,  0,331,  0) x (  0,  0,  0,  0)
( 40, 41,  0,  0) x ( 44,  0,  0,  0)
(  0,  0, 21, 22) x (  0,  0, 24,  0)
Sample 2:
(331,  0,330,  0) x (  0,  0,  0,  0)
(  0,  0, 40, 41) x (  0,  0,  0, 44)
( 21, 22,  0,  0) x ( 24,  0,  0,  0)

The permutation is:

(3 4 1 2) x (8 7 5 6)
```

Alignment example 2

```
Sample 1:
(330,331,  0,  0) x (  0,  0,  0,  0)
( 40, 41,  0,  0) x ( 44,  0,  0,  0)
(  0,  0, 21, 22) x (  0,  0, 24,  0)
Sample 2:
(331,  0,330,  0) x (  0,  0,  0,  0)
(  0,  0, 40, 41) x (  0,  0,  0, 44)
( 21, 22,  0,  0) x ( 24,  0,  0,  0)

Marker 1: 330,331 on same sub-genome in sample 1
but different sub-genomes in sample 2 (incnsistent).

Marker 1: perm. (3 1 2 4) or (3 1 4 2) (not valid).

Marker 2: perm. (3 4 1 2) or (3 4 2 1) (inconsistent).
```

Alignment example 3

```
Sample 1:
(330,331,  0,  0) x (  0,  0,  0,  0)
(  0,  0, 40, 41) x ( 44,  0,  0,  0)
( 21,  0,  0,  0) x ( 22,  0, 24,  0)
Sample 2:
(331,330,  0,  0) x (  0,  0,  0,  0)
(  0,  0, 40, 41) x (  0,  0,  0, 44)
( 21,  0,  0,  0) x ( 24,  0,  0, 22)

Marker 1: perm. (2 1 3 4) or (2 1 4 3).

Marker 2: perm.(1 2 3 4) or (2 1 3 4).

Markers 1,2: perm. (2 1 3 4) (unique).

Marker 3: perm. (1 2 3 4) or (2 1  2 3) (inconsistent).
```

Alignment problem.

- Want to reduce to a unique permutation but:
  - may not be enough information $\Rightarrow$ uncertainty
  - marker data may be inconsistent $\Rightarrow$ no solution possible
- Multiple samples from conditional peeling for tetraploid clover would not align.
- Without alignment QTL phase would effectively be random
- $\Rightarrow$ estimated QTL effects would cancel

The alignment problem—proposed solution

- possibly a problem with conditional peeling, use HMM peeling

- if they still don't align
  - compute equivalence classes of mutually alignable samples
  - analyse alignment classes separately
  - $\Rightarrow$ multiple runs of the BIC method
  - estimate probabilities for each alignment class $\propto$ number of samples in each class
  - combine information across alignment classes[31] according to their probabilities

Implementation progress

- HMM peeling implemented
- HMM peeling incremental approach implemented
- parallel computing (OpenMP) tried
- algorithms modified to store conditional probability calculations
- test runs on tetraploid clover chromosomes A1A2, B1B2

Results

- No speedup with OpenMP as yet (computations dominated by memory access or shared memory conflicts?)
  - OpenMP easy in theory, but . . .

- HMM peeling for clover B1,B2.
  - Time still increases rapidly with increasing numbers of markers (perhaps just fortuitous not many combinations for initial markers?)
  - Time in minutes for 3 markers, days for 6–7 markers.

Next steps

- Integrate alignment permutations integrated into HMM peeling (in progress),
  - expect up to 64$\times$ speedup for tetraploids
- Further testing, runs on other chromsomes.
- QTL mapping actual and simulated QTL data $\rightsquigarrow$ paper.

HMM peeling output for clover B1B2

Single marker peeling:

```
single marker peeling:
 *** single marker peeling for locus[0] prs312a ***
peeling (peel1):


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

alleles:
0: 0
1: 312
parental phenotypes:
1: 312
2: 0
number of progeny = 92
lc1p->n_unique_phenotypes = 2
```

