Theor Appl Genet (2011) 123:681–691
DOI 10.1007/s00122-011-1618-4

ORIGINAL PAPER

A model for linkage analysis with apomixis

Wei Hou · Shen Lin · Yao Li · Xiaoming Pang · Yanru Zeng · Rongling Wu

Received: 22 February 2011 / Accepted: 9 May 2011 / Published online: 31 May 2011

Abstract Apomixis, or asexual reproduction through seeds, occurs in over 400 species of angiosperms. Although apomixis can favorably perpetuate desired genotypes through successive seed generation, it may also bring about some difficulty for linkage analysis and quantitative trait locus mapping. In this article, we explore the issue of how apomixis affects the precision and power of linkage analysis with molecular markers. We derive a statistical model for estimating the linkage between different markers when some progeny are derived from apomixis. The model was constructed within the maximum likelihood framework and implemented with the EM algorithm. A series of procedures are formulated to test the linkage of markers, the rate of apomixis, and the degree of genetic interference during meiosis. The model was examined and validated through simulation studies. The model will provide a tool for linkage mapping and evolutionary studies for plant species that undergo apomixis.

Communicated by M. Sillanpaa.

W. Hou and S. Lin contributed equally to this work.

W. Hou
Department of Biostatistics, University of Florida, Gainesville, FL 32611, USA

S. Lin · Y. Zeng (✉)
The Nurturing Station for the State Key Laboratory of Subtropical Silviculture, Zhejiang Agricultural and Forestry University, Lin'an 311300, Zhejiang, China
e-mail: yrzeng@zafu.edu.cn

Y. Li
Department of Statistics, West Virginia University, Morgantown, WV 26506, USA

X. Pang · R. Wu (✉)
Center for Computational Biology, Beijing Forestry University, Beijing 100083, China
e-mail: rwu@hes.hmc.psu.edu

X. Pang · R. Wu
National Engineering Laboratory for Tree Breeding, Beijing Forestry University, Beijing 100083, China

X. Pang · R. Wu
Key Laboratory of Genetics and Breeding in Forest Trees and Ornamental Plants, Beijing Forestry University, Beijing 100083, China

Introduction

Apomixis is an asexual type of reproduction in which the embryos of a plant grow from egg cells without being fertilized by pollen (Koltunow and Grossniklaus 2003; Bicknell and Koltunow 2004). As a natural way of cloning plants, apomixis affords us an excellent vehicle to perpetuate any genotype of interest in agriculture, no matter how complex, through successive seed generations (Spillane et al. 2004; d'Erfurth et al. 2009). Just because of its tremendous potential for plant breeding, attempts have been made to study the different cytological mechanisms of apomixis (Koltunow 1993), its genetic control (Grimanelli et al. 2001), its relationship to the sexual pathway (Spillane et al. 2001), and its use in agriculture (Hanna 1995). Apomixis reflects a developmental alteration through which reproductive events in the ovule need to be reprogrammed (Koltunow and Grossniklaus 2003). More recently, molecular analyses have been performed to isolate, characterize, and transfer genes that are directly involved in controlling apomixis

(d'Erfurth et al. 2009), aimed to engineering this trait into crops.

Despite its significant value to agriculture, apomixis presents a problem for linkage analysis in an experimental cross, thus impeding the construction of genetic linkage maps and subsequent quantitative trait locus mapping. The principle of linkage analysis is founded on the crossover of different linked genes between two homologous chromosomes at meiosis, leading to recombinant gametes (Lathro et al. 1984; Lu et al. 2004; Wu et al. 2007). By analyzing the proportion of recombinant gametes to the total amount of gametes (including the recombinants and non-recombinants) from observed zygotic genotypes in a family, the degree of linkage is quantified and tested. When apomixis is a mechanism for reproduction, some of the progenies identical to the female parent will not experience the crossover in meiosis, which are asexually derived from the same genotype of the female parent. Thus, because it is not possible to distinguish the progeny derived from the crossover from those from apomixis, a direct use of traditional linkage analysis purely based on the crossover will not work.

The motivation of this study is to develop a statistical model for estimating and testing the linkage of molecular markers by integrating apomixis into a linkage analysis setting. The model is based on a likelihood of progeny marker data for an experimental cross, with those maternal-like progeny genotypes containing two mixture components, one from the meiosis and the second from apomixis. Two unknown parameters, the linkage and the degree of apomixis, that define the mixture model are estimated by using the maximum likelihood estimating principle. The EM algorithm is implemented to obtain the maximum likelihood estimates of these two parameters. The model is first constructed for two markers and then extended to include three markers in which the impact of genetic interference on the linkage estimation is investigated (Lu et al. 2004). Simulation studies with different sample sizes were performed to test and validate the model. A real example was used to demonstrate the usefulness of the model.

