Theor Appl Genet (2004) 109: 996–1004
DOI 10.1007/s00122-004-1705-x

ORIGINAL PAPER

Xin-Sheng Hu · Carol Goodwillie · Kermit M. Ritland

Joining genetic linkage maps using a joint likelihood function

Received: 5 March 2003 / Accepted: 14 April 2004 / Published online: 25 June 2004

Abstract We present an efficient method to join genetic maps derived from different crosses, which is especially appropriate for dominant markers. In contrast to the "JoinMap" algorithm, which estimates information about recombination in a given cross from the LOD values and then combines estimates among crosses assuming a binomial sampling distribution, we construct a joint likelihood function that combines information across all crosses, to obtain a joint estimate of recombination. Simulations indicate that the difference between these two approaches is small when codominant markers are used, but that the joint likelihood approach shows substantially improved estimates when dominant or a mixture of dominant and codominant markers are used. This is because the joint likelihood implicitly finds the optimal weights among different classes of data, while the former method does not accurately predict the information from crosses involving dominant markers. Application of our method is illustrated by the construction of a linkage map for *Linanthus* using both backcrosses and the F_2 of a cross between *Linanthus jepsonii* and *L. bicolor*, assayed with amplified fragment length polymorphisms (AFLP).

Communicated by O. Savolainen

X.-S. Hu (✉) · K. M. Ritland
Department of Forest Sciences, Faculty of Forestry, University of British Columbia,
Vancouver, BC, Canada, V6T 1Z4
e-mail: xin-sheng.hu@ualberta.ca
Tel.: +1-780-4920715
Fax: +1-780-4924323

C. Goodwillie
Department of Biology, East Carolina University,
Greenville, NC 27858, USA

Present address:
X.-S. Hu
Department of Renewable Resources, University of Alberta,
751 General Service Building,
Edmonton, AB, Canada, T6G 2H1

Introduction

Genetic linkage maps have a wide variety of applications in quantitative genetics and genomics. Numerous genetic maps, based upon various types of genetic markers, have accumulated in recent years. Classically, just a single cross is used to construct linkage maps, resulting in a segregating population (e.g. Maliepaard et al. 1997; Wu et al. 2002), or other types of populations such as recombinant inbred lines (RIL) or advanced intercross (Fisher and Balmakund 1928; Bailey 1961; Mather 1951). Well-established computer programs, such as MAPMAKER and LINKAGE-1 (Lander and Green 1987; Suitter et al. 1983), have been seen an extensive usage for the analysis of this type of data.

With the increasing activity focused on molecular mapping in many species, often more than one genetic map arises. There are at least two sources of multiple maps. First, separate mapping populations arise from separate experimental crosses using independent materials (e.g. Sewell et al. 1999; Lombard and Delourme 2001). Second, when alleles segregate in both parents of outcrossing species, the linkage map specific for each parent is often constructed, such as in the widely used pseudo-testcross mapping strategy (e.g. Grattapaglia and Sederoff 1994; Marques et al. 1998; Testolin et al. 2001). The problem then arises for how to combine multiple genetic maps into a single, optimally informative map.

Stam (1993) proposed an approach, implemented in the computer program "JoinMap", to integrate individual linkage mapping results from independent experiments into a single map. This method has been extensively applied (e.g. Burbridge et al. 2001; Dettori et al. 2001; Gosselin et al. 2002; Ulloa et al. 2002). The basic procedure employed by JoinMap is to start with individual pairwise recombination estimates derived from different experiments. These estimates are linearly combined into a single estimate using weights.

Here, we propose a procedure for integrating maps based upon the joint likelihood of data across experiments. Our point is that if estimates of recombination for a pair of

markers differ among experiments, the joint maximum likelihood estimate involving all experiments is, in most cases, more accurate than that given by the procedure of Stam (1993). The later stages in constructing linkage maps are the same as the preexisting methodologies (see Liu 1998). We use a simulation approach to examine the differences between Stam's and the joint likelihood methods in statistical power. We illustrate the application to multiple crosses with a study involving the wildflower, *Linanthus*. Our results also emphasize the importance of archiving raw data (individual progeny genotypes) from the original experiment for efficient use of new data.

Materials and methods

New methodology

Stam (1993) calculated the joint recombination fraction, r, across T experiments (crosses) as a weighted sum of individual estimates,

Table 1 The expected segregation ratios of phenotypes in the full-sib family of each type of parent mating assayed with one codominant marker and another dominant marker with a recombination fraction r between them. The *six phenotypes* are scored as follows: [a]FF/B_, [b]FF/NN, [c]SF/B_, [d]SF/NN, [e]SS/B_, [f]SS/NN, where F represents a fast band and S represents a slow band for the first codominant marker, and B_ stands for band presence and N for band absence for the second dominant marker