```
table of progeny phenotypes
 i:   yp, count
 0:    0, 38
 1:  312, 36
lc1p->n_genotypes = 16
marginal probabilities for parental genotypes
 i:   mg,   yp,     p1,     p2
 0: 0,0,0,0,    0, 0.0000, 1.0000
 1: 0,0,0,312,  312, 0.2500, 0.0000
2: 0,0,312,0,  312, 0.2500, 0.0000
 3: 0,0,312,312,  312, 0.0000, 0.0000
 4: 0,312,0,0,  312, 0.2500, 0.0000
 5: 0,312,0,312,  312, 0.0000, 0.0000
. . .
 6: 0,312,312,0,  312, 0.0000, 0.0000
 8: 312,0,0,0,  312, 0.2500, 0.0000
 9: 312,0,0,312,  312, 0.0000, 0.0000
10: 312,0,312,0,  312, 0.0000, 0.0000
. . .
11: 312,0,312,312,  312, 0.0000, 0.0000
12: 312,312,0,0,  312, 0.0000, 0.0000
```

```
13: 312,312,0,312,  312, 0.0000, 0.0000
14: 312,312,312,0,  312, 0.0000, 0.0000
15: 312,312,312,312,  312, 0.0000, 0.0000
.
.
.
*** single marker peeling for locus[23] ats067a ***
peeling (peel1):


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

alleles:
0: 0
1: 67
parental phenotypes:
1: 0
2: 67
number of progeny = 92
```

```
lc1p->n_unique_phenotypes = 2
table of progeny phenotypes
 i:   yp, count
 0:    0, 47
 1:   67, 40
lc1p->n_genotypes = 16
marginal probabilities for parental genotypes
 i:   mg,   yp,     p1,     p2
 0: 0,0,0,0,    0, 1.0000, 0.0000
 1: 0,0,0,67,   67, 0.0000, 0.2500
 2: 0,0,67,0,   67, 0.0000, 0.2500
 3: 0,0,67,67,   67, 0.0000, 0.0000
 4: 0,67,0,0,   67, 0.0000, 0.2500
 5: 0,67,0,67,   67, 0.0000, 0.0000
. . .
 6: 0,67,67,0,   67, 0.0000, 0.0000
 8: 67,0,0,0,   67, 0.0000, 0.2500
 9: 67,0,0,67,   67, 0.0000, 0.0000
10: 67,0,67,0,   67, 0.0000, 0.0000
. . .
11: 67,0,67,67,   67, 0.0000, 0.0000
```

```
12: 67,67,0,0,    67, 0.0000, 0.0000
13: 67,67,0,67,    67, 0.0000, 0.0000
14: 67,67,67,0,    67, 0.0000, 0.0000
15: 67,67,67,67,    67, 0.0000, 0.0000
```

HMM peeling

```
*** joint peeling with hmm for loci 0...1 ***
 ***> joint_peel_relative_hmm:  parental combination [0] ***
locus 0, (0,0,0,312) x (0,0,0,0)
locus 1, (0,548,0,0) x (0,0,0,544)
locus 2, (0,0,0,77) x (0,0,0,0)
.
.
.

(*** HMM transition probability matrix [xt+1|xt] ***)
logfmp1mp2_locus_set_hmm: before hmm;
    fxx values (t = 1, theta = 0.0099):
             0        1        2        3        4        5        6
0  [ 0.96098 0.00961 0.00961 0.00010 0.00961 0.00010 0.00010 . . .
```

```
1   [ 0.00961 0.96098 0.00010 0.00961 0.00010 0.00961 0.00000
2   [ 0.00961 0.00010 0.96098 0.00961 0.00010 0.00000 0.00961
3   [ 0.00010 0.00961 0.00961 0.96098 0.00000 0.00010 0.00010
4   [ 0.00961 0.00010 0.00010 0.00000 0.96098 0.00961 0.00961
5   [ 0.00010 0.00961 0.00000 0.00010 0.00961 0.96098 0.00010
6   [ 0.00010 0.00000 0.00961 0.00010 0.00961 0.00010 0.96098
.
.
.