Table 1 The frequencies of four genotypes at two markers in a backcross with progeny arising from a mix of meiotic crossovers and apomixis

Genotype	Diplotype configuration	Frequency	Observation
AaBb	$\begin{cases} AaBb \\ AB\|ab \end{cases}$	$\begin{cases} \theta \\ \frac{1}{2}(1-r)(1-\theta) \end{cases}$	n_{11}
Aabb	*Ab|ab*	$\frac{1}{2}r(1-\theta)$	n_{10}
aaBb	*aB|ab*	$\frac{1}{2}r(1-\theta)$	n_{01}
aabb	*ab|ab*	$\frac{1}{2}(1-r)(1-\theta)$	n_{00}

Model and method

Two-point analysis

We derive the model for linkage analysis for a backcross, initiated with two inbred lines. It can be extended to an F_2 design without technical difficulty although the statistical properties of the model may change due to the complexity of data structure. Consider two markers, **A** with two alleles *A* and *a* and **B** with two alleles *B* and *b*. The F_1 progeny *AaBb* derived from inbred lines *AABB* and *aabb* will produce four gametes in meiosis, *AB*, *Ab*, *aB*, and *ab*. These four gametes have a frequency of $\frac{1}{2}(1-r), \frac{1}{2}r, \frac{1}{2}r$, and $\frac{1}{2}(1-r)$, respectively, if the two markers have a recombination fraction of *r*. If the F_1 as a female parent is mated to original male parent *aabb*, then we have four backcross genotypes, *AaBb*, *Aabb*, *aaBb*, and *aabb* with respective observations n_{11}, n_{10}, n_{01}, and n_{00} summed to *n*. Apomixis means that a female parent produces progeny with the same genotype as hers. Thus, of progeny genotype *AaBb*, some are generated from meiosis, whereas the others from apomixis. Let θ denote the proportion of the progeny that are derived due to apomixis, and thus, $1-\theta$ denotes the progeny generated due to the meiosis. Tables 1 and 2 give the frequencies of the four backcross genotypes at two and three markers when the progenies are formed through meiosis and apomixis.

The multinomial log-likelihood for marker observation **M** can be expressed as

$$\begin{aligned} \log L(\boldsymbol{\Theta}|\mathbf{M}) &= \text{constant} \\ &+ n_{11}\log\left[\theta + \frac{1}{2}(1-r)(1-\theta)\right] \\ &+ (n_{10}+n_{01})\log[r(1-\theta)] \\ &+ n_{00}\log[(1-r)(1-\theta)], \end{aligned} \tag{1}$$

Table 2 The frequencies of four genotypes at three markers in a backcross with progeny arising from a mix of meiotic crossovers and apomixis

Genotype	Diplotype configuration	Frequency	Observation
AaBbCc	$\begin{cases} AaBbCc \\ ABC\|abc \end{cases}$	$\begin{cases} \theta \\ \frac{1}{2}g_{00}(1-\theta) \end{cases}$	n_{111}
aabbcc	*abc|abc*	$\frac{1}{2}g_{00}(1-\theta)$	n_{000}
AaBbcc	*ABc|abc*	$\frac{1}{2}g_{01}(1-\theta)$	n_{110}
aabbCc	*abC|abc*	$\frac{1}{2}g_{01}(1-\theta)$	n_{001}
Aabbcc	*Abc|abc*	$\frac{1}{2}g_{10}(1-\theta)$	n_{100}
aaBbCc	*aBC|abc*	$\frac{1}{2}g_{10}(1-\theta)$	n_{011}
AabbCc	*AbC|abc*	$\frac{1}{2}g_{11}(1-\theta)$	n_{101}
aaBbcc	*aBc|abc*	$\frac{1}{2}g_{11}(1-\theta)$	n_{010}

where $\mathbf{\Theta} = (r, \theta)$ contains the unknown parameters. The second term of the log-likelihood (Eq. 1) contains two components due to the apomixis and meiosis, respectively, that are mixed together to define the frequency of maternal-like progeny genotypes. The last two terms are log-likelihoods for the progeny genotypes that are different from the maternal genotype.