Genotypes	A_1B_1/A_1B_1	A_1B_1/A_1B_2	A_1B_1/A_2B_1	A_1B_1/A_2B_2	A_1B_2/A_2B_1	A_1B_2/A_1B_2	A_2B_1/A_2B_1	A_1B_2/A_2B_2	A_2B_1/A_2B_2	A_2B_2/A_2B_2
A_1B_1/A_1B_1	[a]1									
A_1B_1/A_1B_2	[a]1	[a]3/4 [b] 1/4								
A_1B_1/A_2B_1	[a]1/2 [c]1/2	[a]1/2 [c]1/2	[a]1/4 [c]1/2 [e]1/4							
A_1B_1/A_2B_2	[a]1/2 [c]1/2	[a](2−r)/4 [b]r/4 [c](1+r)/4 [d](1−r)/4	[a]1/4 [c]1/2 [e]1/4	[a](1−r^2)/4 [b]r^2/4 [c](1−r+r^2)/2 [d]r(1−r)/2 [e](2r−r^2)/4 [f](1−r^2)/4						
A_1B_2/A_2B_1	[a]1/2 [c]1/2	[a](1+r)/4 [b](1−r)/4 [c](2−r)/4 [d]r/4	[a]1/4 [c]1/2 [e]1/4	[a](1−r+r^2)/4 [b]r(1−r)/4 [c](1+2r−2r^2)/4 [d](1−2r+2r^2)/4 [e](1−r+r^2)/4 [f]r(1−r)/4	[a](2r−r^2)/4 [b](1−r)^2/4 [c](1−r+r^2)/2 [d]r(1−r)/2 [e](1−r^2)/4 [f]r^2/4					
A_1B_2/A_1B_2	[a]1	[a]1/2 [b]1/2	[a]1/2 [c]1/2	[a](1−r)/2 [b]r/2 [c]r/2 [d](1−r)/2	[a]r/2 [b](1−r)/2 [c](1−r)/2 [d]r/2	[b]1				
A_2B_1/A_2B_1	[c]1	[c]1	[c]1/2 [e]1/2	[c]1/2 [e]1/2	[c]1/2 [e]1/2	[c]1	[e]1			
A_1B_2/A_2B_2	[a]1/2 [c]1/2	[a]1/4 [b]1/4 [c]1/4 [d]1/4	[a]1/4 [c]1/2 [e]1/4	[a](1−r)/4 [b]r/4 [c]1/4 [d]1/4 [e]r/4 [f](1−r)/4	[a]r/4 [b](1−r)/4 [c]1/4 [d]1/4 [e](1−r)/4 [f]r/4	[b]1/2 [d]1/2	[c]1/2 [e]1/2	[b]1/4 [d]1/2 [f]1/4		
A_2B_1/A_2B_2	[c]1	[c]3/4 [d]1/4	[c]1/2 [e]1/2	[c](2−r)/4 [d]r/4 [e](1+r)/4 [f](1−r)/4	[c](1+r)/4 [d](1−r)/4 [e](2−r)/4 [f]r/4	[c]1/2 [d]1/2	[e]1	[c]1/4 [d]1/4 [e]1/4 [f]1/4	[e]3/4 [f]1/4	
A_2B_2/A_2B_2	[c]1	[c]1/2 [d]1/2	[c]1/2 [e]1/2	[c](1−r)/2 [d]r/2 [e]r/2 [f](1−r)/2	[c]r/2 [d](1−r)/2 [e](1−r)/2 [f]r/2	[d]1	[e]1	[d]1/2 [f]1/2	[e]1/2 [f]1/2	[f]1

$$\hat{r} = \sum_{i=1}^{T} f_i \hat{r}_i, \tag{1}$$

where f_i is the weight applied to the ith estimate of recombination, r_i. The weights for each r_i rely upon either an associated LOD value (derived from "population data") or a standard error (derived from "independent data"), and are obtained by finding the hypothetical binomial sample size which yields the same LOD or standard error. Thus for example, for the LOD case with sample size N, the expression $\mathrm{LOD} = \ln\left(r^{rN}(1-r)^{(1-r)N} \Big/ 0.5^{N}\right)$ is solved for N as $N = \mathrm{LOD}\big/(r\ln(r) - r\ln(1-r) + \ln(2-2r))$. This N is used as the relative weight for this r in the summation of Eq. 1. Note that the binomial probability is used regardless of the type of cross (backcross, F_2) and the type of marker (dominant, co-dominant).

The joint likelihood method proposed here is as follows. Consider two linked codominant markers A and B, and let the two parents of a cross be represented by the configuration A_iB_k/A_jB_l and $A_{i'}B_{k'}/A_{j'}B_{l'}$ ($i,i',j,j'=1,2,...,N_A$; $k,k',l,l'=1,2,...,N_B$), where N_A and N_B are the number of alleles for loci A and B, respectively. Similarly, denote the observed offspring configuration of the cross by A_yB_n/A_zB_m ($y,z=i,i'$, j,j'; $n,m=k,k',l,l'$). For the i-th cross of the total T crosses, the likelihood of the data is

$$\pi_i = \binom{N_i}{x_{i1}\ldots x_{iK}} \sum_{f=1}^{N_{AB}} \sum_{m=1}^{N_{AB}} P(f,m) \prod_{k=1}^{K} [P(k|f,m)]^{x_{ik}}, \tag{2}$$

where $P(f,m)$ is the *prior* probability of parental mating $f\times m$, $P(k|f,m)$ is the expected probability of phenotype k, given the parental mating type of $f\times m$, in which f and m refer to female and male genotypes, respectively, K is the total number of phenotypes that can be scored, x_{ik} is the number of the kth phenotype in the ith cross, and N_{AB} ($=N_AN_B$ $(N_AN_B+1)/2$) is the total number of combined parent genotype for these two loci. An example of $P(k|f,m)$ is given in Table 1 for a full-sib family, assayed for one codominant and one dominant marker with a recombination fraction r between them. A generalized probability for each progeny genotype, conditioned on parental genotype, can be expressed with the "Kronecker operator" (see Ritland 2002; Hu and Ritland, in preparation).