(*** Vector of observations probabilities for locus 0 ***)
[0] fyx[iy=1,ix=0), =  0.00000
[0] fyx[iy=1,ix=1), =  0.00000
[0] fyx[iy=1,ix=2), =  1.00000
[0] fyx[iy=1,ix=3), =  1.00000
[0] fyx[iy=1,ix=4), =  0.00000
[0] fyx[iy=1,ix=5), =  0.00000
[0] fyx[iy=1,ix=6), =  1.00000
[0] fyx[iy=1,ix=7), =  1.00000
[0] fyx[iy=1,ix=8), =  0.00000
```

```
[0] fyx[iy=1,ix=9), =  0.00000
[0] fyx[iy=1,ix=10), =  1.00000
[0] fyx[iy=1,ix=11), =  1.00000
[0] fyx[iy=1,ix=12), =  0.00000
[0] fyx[iy=1,ix=13), =  0.00000
[0] fyx[iy=1,ix=14), =  1.00000
[0] fyx[iy=1,ix=15), =  1.00000
.
.
.
--***   32 of 64 values selected with prob >=  0.000500000,
    accounting for 100.00000 percent of probability
Date: Fri Sep  9 18:14:22 2011

*** relative joint peeling hmm for loci 2|0...1 ***
 ***> joint_peel_relative_hmm:  parental combination [0] ***
locus 0, (0,0,0,312) x (0,0,0,0)
locus 1, (0,548,0,0) x (0,0,0,544)
locus 2, (0,0,0,77) x (0,0,0,0)

--***   32 of 128 values selected with prob >=  0.000500000,
```

```
    accounting for 100.00000 percent of probability
Date: Fri Sep  9 18:14:24 2011

*** relative joint peeling hmm for loci 3|0...2 ***
 ***> joint_peel_relative_hmm:  parental combination [0] ***
locus 0, (0,0,0,312) x (0,0,0,0)
locus 1, (0,548,0,0) x (0,0,0,544)
locus 2, (0,77,0,0) x (0,0,0,0)
locus 3, (0,0,0,5) x (0,0,0,7)
.
.
.

--****   64 of 256 values selected with prob >=  0.000500000,
    accounting for 100.00000 percent of probability
Date: Fri Sep  9 22:33:28 2011

*** relative joint peeling hmm for loci 6|0...5 ***
 ***> joint_peel_relative_hmm:  parental combination [0] ***
locus 0, (0,0,0,312) x (0,0,0,0)
locus 1, (0,548,0,0) x (0,0,0,544)
```

```
locus 2, (0,77,0,0) x (0,0,0,0)
locus 3, (0,5,0,0) x (0,7,0,0)
locus 4, (0,0,584,0) x (0,0,0,0)
locus 5, (0,123,0,0) x (0,0,0,0)
locus 6, (0,0,0,0) x (0,0,0,55)
.
.
.
--*** joint_peel_relative_hmm, max_locus = 6, min_prob= 0.00050:
    selecting and storing index vectors ***--
--***   64 of 256 values selected with prob >=  0.000500000,
    accounting for 100.00000 percent of probability
Date: Mon Sep 12 13:09:17 2011

*** relative joint peeling hmm for loci 7|0...6 ***
 ***> joint_peel_relative_hmm:  parental combination [0] ***
locus 0, (0,0,0,312) x (0,0,0,0)
locus 1, (0,548,0,0) x (0,0,0,544)
locus 2, (0,77,0,0) x (0,0,0,0)
locus 3, (0,5,0,0) x (0,7,0,0)
locus 4, (0,0,584,0) x (0,0,0,0)
```

```
locus 5, (0,123,0,0) x (0,0,0,0)
locus 6, (0,0,0,0) x (0,55,0,0)
locus 7, (0,0,0,648) x (0,0,0,0)
logfmp1mp2_locus_set_hmm: on exit: logfmp1mp2 = -422.11559,
    fmp1mp2 = 4.7591e-184
 ***> joint_peel_relative_hmm:  parental combination [1] ***
locus 0, (0,0,0,312) x (0,0,0,0)
locus 1, (0,548,0,0) x (0,0,0,544)
locus 2, (0,77,0,0) x (0,0,0,0)
locus 3, (0,5,0,0) x (0,7,0,0)
locus 4, (0,0,584,0) x (0,0,0,0)
locus 5, (0,123,0,0) x (0,0,0,0)
locus 6, (0,0,0,0) x (0,55,0,0)
locus 7, (0,0,648,0) x (0,0,0,0)
.
.
.
```

Sample output, 10 samples (0-9) for each locus:

```
*
 *** Joint peeling/hmm: sample data for locus  0 ***
*
0: 0,0,0,312; 0,0,0,0; 1,5,4,8;1,5,4,8;2,5,4,7;1,6,3,7;1,5,4,8;1,5,3,7;2,6,4,7; . . .
1: 0,0,0,312; 0,0,0,0; 1,6,4,7;1,6,4,7;2,5,4,7;1,5,3,8;1,6,4,7;1,5,3,7;2,5,4,8; . . .
2: 0,0,312,0; 0,0,0,0; 1,5,3,8;1,5,3,8;2,6,3,8;1,6,4,7;1,5,3,8;1,6,4,8;2,5,3,7; . . .
3: 312,0,0,0; 0,0,0,0; 1,5,4,8;1,5,4,7;1,5,3,8;2,6,4,8;1,5,4,7;2,5,4,8;1,6,3,7; . . .
4: 0,0,0,312; 0,0,0,0; 2,5,4,7;2,6,4,7;1,6,4,7;2,6,3,8;2,5,4,7;2,5,3,7;1,5,4,8; . . .
5: 0,312,0,0; 0,0,0,0; 2,6,4,7;2,5,4,7;2,5,3,7;1,6,4,8;2,5,4,7;1,5,4,7;2,5,3,8; . . .
6: 312,0,0,0; 0,0,0,0; 1,6,3,8;1,6,3,8;1,6,4,7;2,5,3,7;1,6,3,8;2,6,3,8;1,5,4,8; . . .
7: 0,312,0,0; 0,0,0,0; 2,6,3,8;2,5,3,8;2,6,4,8;1,6,3,7;2,5,3,8;1,5,3,8;2,6,4,7; . . .
8: 312,0,0,0; 0,0,0,0; 1,5,3,7;1,5,3,7;1,5,4,8;2,6,3,8;1,5,3,7;2,5,3,7;1,6,4,8; . . .
9: 0,0,312,0; 0,0,0,0; 1,6,3,7;1,5,3,7;2,6,3,7;1,6,4,8;1,5,3,7;1,6,4,7;2,5,3,8; . . .
*
 *** Joint peeling/hmm: sample data for locus  1 ***
*
0: 548,0,0,0; 544,0,0,0; 1,5,4,8;1,5,4,8;2,5,4,7;1,6,3,7;1,5,4,8;1,5,3,7;2,6,4,7; . .
1: 548,0,0,0; 0,0,544,0; 1,6,4,7;1,6,4,7;2,5,4,7;1,5,3,8;1,6,4,7;1,5,3,7;2,5,4,8; . .
2: 548,0,0,0; 0,0,0,544; 1,5,3,8;1,5,3,8;2,6,3,8;1,6,4,7;1,5,3,8;1,6,4,8;2,5,3,7; . .
3: 0,0,0,548; 544,0,0,0; 1,5,4,8;1,5,4,7;1,5,3,8;2,6,4,8;1,5,4,7;2,5,4,8;1,6,3,7; . .
4: 0,548,0,0; 0,0,544,0; 2,5,4,7;2,5,4,7;1,6,4,7;2,6,3,8;2,5,4,7;2,5,3,7;1,5,4,8; . .
5: 0,0,0,548; 0,0,544,0; 2,6,4,7;2,5,4,7;2,6,3,7;1,6,4,8;2,5,4,7;1,5,4,7;2,5,3,8; . .
6: 0,0,548,0; 0,544,0,0; 1,6,3,8;1,6,3,8;1,6,4,7;2,5,3,7;1,6,3,8;2,6,3,8;1,5,4,8; . .
7: 0,0,548,0; 0,0,0,544; 2,6,3,8;2,5,3,8;2,6,4,8;1,6,3,7;2,5,3,8;1,5,3,8;2,6,4,7; . .
8: 0,0,548,0; 544,0,0,0; 1,5,3,7;1,5,3,7;1,5,4,8;2,6,3,8;1,5,3,7;2,5,3,7;1,6,4,8; . .
9: 548,0,0,0; 0,0,544,0; 1,6,3,7;1,5,3,7;2,6,3,7;1,6,4,8;1,5,3,7;1,6,4,7;2,5,3,8; . .
*
```