To estimate the unknowns in log-likelihood (Eq. 1), we implement the EM algorithm to handle the missing data of the log-likelihood, i.e., the origin of the maternal-like progeny genotype, either apomixic or meiotic. In the E step, we calculate the posterior probability of the progeny with genotype *AaBb* being derived from apomixis using

$$\psi = \frac{\theta}{\frac{1}{2}(1-r)(1-\theta)+\theta}, \tag{2}$$

In the M step, the parameters are estimated using

$$\hat{\theta} = \frac{n_{11}\psi}{n}, \tag{3}$$

$$\hat{r} = \frac{n_{10}+n_{01}}{n-n_{11}\psi}. \tag{4}$$

By taking a value within the space of each unknown parameter as a starting value, we calculate the posterior probability in the E step (Eq. 2) and estimate the parameters in the M step (Eqs. 3, 4). Both the steps are iterated until stable estimates of these parameters are obtained. These stable estimates are the maximum likelihood estimates of the recombination fraction and apomixis proportion.

After the parameters are estimated, the next step is to test the significance of the linkage and apomixis occurrence, respectively. The log-likelihood under H_0: $\theta = 0$ is reduced to one for a traditional linkage analysis, expressed as

$$\log L(r|\mathbf{M}) = \text{constant} + (n_{11}+n_{00})\log(1-r) + (n_{10}+n_{01})\log r$$

for which no EM algorithm is needed to estimate the linkage for the backcross design (Wu et al. 2007). The log-likelihood ratio calculated as $-2(\log L(r|\mathbf{M}) - \log L(\mathbf{\Theta}|\mathbf{M}))$ is used to test the significance of apomixis rate. Since the apomixis rate is at the boundary of its parameter space under the null hypothesis, a standard χ^2 test is not appropriate for testing the significance of this parameter. As a result, we use a simulation approach to determine the critical threshold for the significance test of the apomixis rate.

Under $H_0 : r = \frac{1}{2}$ (there is no linkage), the log-likelihood is changed as

$$\log L(\theta|\mathbf{M}) = \text{constant} + n_{11}\log\left[\theta + \frac{1}{4}(1-\theta)\right] + (n_{10}+n_{01}+n_{00})\log(1-\theta). \tag{5}$$

The EM algorithm is required to estimate the rate of apomixis. In the E step, the expected proportion of apomixis within genotype *AaBb* is calculated by

$$\phi = \frac{\theta}{\theta + \frac{1}{4}(1-\theta)}.$$

In the M step, the apomixis rate is calculated as

$$\theta = \frac{n_{11}\phi}{n}.$$

The log-likelihood ratio is then calculated using $\log L(\theta|\mathbf{M})$ under the null hypothesis (5) and $\log L(\Theta|\mathbf{M})$ under the alternative hypothesis (Eq. 1). Such a test statistic can be used to test the significance of the linkage based on a standard χ^2 test approach.

If the F_1 is used as the male parent to cross with the original female parent *aabb*, then the same four backcross genotypes are produced. Yet, in this case, backcross genotype *aabb* contains two types, one from meiosis and the second from apomixis. The same procedure described above can be used for parameter estimation and test.

Three-point analysis

Consider three markers in order **A**–**B**–**C** that are segregating in the backcross. The triple heterozygote F_1 produces eight different gametes in meiosis which are classified into four types: (1) *ABC* and *abc* in which there is neither crossover between markers **A** and **B**, nor between markers **B** and **C**, (2) *ABc* and *abC* in which there is no crossover between markers **A** and **B** but one crossover between markers **B** and **C**, (3) *aBC* and *Abc* in which there is one crossover between markers **A** and **B** but no crossover between markers **B** and **C**, and (4) *AbC* and *aBc* in which there is one crossover between markers **A** and **B** as well as between markers **B** and **C**. Let g_{00}, g_{01}, g_{10}, and g_{11} denote the frequencies of these four types, respectively. Thus, we can express the frequencies of the eight gametes in terms of these g values.

If the recombination fractions between markers **A** and **B**, markers **B** and **C**, and markers **A** and **C** are expressed as r_1, r_2, and r_3, respectively, then we have

$$\begin{aligned} r_1 &= g_{11} + g_{10} \\ r_2 &= g_{11} + g_{01} \\ r_3 &= g_{10} + g_{01}. \end{aligned} \tag{6}$$

It is easy to get

$$\begin{aligned} g_{10} &= \frac{1}{2}(r_1 - r_2 + r_3) \\ g_{11} &= \frac{1}{2}(r_1 + r_2 - r_3) \\ g_{01} &= \frac{1}{2}(r_3 - r_1 + r_2) \\ g_{00} &= 1 - g_{10} - g_{11} - g_{01}. \end{aligned} \tag{7}$$

If the coefficient of coincidence between marker intervals **A–B** and **B–C** is defined as C, we have $r_3 = r_1 + r_2 - 2Cr_1r_2$ and

$$C = \frac{r_1 + r_2 - r_3}{2r_1r_2}, \tag{8}$$

with the bound of C being $[0, min(\frac{1}{r_1}, \frac{1}{r_2})]$.