In Eq. 2, the summation is overall all parent linkage phases. Alternatively, only the most likely phase can be used. This is likely to be more efficient with larger progeny sizes, but at these sizes, the incorrect phases would have likelihoods much smaller than the correct phase, and have little effect on the overall likelihood. With smaller progeny sizes, choosing the most likely phase may introduce biases due to small sample size effects. Thus we prefer using the unmodified summation over alternative linkage phases. These statistical properties are very much like those encountered in the inference of maternal parentage in mating system estimation using open-pollinated progeny arrays (Ritland 1986).

The prior $P(f,m)$ can be the population genotypic frequencies, and include even the phase (linkage disequilibrium). However, usually the genotype frequencies have not been assayed, and natural levels of linkage disequilibrium are likely to be nearly zero between even the closest detectably linked markers. Thus, we use the simplest prior—a uniform distribution. Furthermore, with any reasonable progeny size (approximately 20 or greater), the prior carries little weight compared to the progeny segregation pattern.

In Eq. 2, the number of all possible offspring phenotypes (K) depends upon the number of alleles at the loci and whether the loci exhibit dominance or co-dominance (Table 2). With the assumption of *prior* uniform distribution for the parental genotype distribution in different crosses, the expected frequency of informative crosses also depends upon the type of marker (Table 2). When both parents and offspring are genotyped, priors are not necessary, and Eq. 2 can be reduced to

$$\pi_i = \binom{N_i}{x_{i1}\ldots x_{iK}} \prod_{k=1}^{K} [P(k|f,m)]^{x_{ik}}. \tag{3}$$

In each of the above two cases, the joint likelihood of T crosses is the product of π_i ($i=1,...,T$),

$$L = \prod_{i=1}^{T} \pi_i, \tag{4}$$

which is maximized with respect to a single r across all T crosses. Note that when parent genotypes are available and $T=1$, Eq. 4 reduces to the traditional case involving just a single cross (e.g. Maliepaard et al. 1997).

Maximum likelihood estimates (MLEs) of the recombination fraction can be obtained via the Newton-Raphson (NR) method. Given the t-th estimate of recombination fraction (r^t), the recombination fraction at the next iterative step (r^{t+1}) can be calculated by

$$\hat{r}^{t+1} = \hat{r}^{t} - \left(\frac{\partial^2 \ln L}{\partial r^2}\Big|_{r=\hat{r}^t}\right)^{-1} \frac{\partial \ln L}{\partial r}\Big|_{r=\hat{r}^t}, \tag{5}$$

where $\partial\ln L/\partial r$ and $\partial^2\ln L/\partial r^2$ can be obtained according to Hu and Ritland (in preparation) with minor modifications. The above is iterated until convergence ($(|\hat{r}^{t+1} - \hat{r}^{t}|$ is small). Our numerical simulations showed any initial setting of r within the range of (0, 0.5) results in the same convergence value.

Table 2 The number of phenotypes and the number of informative crosses in the two-locus case, denoted by *A* and *B* with N_A and N_B alleles, respectively, under different combinations of inheritance models

Inheritance models	Markers	Number of phenotypes	Number of informative crosses
Both codominant	e.g. Isozymes, RFLPs, microsatellites	$N_A(N_A+1)/2 \times N_B(N_B+1)/2$	$N_AN_B(N_A-1)(N_B-1)/2 \times N_AN_B(N_AN_B+1)/2$
Co-dominant–dominant	e.g. Microsatellite-RAPD	$N_A(N_A+1)/2 \times 2$	$N_A(N_A-1) \times N_A(3N_A+1)/2$
Both dominant	e.g. AFLP, RAPD	4	10

Comparison of alternative methods

To examine the statistical properties of Stam's versus our method, we conducted Monte Carlo simulations involving several independent crosses, with linkage phase known. Nine types of crosses were generated by crossing either A_1B_1/A_2B_2 and A_1B_2/A_2B_1 (informative coupling or repulsion double heterozygote parents) with either of A_1B_1/A_2B_2, A_1B_2/A_2B_1, A_1B_2/A_2B_2, A_2B_1/A_2B_2 and A_2B_2/A_2B_2 (various informative and non-informative parents). These crosses represent a variety of mating designs, such as F_2 and BC_1 progeny, expected to occur in outbred populations. An additional two crosses ($A_1B_1/A_2B_2 \times A_1B_1/A_1B_1$ and $A_1B_2/A_2B_1 \times A_1B_1/A_1B_1$), the backcross to another parent BC_2, were also considered for the case of codominant markers. With diallelic loci, i.e. N_A=N_B=2 and N_{AB}=9, the number of progeny phenotypes is K=9 for the codominant markers, K=4 for the dominant markers, and K=6 for the mixture.

The results of these simulations indicate that these two methods differ in their precision of estimation, although the average estimates with both methods are close to the true values. When codominant markers are used, the joint likelihood method has slightly better precision than Stam's method (Fig. 1a). The difference between the two methods becomes considerably greater for the case of one codominant marker linked with one dominant marker (Fig. 1b), and even greater for the case of two linked dominant markers (Fig. 1c).