```
 *** Joint peeling/hmm: sample data for locus  2 ***
*
0: 77,0,0,0; 0,0,0,0; 1,5,4,8;1,5,4,8;2,5,4,7;1,6,3,7;1,5,4,8;1,5,3,7;2,6,4,7; . . .
1: 77,0,0,0; 0,0,0,0; 1,6,4,7;1,6,4,7;2,5,4,7;1,5,3,8;1,6,4,7;1,5,3,7;2,5,4,8; . . .
2: 77,0,0,0; 0,0,0,0; 1,5,3,8;1,5,3,8;2,6,3,8;1,6,4,7;1,5,3,7;1,6,4,8;2,5,3,7; . . .
3: 0,0,0,77; 0,0,0,0; 1,5,4,8;1,5,4,7;1,5,3,8;2,6,4,8;1,5,4,7;2,6,4,8;1,6,3,7; . . .
4: 0,77,0,0; 0,0,0,0; 2,5,4,7;2,5,4,7;1,6,4,7;2,6,3,8;2,5,4,7;2,5,3,7;1,5,4,8; . . .
5: 0,0,0,77; 0,0,0,0; 2,6,4,7;2,5,4,7;2,6,3,7;1,6,4,8;2,5,4,7;1,5,4,7;2,5,3,8; . . .
6: 0,0,77,0; 0,0,0,0; 1,6,3,8;1,6,3,8;1,6,4,7;2,5,3,7;1,6,3,8;2,6,3,8;1,5,4,8; . . .
7: 0,0,77,0; 0,0,0,0; 2,6,3,8;2,5,3,8;2,6,4,8;1,6,3,7;2,5,3,8;1,5,3,8;2,6,4,7; . . .
8: 0,0,77,0; 0,0,0,0; 1,5,3,7;1,5,3,7;1,6,4,8;2,6,3,8;1,5,3,7;2,5,3,7;1,6,4,8; . . .
9: 77,0,0,0; 0,0,0,0; 1,6,3,7;1,5,3,7;2,6,3,7;1,6,4,8;1,5,3,7;1,6,4,7;2,5,3,8; . . .
*
 *** Joint peeling/hmm: sample data for locus  3 ***
*
0: 5,0,0,0; 0,0,0,7; 1,5,4,8;1,5,4,8;2,5,4,7;1,6,3,7;1,5,4,8;1,5,3,7;2,6,4,7; . . .
1: 5,0,0,0; 0,7,0,0; 1,6,4,7;1,6,4,7;2,5,4,7;1,5,3,8;1,6,4,7;1,5,3,7;2,5,4,8; . . .
2: 5,0,0,0; 7,0,0,0; 1,5,3,8;1,6,3,8;2,6,3,8;1,6,4,7;1,5,3,7;1,6,4,8;2,5,3,7; . . .
3: 0,0,0,5; 0,0,7,0; 1,5,4,8;1,5,4,7;1,5,3,8;2,6,4,8;1,5,4,7;2,6,4,8;1,6,3,7; . . .
4: 0,5,0,0; 7,0,0,0; 2,5,4,7;2,5,4,7;1,6,4,7;2,6,3,8;2,5,4,7;2,5,3,7;1,5,4,8; . . .
5: 0,0,0,5; 7,0,0,0; 2,6,4,7;2,5,4,7;2,6,3,7;1,6,4,8;2,5,4,7;1,5,4,7;2,5,3,8; . . .
6: 0,0,5,0; 0,0,0,7; 1,6,3,8;1,6,3,8;1,6,4,7;2,5,3,7;1,6,3,8;2,6,3,8;1,5,4,8; . . .
7: 0,0,5,0; 7,0,0,0; 2,6,3,8;2,5,3,8;2,6,4,8;1,6,3,7;2,5,3,8;1,5,3,8;2,6,4,7; . . .
8: 0,0,5,0; 0,0,7,0; 1,5,3,7;1,5,3,7;1,6,4,8;2,6,3,8;1,5,3,7;2,5,3,7;1,6,4,8; . . .
9: 5,0,0,0; 7,0,0,0; 1,6,3,7;1,5,3,7;2,6,3,7;1,6,4,8;1,5,3,7;1,6,4,7;2,5,3,8; . . .
*
 *** Joint peeling/hmm: sample data for locus  4 ***
*
```