In the backcross with eight genotypes, triple heterozygote *AaBbCc* may be formed from meiosis and apomixis. With the apomixis proportion θ, we express the log-likelihood of eight genotypic observations **M** as

$$\begin{aligned}\log L(\mathbf{\Theta}|\mathbf{M}) &= \text{constant} \\ &+ n_{111}\log[\theta + \frac{1}{2}g_{00}(1-\theta)] \\ &+ n_{000}\log[\frac{1}{2}g_{00}(1-\theta)] \\ &+ (n_{110}+n_{001})\log[g_{01}(1-\theta)] \\ &+ (n_{100}+n_{011})\log[g_{10}(1-\theta)] \\ &+ (n_{010}+n_{101})\log[g_{11}(1-\theta)]\end{aligned} \tag{9}$$

The EM algorithm is implemented to estimate the parameters. In the E step, we calculate the proportion of the progeny that is derived from apomixis within genotype *AaBbCc* by

$$\psi = \frac{\theta}{\frac{1}{2}g_{00}(1-\theta)+\theta}. \tag{10}$$

In the M step, the parameters are estimated using

$$\begin{aligned}\hat{\theta} &= \frac{n_{111}\psi}{n} \\ \hat{g_{00}} &= \frac{n_{111}(1-\psi)+n_{000}}{n-n_{111}\psi} \\ \hat{g_{01}} &= \frac{n_{110}+n_{001}}{n-n_{111}\psi} \\ \hat{g_{10}} &= \frac{n_{100}+n_{011}}{n-n_{111}\psi} \\ \hat{g_{11}} &= 1-\hat{g_{00}}-\hat{g_{01}}-\hat{g_{10}}.\end{aligned} \tag{11}$$

There are three hypothesis tests for three-point analysis. The first is about the proportion of apomixis. When there is no apomixis, the log-likelihood is reduced to the traditional three-point analysis model (Wu et al. 2007). The second test is about the linkage between each pair of markers. Under $H_0 : r_1 = r_2 = r_3 = \frac{1}{2}$, the estimate of θ needs the implementation of the EM algorithm as shown for the two-point analysis. If the test is made for one or two of these recombination fractions, the same EM procedure can be used, with the corresponding constraints constructed from Eq. 6.

The third test is about the degree of genetic interference defined as $1 - C$. We are interested in two cases, $C = 0$ (no double crossover) and $C = 1$ (no interference). The EM algorithm can be implemented to estimate other parameters under a specific null hypothesis, with the constraints derived from the relationships among different recombination fractions.

Computer simulation

We carried out simulation studies to examine the statistical properties of the model. Backcross populations with different sample sizes (100, 200, 300, 400, 500, and 600) were considered. Marker genotypes were simulated by assuming different proportions of apomixis ($\theta = 0$, 0.05, 0.10, 0.20, 0.30, 0.40, 0.50, 0.60) and different degrees of linkage ($r = 0.05$, 0.10, 0.15, 0.20, 0.25). The simulated data were analyzed by the new model (incorporating apomixis) and traditional model (not incorporating apomixis). Table 3 lists part of the results about parameter estimation under several representative apomixis proportions and linkage degrees under sample size 100, 200, and 500. In general, the new model provides reasonably precise estimates of the apomixis proportion and linkage. If there is no apomixis, the new model will incorrectly obtain a small estimate of apomixis proportion, but this false positive rate can be very low when a sample size increases to 400–500. If there exists a low proportion of apomixis, the new model can well estimate this parameter even with a modest sample size 100–200. A high value of apomixis proportion can well be estimated even with a small sample size 100.

If there is no apomixis, the new and traditional models estimate the linkage with similar precision and accuracy (Table 3). As expected, the estimation of the linkage is increasingly biased with increasing apomixis proportions using the traditional model. For example, with $\theta = 0.5$, the traditional model obtains the estimate of 0.0238 for a true recombination fraction of 0.05 when 100 samples are used. Under the same condition, the estimate is 0.0502 by the new model. In general, the new model provides a good estimate of the recombination rate. The discrepancy between the two models is magnified when two markers are weakly linked.