In addition to the differences in statistical power, the sample size needed for a given precision of estimates also varies between these two methods. For example, with codominant markers, for both methods, 30 individuals per cross for approximately ten crosses can generate good estimates (unbiased average together with small standard deviations). However with dominant markers or mixture of marker types there are substantial differences of adequate sample sizes between the two methods (Fig. 2a, b). Fifty individuals per cross and 10 crosses are appropriate when the joint likelihood method is applied, with either a mixture of markers (Fig. 3a, b) or pure dominant markers (Fig. 4a, b); while double the number (more than one hundred individuals) are required for Stam's method in either case.

Computer progams

We have written a series of computer programs in C that implement our method for the estimation of pairwise recombination rates and their associated *P*-values (as determined by bootstrapping), as well as for linkage map construction. The programs can handle diallelic markers for either dominant or codominant markers, or a mixture of the two. Programs are available from X.-S. Hu. The criteria for grouping the markers are the same as those in previous studies (e.g. Stam 1993). The first stage is the generation of a preliminary linkage marker order for each group. If the *P*-value of a pair of markers is less than a

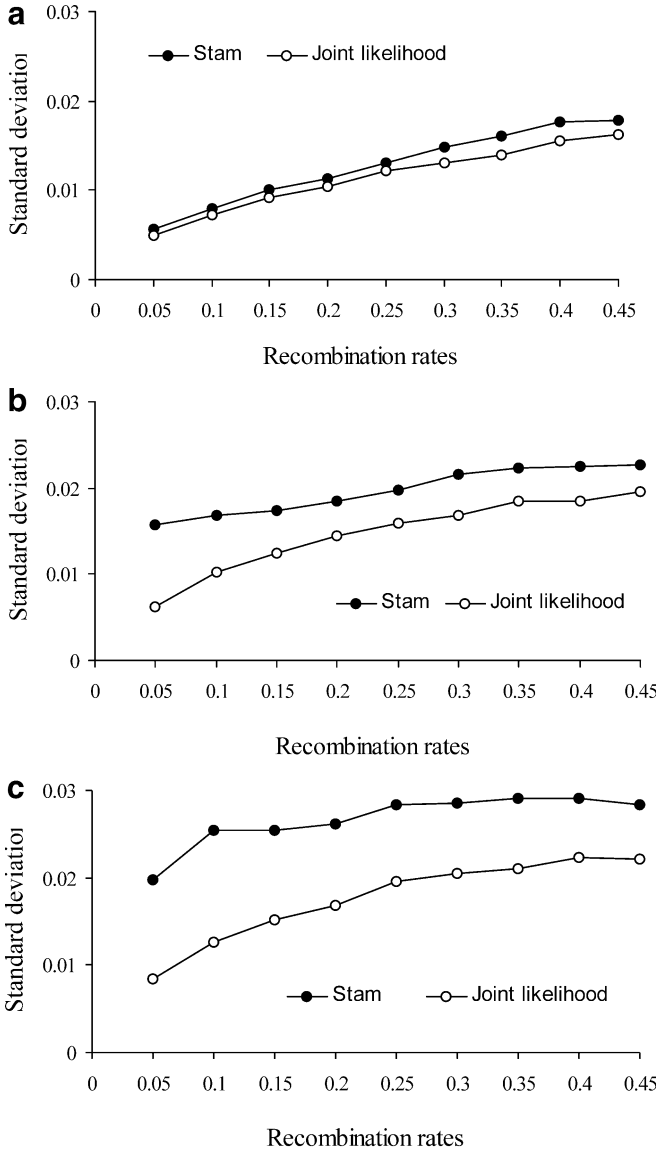

Fig. 1a–c Comparison between Stam's and the joint likelihood methods in terms of the standard deviation: **a** two codominant markers, with 11 crosses; **b** one codominant marker paired with another dominant marker, with nine crosses; and **c** two dominant markers, with nine crosses. Estimates are based upon 1,000 independent runs; sample sizes are 200 for F_2 populations and 100 for other crosses

given threshold, the markers are assigned to the same linkage group. The preliminary order of markers in a linkage group is then found by first choosing the pair of markers with the smallest *r*, then progressively adding markers with the smallest *r* to the ends of the growing linkage group until all linked markers are incorporated (see Liu 1998). The order of adjacent markers are then permuted, and the order with the minimum sum of adjacent recombination fractions (SARF) is chosen as the final order. Recombination fractions are then transformed into mapping distances according to either Kosambi's (1944) or Haldane's (1919) formulae; the least squares (LS) method is then used to estimate the mapping distance between any adjacent markers (Stam 1993).