```
0: 0,0,584,0; 0,0,0,0; 1,5,4,8;1,5,4,8;2,5,4,7;1,6,3,7;1,5,4,8;1,5,3,7;2,6,4,7; . . .
1: 0,0,584,0; 0,0,0,0; 1,6,4,7;1,6,4,7;2,5,4,7;1,5,3,8;1,6,4,7;1,5,3,7;2,5,4,8; . . .
2: 0,0,0,584; 0,0,0,0; 1,5,3,8;1,6,3,8;2,6,3,8;1,6,4,7;1,5,3,7;1,6,4,8;2,5,3,7; . . .
3: 0,584,0,0; 0,0,0,0; 1,5,4,8;1,5,4,7;1,6,3,8;2,6,4,8;1,5,4,7;2,6,4,8;1,6,3,7; . . .
4: 0,0,584,0; 0,0,0,0; 2,5,4,7;2,5,4,7;1,6,4,7;2,6,3,8;2,5,4,7;2,5,3,7;1,5,4,8; . . .
5: 584,0,0,0; 0,0,0,0; 2,6,4,7;2,5,4,7;2,6,3,7;1,6,4,8;2,5,4,7;1,5,4,7;2,5,3,8; . . .
6: 0,584,0,0; 0,0,0,0; 1,6,3,8;1,6,3,8;1,6,4,7;2,5,3,7;1,6,3,8;2,6,3,8;1,5,4,8; . . .
7: 584,0,0,0; 0,0,0,0; 2,6,3,8;2,5,3,8;2,6,4,8;1,6,3,7;2,5,3,8;1,5,3,8;2,6,4,7; . . .
8: 0,584,0,0; 0,0,0,0; 1,5,3,7;1,5,3,7;1,6,4,8;2,6,3,8;1,5,3,7;2,5,3,7;1,6,4,8; . . .
9: 0,0,0,584; 0,0,0,0; 1,6,3,7;1,5,3,7;2,6,3,7;1,6,4,8;1,5,3,7;1,6,4,7;2,5,3,8; . . .
```

Sample alignments and recombination rates for clover B1B2

```
Sample 1: (1 2 3 4) x (4 3 1 2)
Sample 2: (1 2 4 3) x (4 3 2 1)
Sample 3: (4 3 2 1) x (1 2 4 3)
Sample 4: (2 1 3 4) x (3 4 2 1)
Sample 5: (4 3 2 1) x (3 4 2 1)
Sample 6: (3 4 2 1) x (2 1 3 4)
Sample 7: (3 4 1 2) x (4 3 2 1)
Sample 8: (3 4 2 1) x (1 2 4 3)
Sample 9: (1 2 4 3) x (3 4 1 2)

> # est. recombination rate (%)
> 100*cloverB1B2.recombinations/3680
     [,1] [,2] [,3] [,4] [,5]
[1,]  0.0 1.85 1.71 2.31 2.74
[2,]  1.8 0.00 0.41 1.06 1.49
[3,]  1.7 0.41 0.00 0.65 1.09
[4,]  2.3 1.06 0.65 0.00 0.43
[5,]  2.7 1.49 1.09 0.43 0.00
```

VISG Virtual Institute of Statistical Genetics

Acknowledgements

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).