As compared with two-point analysis, three-point analysis is more informative in terms of the estimation of genetic interference. Three ordered markers were simulated under three assumptions, i.e., no double crossover ($C = 0$), independent meiosis ($C = 1$), and genetic interference ($C = 2$). The results from three-point analysis support those from two-point analysis. Figures 1, 2 and 3 compare the estimates of the linkage between the first and third markers with its true values under three different assumptions. In any case, the new model provides unbiased estimates of the linkage compared to the

Table 3 The maximum likelihood estimates (MLEs) of the apomixis proportion and linkage for a simulated backcross of different sample sizes obtained by the new and traditional models

True value		New model		Traditional model
θ	r	$\hat{\theta}$	$\hat{r}$	$\hat{r}$
$n = 100$				
0.00	0.05	0.0388 (0.0566)	0.0519 (0.0231)	0.0497 (0.0219)
0.00	0.25	0.0338 (0.0498)	0.2590 (0.0469)	0.2496 (0.0434)
0.10	0.05	0.1082 (0.0844)	0.0512 (0.0242)	0.0453 (0.0209)
0.10	0.25	0.1060 (0.0770)	0.2528 (0.0515)	0.2244 (0.0420)
0.20	0.05	0.2008 (0.0942)	0.0505 (0.0254)	0.0399 (0.0195)
0.20	0.25	0.2009 (0.0861)	0.2528 (0.0550)	0.2001 (0.0396)
0.40	0.05	0.4007 (0.0903)	0.0506 (0.0295)	0.0299 (0.0170)
0.40	0.25	0.4003 (0.0830)	0.2531 (0.0633)	0.1499 (0.0356)
0.60	0.05	0.6006 (0.0782)	0.0513 (0.0362)	0.0201 (0.0139)
0.60	0.25	0.6001 (0.0728)	0.2538 (0.0798)	0.0996 (0.0302)
$n = 200$				
0.00	0.05	0.0270 (0.0396)	0.0516 (0.0160)	0.0501 (0.0153)
0.00	0.25	0.0235 (0.0347)	0.2565 (0.0325)	0.2501 (0.0303)
0.10	0.05	0.1026 (0.0640)	0.0506 (0.0169)	0.0451 (0.0147)
0.10	0.25	0.1007 (0.0592)	0.2508 (0.0361)	0.2247 (0.0296)
0.20	0.05	0.1992 (0.0679)	0.0503 (0.0178)	0.0401 (0.0139)
0.20	0.25	0.2004 (0.0614)	0.2511 (0.0386)	0.1999 (0.0283)
0.40	0.05	0.4004 (0.0639)	0.0506 (0.0209)	0.0301 (0.0121)
0.40	0.25	0.4003 (0.0591)	0.2527 (0.0446)	0.1506 (0.0251)
0.60	0.05	0.5996 (0.0554)	0.0508 (0.0256)	0.0201 (0.0100)
0.60	0.25	0.6005 (0.0525)	0.2531 (0.0555)	0.1001 (0.0210)
$n = 500$				
0.00	0.05	0.0177 (0.0257)	0.0511 (0.0100)	0.0501 (0.0097)
0.00	0.25	0.0158 (0.0228)	0.2544 (0.0208)	0.2503 (0.0195)
0.10	0.05	0.0999 (0.0431)	0.0500 (0.0105)	0.0449 (0.0093)
0.10	0.25	0.1003 (0.0387)	0.2502 (0.0228)	0.2247 (0.0188)
0.20	0.05	0.2005 (0.0426)	0.0501 (0.0111)	0.0399 (0.0087)
0.20	0.25	0.1998 (0.0391)	0.2503 (0.0245)	0.2000 (0.0181)
0.40	0.05	0.4003 (0.0404)	0.0502 (0.0129)	0.0300 (0.0076)
0.40	0.25	0.3996 (0.0373)	0.2506 (0.0279)	0.1501 (0.0159)
0.60	0.05	0.6000 (0.0352)	0.0502 (0.0159)	0.0200 (0.0063)
0.60	0.25	0.6002 (0.0332)	0.2503 (0.0343)	0.0997 (0.0134)

The standard deviations of the MLEs estimated from 10,000 simulation replicates are given in parentheses

traditional model. This can be seen from Figs. 1, 2 and 3, in which the estimate curve from the new model tends to be on the diagonal, whereas the curve from the traditional model deviates from the diagonal. If crossovers in adjacent marker intervals are not independent, the new model shows an advantage in the linkage estimation over the traditional model.