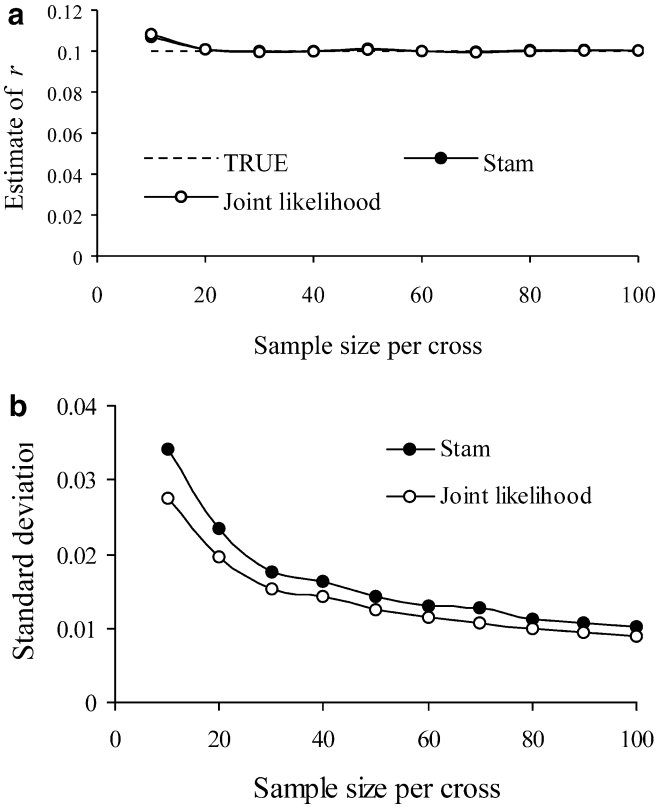

Fig. 2a, b Comparison between Stam's and the joint likelihood methods in terms of sample size, assayed with two codominant markers: **a** average estimates and **b** standard deviation. Eleven crosses are involved, including F_2, BC_1, BC_2, and the crosses between parents with a heterozygote at one locus (results are based upon 1,000 independent runs)

Results

An example

As a practical example, we applied our joint likelihood method to examine two annual plant species of *Linanthus* (Polemoniaceae): *L. jepsonii* and *L. bicolor*. *L. jepsonii* is mainly distributed in grassland and oak woodlands in North California, whereas *L. bicolor* occupies drier sites throughout California, Oregon and Washington. Although the two species are recently diverged, they can be distinguished with both vegetative and floral characters (Schemske and Goodwillie 1996; Goodwillie 2000). We obtained progeny populations F_2 of the interspecific hybrid *L. bicolor* × *L. jepsonii*, the backcross to the parent *L. bicolor* (BC_1), and to the parent *L. jepsonii* (BC_2). Seven populations in three families were assayed using amplified fragment length polymorphism (AFLP) markers: two populations in each backcross family and three F_2 populations. One hundred individuals were genotyped in each population, except that 257 individuals were genotyped in one F_2 population. A total of 162 polymorphic loci were scored.

Because there were some monomorphic markers in the BC_2, there were in total 9,150 pairs of markers for linkage analysis. Bootstrapping with 1,000 resamplings was used

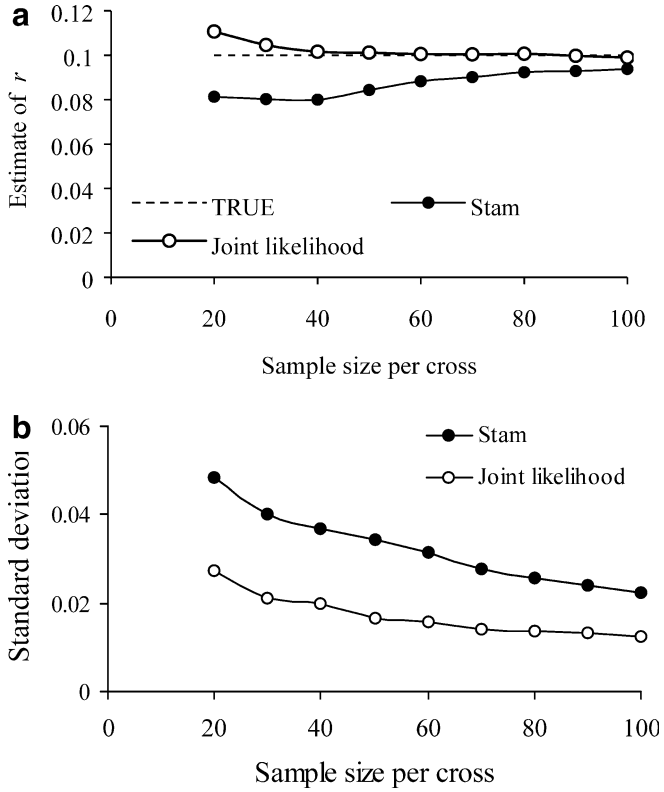

Fig. 3a, b Comparison between Stam's and the joint likelihood methods in terms of sample size, assayed with one codominant marker paired with one dominant marker: **a** average estimates and **b** standard deviation. Nine crosses are involved, including F_2, BC_1, and the crosses with parents heterozygous at one locus (1,000 independent runs)

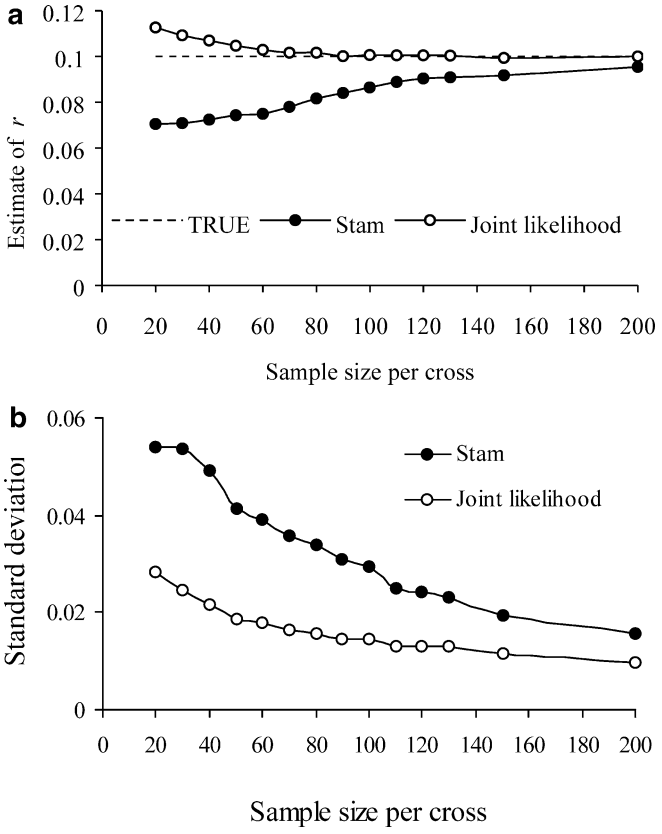

Fig. 4a, b Comparison between Stam's and the joint likelihood methods in terms of sample size, assayed with two dominant markers: **a** average estimates and **b** standard deviation. Nine crosses are involved, including F_2, BC_1, and the crosses with parents heterozygous at one locus (1,000 independent runs)

to calculate *P*-values. The pairwise significance level was set to 1×10^{-5} giving a global 5% level (Churchill and Doerge 1994). Linkages were found among 79 markers, or 49.8% of the total numbers of markers. These occurred in nine linkage groups, each with more than four markers, with three additional doublets and one triplet (Fig. 5). These linkage groups cover a genome size of 864.66 cM in total, with an average map distance of 13.3 cM between adjacent markers. The size of linkage groups ranged from 9.53 to 190.05 cM.

Discussion

Linkage mapping generally consists of three stages: estimation of the recombination rate of pairwise markers, marker grouping, and marker ordering. The originality of the present study is the development of a new approach at the first stage. We have described an approach that efficiently incorporates data from multiple crosses, and also can infer linkage phases, thus obviating the need for genotyping grandparents, and even parents, in outbred pedigrees. Our method can simultaneously estimate the recombination rate from multiple experiments, via maximizing the likelihood across experiments. Compared with Stam's procedure (Stam 1993; van Ooijen and Voorrips 2001), our simulations found that the joint likelihood approach gives estimates with higher precision and accuracy, particularly when dominant markers are used. With dominant markers, the required sample size can hence be significantly reduced when the joint likelihood method is used. The reason for the greater efficiency of our method over Stam's for dominant markers and/or F_2 populations is that the binomial sampling is implicitly designed for codominant markers and for when recombination events are directly observed in the data (e.g. backcross). This sampling distribution for r is not appropriate for the dominant markers, nor for the codominant markers in an F_2 population, nor for the combined dominant-codominant markers.

For the second and third stages, we have adopted the methods used in Stam's (1993) JoinMap, although several alternatives can be chosen (e.g. Liu 1998; Jansen et al. 2001; Rosa et al. 2002). Wu et al. (2003) recently showed that the statistical powers of alternative marker ordering methods are almost identical among five methods: maximum likelihood (ML), sum of adjacent LOD scores (SALOD), SARF, and one approximation algorithm-seriation (SER). Thus, our use of the SARF method seems reliable.

The use of joint likelihood is similar to previous methods applied to human genetic linkage mapping with