To compare the difference in the estimation precision of the apomixis proportion, a simulation study was performed by assuming three ordered markers, with the recombination fraction 0.05 between the first two and 0.25 between the second two, respectively, as well as a coincidence coefficient $C = 2$ under different θ values from 0 to 0.6. Such three-marker data were analyzed by two- and three-point models. For the two-point analysis, we have two estimates of the apomixis proportion based on the linkage of the first two markers and second two markers, respectively, from which an average was obtained. As shown in Table 4, the three-point analysis provide a better estimate of the apomixis proportion than the two-point analysis, in terms of accuracy and precision (Table 4). The advantage of the former over the latter increases with increasing apomixis proportion.

A simulation study was performed to investigate the power of the model to detect apomixis and its false positive rates (FPR) when no apomixis actually exists. The data were simulated by assuming three ordered markers, with the recombination fraction 0.05 between the first two and 0.25 between the second two, respectively, as well as a coincidence coefficient $C = 2$ under different θ values from 0 to 0.6. Table 5 lists the proportions of those simulation replicates, in which a significant apomixis rate was detected on the basis of the critical threshold determined by a simulation approach, over a total of simulation replicates (200). Thus, the proportion under $\theta = 0$ reflects the FPR, whereas those under $\theta > 0$ are the power for apomixis detection. The FPR was detected to be low for both two- and three-point analysis (0.02–0.03). If the apomixis rate is below 0.10, three-point analysis has a little bit greater power for apomixis detection than two-point analysis. From these simulation results (Table 5), we can see that three-point analysis is more conservative for apomixis detection than two-point analysis partly because three-point analysis has an additional parameter C that needs to be estimated.

We also performed a simulation study to examine the power for linkage detection (Table 6). We simulated three ordered markers, with the recombination fraction 0.05 between the first two and 0.25 between the second two, respectively, as well as a coincidence coefficient $C = 2$ under different θ values from 0 to 0.6. Under the circumstance of no apomixis, the power of linkage detection is similar between two- and three-point analysis if there is no genetic interference (Lu et al. 2004). Our simulation

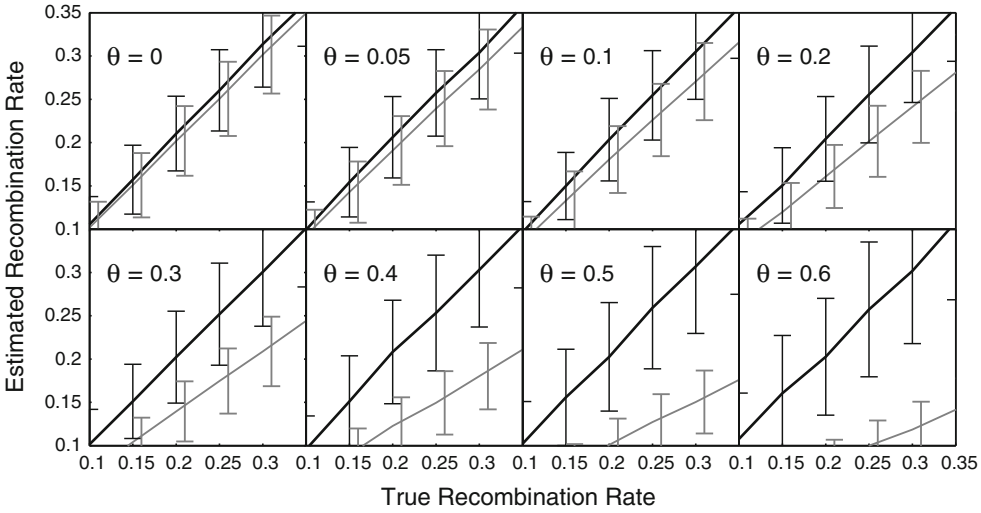

Fig. 1 Plots of the estimated values of the recombination fraction against its true value under different degrees of linkage obtained from three-point analysis in the case of no double crossovers ($C = 0$). The *dark colored curves* are the results from the new model incorporating apomixis, and the *light colored curves* are the results from the traditional model not incorporating apomixis. The *vertical bars* are the standard errors of each estimate