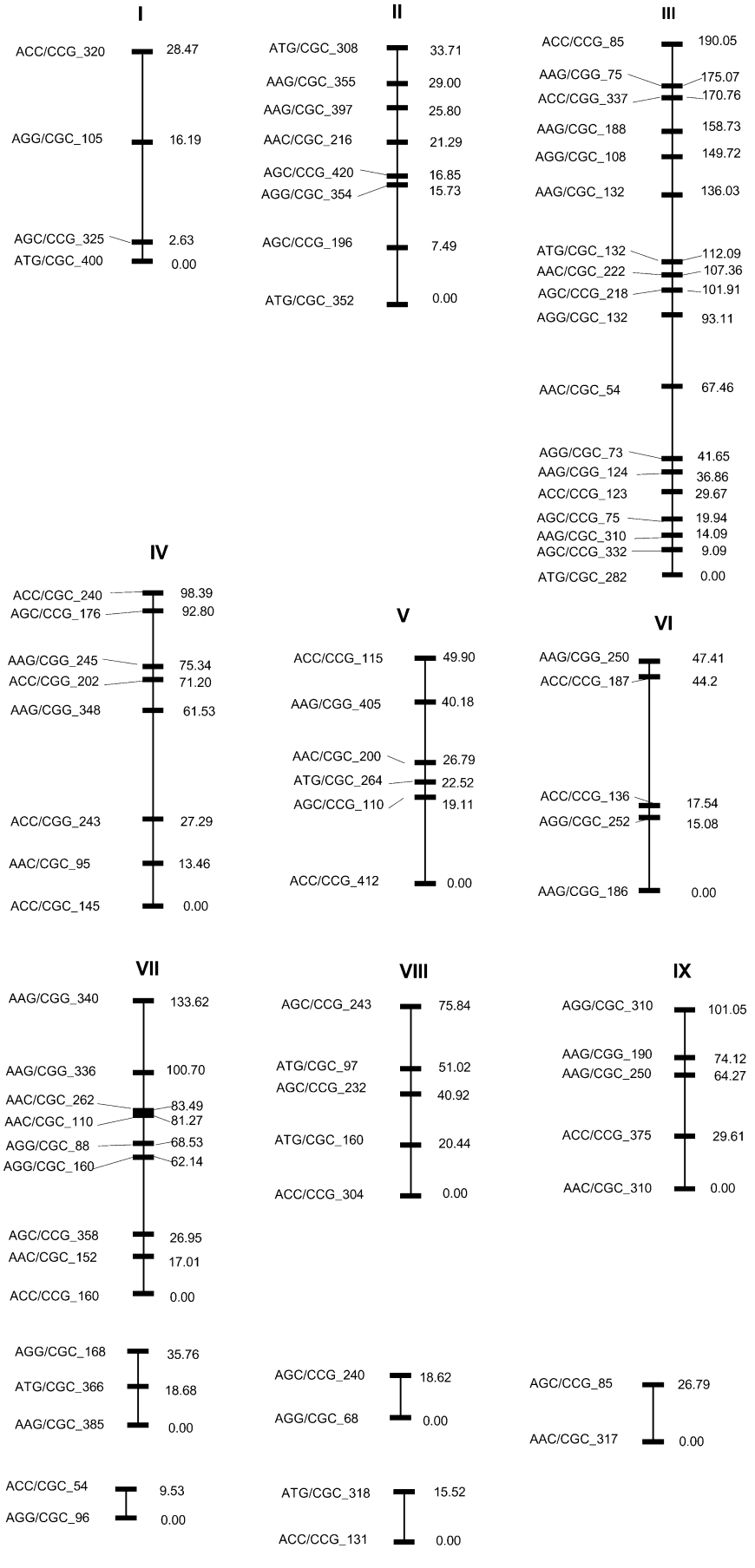

Fig. 5 Preliminary joint linkage maps obtained on the basis of seven progeny populations (one-generation data genotyped with AFLP markers): three F_2 populations of the interspecific hybrid *L. bicolor* × *L. jepsonii*, two populations of the backcross to *L. bicolor* (BC_1), and two populations of the backcross to *L. jepsonii* (BC_2)

multiple pedigrees, such as CRIMAP (Lander and Green 1987). The advantage of CRIMAP is that it considers the estimation of recombination rates and ordering of multiple markers simultaneously. The disadvantage of CRIMAP is that it needs two- or three-generation pedigree data and cannot provide the probability for the inference of parent linkage phases. Compared with CRIMAP, the present method can deal with one-generation data with the parent genotypes being either known or unknown, and provide flexibility in the choice of different types of markers. Furthermore, the parental linkage phases can be inferred when parent genotypes of each pedigree are not available.

Butcher et al. (2002) recently developed a linkage analysis package for outcrossed forest trees, OUTMAP, and demonstrated that their approach was better than JoinMap. They pointed out that an increase in the estimated genome length occurs because the distances between markers often differ from those calculated by JoinMap. Compared with JoinMap, the advantage of OUTMAP is the ability to conduct multi-locus likelihood analysis rather than pairwise analysis for the purpose of marker ordering. However, both OUTMAP and JoinMap are essentially indistinct from the mapping strategy point of view, in that they both first construct individual linkage maps and then integrate them. Moreover, there are several requirements when OUTMAP is applied: codominant markers and two generations of genotype data. These restrictions are not present in our method.

In addition to several concerns mentioned by Stam (1993) and Butcher et al. (2002), one caveat about the JoinMap/OUTMAP approach is that the accuracy of the integrated map depends upon how many segregating markers are in common among different maps. It remains to be examined how important these common markers are, particularly their distribution along linkage groups. With the present method, analyses of pairwise recombination rates for those unshared markers among all crosses remains restricted to those crosses that have "private" informative markers. Also, analyses of shared markers paired with unshared markers remain confined to those crosses that have both shared and unshared markers. Only "globally" shared informative markers are jointly analyzed.

Note that our method is suitable for data from multiple interspecific crosses and their variants (BC and F_2), rather than for data from half-sibs (e.g. open-pollinated progenies sampled from natural populations). There are no specific parental requirements; they can be either homozygous or heterozygous. The relatively posterior probability for a given set of progeny data from a single cross can actually be used to infer the linkage phases of corresponding individual parents (repulsion and coupling), unlike the method of Butcher et al. (2002).

Our method implicitly assumes that the true recombination fractions are constant across all crosses, unlike Stam (1993), who linearly combines the estimates of recombination across experiments in which recombination rates may vary. Events of abnormal meiosis, say deletions or translocations, could likely create divergence in map distance of some pairs of markers among different experimental populations. Also, an obvious case is a difference in recombination rates between male versus female meioses, and also differences between experiments conducted under differing environments such as temperature. A homogeneity test of the recombination fraction (e.g. Beavis and Grant 1991) may be performed before pooling individual maps, and only those individual maps with homogeneity in recombination fraction should be pooled. However, when the number of crosses with abnormal meiosis is small, which is probably the normal situation, the estimate of the recombination fraction would not be affected seriously. Sex-specific differences could pose larger problems, and again this issue warrants further investigation.

Acknowledgements We are grateful to Outi Savolainen and two anonymous referees for constructive comments on the earlier version of this article, and Dr. Carol Ritland for helpful discussion during the preparation of this article. The work was supported by a grant from Natural Sciences and Engineering Research Council (NSERC) to K.R.

References

Bailey NTJ (1961) Introduction to the mathematical theory of genetic linkage. Oxford University Press, London

Beavis WD, Grant D (1991) A linkage map based on information from four F_2 populations of maize (*Zea mays* L.). Theor Appl Genet 82:636–644

Burbridge A, Lindhout P, Grieve TM, Schumacher K, Theres K, van Heusden AW, Bonnema AB, Woodman KJ, Taylor IB (2001) Re-orientation and integration of the classical and interspecific linkage maps of the long arm of tomato chromosome 7. Theor Appl Genet 103:443–454

Butcher PA, Williams ER, Whitaker D, Ling S, Speed TP, Moran GF (2002) Improving linkage analysis in outcrossed forest trees - an example from *Acacia mangium*. Theor Appl Genet 104:1185–1191

Churchill GA, Doerge RW (1994) Empirical threshold values for quantitative trait mapping. Genetics 138:963–971

Dettori MT, Quarta R, Verde I (2001) A peach linkage map integrating RFLPs, SSRs, RAPDs, and morphological markers. Genome 44:783–790

Fisher RA, Balmakund B (1928) The estimation of linkage from the offspring of selfed heterozygotes. J Genet 20:79–92

Goodwillie C (2000) Inbreeding depression and mating systems in two species of *Linanthus* (Polemoniaceae). Heredity 84:283–293

Gosselin I, Zhou Y, Bousquet J, Isabel N (2002) Megagametophyte-derived linkage maps of white spruce (*Picea glauca*) based on RAPD, SCAR and ESTP markers. Theor Appl Genet 104:987–997

Grattapaglia D, Sederoff R (1994) Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross mapping strategy and RAPD markers. Genetics 137:1121–1137

Haldane JBS (1919) The combination of linkage values and the calculation of distance between the loci of linked factors. J Genet 8:299–309

Jansen J, de Jong AG, van Ooijen JW (2001) Constructing dense genetic linkage maps. Theor Appl Genet 102:1113–1122

Kosambi DD (1944) The estimation of map values from recombination values. Ann Eugen 12:172–175

Lander ES, Green P (1987) Construction of multilocus genetic linkage maps in humans. Proc Natl Acad Sci USA 84:2363–2367

Liu BH (1998) Statistical genomics: linkage, mapping, and QTL analysis. CRC Press, Boca Raton

Lombard V, Delourme R (2001) A consensus linkage map for rapeseed (*Brassica napus* L.): construction and integration of three individual maps from DH populations. Theor Appl Genet 103:491–507

Maliepaard C, Jansen J, Van Ooijen JW (1997) Linkage analysis in a full-sib family of an outbreeding plant species: overview and consequences for applications. Genet Res 70:237–250

Marques CM, Araujo JA, Ferreira JG, Whetten R, O'Malley DM, Liu BH, Sederoff R (1998) AFLP genetic maps of *Eucalyptus globules* and *E. tereticornis*. Theor Appl Genet 96:727–737

Mather K (1951) The measurement of linkage in heredity, 2nd edn. Methuen, London

Ooijen JW van, Voorrips RE (2001) JoinMap 3.0: software for the calculation of genetic linkage maps. Plant Research International, Wageningen

Ritland K (1986) Joint maximum likelihood estimation of genetic and mating structure using open-pollinated progenies. Biometrics 42:25–43

Ritland K (2002) Extensions of models for the estimation of mating systems using *n* independent loci. Heredity 88:221–228

Rosa GJM, Yandell BS, Gianola D (2002) A Bayesian approach for constructing genetic maps when markers are miscoded. Genet Sel Evol 34:353–369

Schemske DW, Goodwillie C (1996) Morphological and reproductive characteristics of *Linanthus jepsonii* (Polemoniaceae), a newly described, geographically restricted species from Northern California. Madrono 43:453–463

Sewell MM, Sherman BK, Neale DB (1999) A consense map for loblolly pine (*Pinus taeda* L.) I. Construction and integration of individual linkage maps from two outbred three-generation pedigrees. Genetics 151:321–330

Stam P (1993) Construction of integrated genetic linkage maps by means of a new computer package: JoinMap. Plant J 3:739–744

Suitter KA, Wendel JF, Case JS (1983) LINKAGE 1: a PASCAL computer program for the detection and analysis of genetic linkage. J Heredity 74:203–204

Testolin R, Huang WG, Lain O, Messian R, Vecchione A, Cipriani G (2001) A kiwifruit (*Actinidia* spp.) linkage map based on microsatellites and integrated with AFLP markers. Theor Appl Genet 103:30–36

Ulloa M, Meredith WR, Shappley ZW, Kahler AL (2002) RFLP genetic linkage maps from F2.3 populations and a joinmap of *Gossypium hirsutum*. Theor Appl Genet 104:200–208

Wu J, Jenkins J, Zhu J, McCarty J, Watson C (2003) Monte Carlo simulations on marker grouping and ordering. Theor Appl Genet 107:568–573

Wu RL, Ma CX, Wu SS, Zeng ZB (2002) Linkage mapping of sex-specific differences. Genet Res 79:85–96