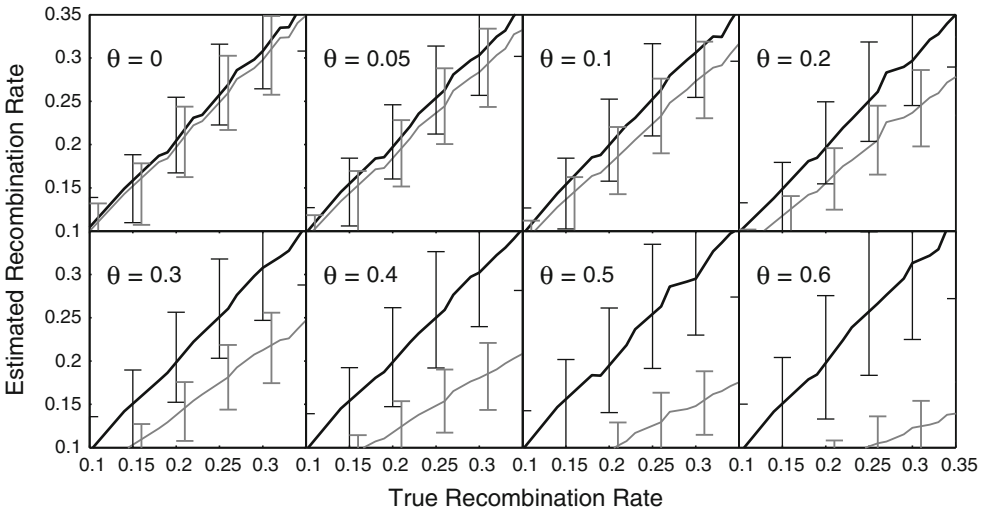

Fig. 2 Plots of the estimated values of the recombination fraction against its true value under different degrees of linkage obtained from three-point analysis in the case of no double crossovers ($C = 1$). The *dark colored curves* are the results from the new model incorporating apomixis, and the *light colored curves* are the results from the traditional model not incorporating apomixis. The *vertical bars* are the standard errors of each estimate

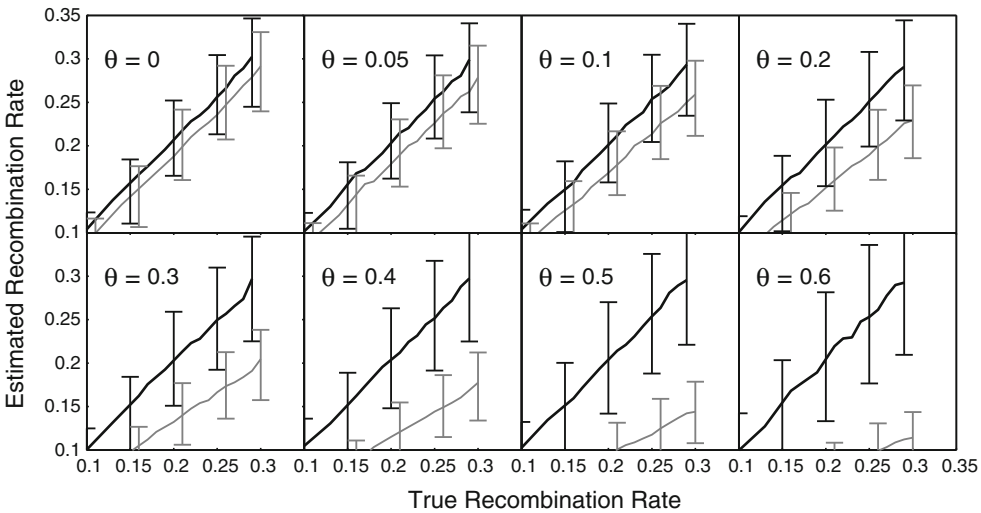

Fig. 3 Plots of the estimated values of the recombination fraction against its true value under different degrees of linkage obtained from three-point analysis in the case of no double crossovers ($C = 2$). The *dark colored curves* are the results from the new model incorporating apomixis, and the *light colored curves* are the results from the traditional model not incorporating apomixis. The *vertical bars* are the standard errors of each estimate

assumes a strong genetic interference ($C = 2$), which causes three-point analysis to be less powerful than two-point analysis (Table 6). If apomixis exists, the power of linkage test tends to reduce with increasing apomixis degree. When genetic interference exists, three-point analysis shows lower power for linkage detection than two-

Table 4 The maximum likelihood estimates (MLEs) of the apomixis proportion by two- and three-point analyses

True value	Two-point	Three-point
$n = 100$		
0.00	0.0345 (0.0510)	0.0340 (0.0503)
0.05	0.0648 (0.0664)	0.0644 (0.0655)
0.10	0.1051 (0.0784)	0.1051 (0.0775)
0.20	0.2003 (0.0864)	0.2001 (0.0852)
0.30	0.3012 (0.0846)	0.3012 (0.0836)
0.40	0.3993 (0.0831)	0.3992 (0.0822)
0.50	0.4999 (0.0800)	0.4999 (0.0791)
0.60	0.6008 (0.0730)	0.6008 (0.0723)
$n = 200$		
0.00	0.0243 (0.0357)	0.0239 (0.0351)
0.05	0.0589 (0.0514)	0.0585 (0.0506)
0.10	0.1025 (0.0597)	0.1024 (0.0590)
0.20	0.1997 (0.0617)	0.1998 (0.0609)
0.30	0.2990 (0.0617)	0.2990 (0.0610)
0.40	0.4000 (0.0582)	0.3999 (0.0576)
0.50	0.5006 (0.0557)	0.5006 (0.0551)
0.60	0.6003 (0.0522)	0.6003 (0.0517)
$n = 500$		
0.00	0.0156 (0.0229)	0.0154 (0.0225)
0.05	0.0514 (0.0356)	0.0513 (0.0350)
0.10	0.0994 (0.0389)	0.0995 (0.0384)
0.20	0.1996 (0.0392)	0.1995 (0.0387)
0.30	0.3002 (0.0378)	0.3000 (0.0374)
0.40	0.4006 (0.0376)	0.4005 (0.0372)
0.50	0.5006 (0.0353)	0.5006 (0.0349)
0.60	0.6003 (0.0329)	0.6003 (0.0326)

Table 5 The false positive rate (in boldface) and power for apomixis detection calculated from 10,000 simulation replicates by two- and three-point analyses

	$n = 100$		$n = 200$		$n = 500$	
	Two	Three	Two	Three	Two	Three
0	**0.02**	**0.03**	**0.02**	**0.03**	**0.02**	**0.03**
0.05	0.07	0.07	0.11	0.13	0.24	0.25
0.10	0.17	0.18	0.32	0.38	0.72	0.73
0.20	0.52	0.52	0.82	0.88	1	1
0.30	0.85	0.73	0.99	0.96	1	1
0.40	0.94	0.70	1	0.93	1	1
0.50	0.92	0.62	0.99	0.88	1	1
0.60	0.89	0.55	0.98	0.83	1	1

point analysis, especially when sample size is small. With a sample size of 200, three-point analysis can provide adequate power even when the apomixis rate is high.

Table 6 Power for linkage test calculated from 10,000 simulation replicates by two- and three-point analyses

	Two-point		Three-point		
	r_1	r_2	r_1	r_2	r_3
$n = 100$					
0	0.99	0.99	0.91	0.84	0.84
0.05	0.99	0.99	0.88	0.81	0.81
0.10	0.99	0.98	0.87	0.79	0.79
0.20	0.98	0.97	0.84	0.73	0.73
0.30	0.97	0.95	0.79	0.64	0.64
0.40	0.95	0.91	0.74	0.55	0.55
0.50	0.92	0.85	0.67	0.44	0.44
0.60	0.86	0.77	0.58	0.31	0.31
$n = 200$					
0	1	1	0.99	0.99	0.99
0.05	1	1	0.99	0.98	0.98
0.10	1	1	0.99	0.98	0.98
0.20	1	1	0.98	0.96	0.97
0.30	1	1	0.96	0.94	0.94
0.40	1	1	0.94	0.90	0.90
0.50	0.99	0.99	0.90	0.83	0.84
0.60	0.98	0.97	0.84	0.72	0.72
$n = 500$					
0	1	1	1	1	1
0.05	1	1	1	1	1
0.10	1	1	1	1	1
0.20	1	1	1	1	1
0.30	1	1	1	1	1
0.40	1	1	1	1	1
0.50	1	1	1	1	1
0.60	1	1	0.99	0.99	0.99

The recombination fractions between the first and two markers, the second and third markers, and the first and third markers are expressed as r_1, r_2, and r_3, respectively

Worked example

The method was used to analyze a real data set collected from a genetic study in hickory (*Carya cathayensis*). Hickory is one of the species in which apomixis is believed to occur in pollination (Zhang et al., unpublished results). The model was used to test if apomixis exists in this species. A full-sib family of 100 progenies was made by crossing hickory (father) and pecan (*Carya illinoensis*) (mother). DNAs were isolated from young leaves of these hybrids and the two parents, from which dominant markers were detected with SRAP, ISSR, and RAPD techniques. Although these markers, expressed as presence/absence of a polymorphic band, confound the two genotypes containing the dominant allele, they are fully informative for a backcross design. For a full-sib family derived from

Table 7 The MLEs of the recombination fractions (r) and apomixis rate (θ) using three markers from a linkage group in a hickory × pecan F_1 hybrid progeny

	Two-point analysis		Three-point analysis	
	r	θ	r	θ
F1_R3_1				
|	0.20	0.30	0.20	
F4_R1_5	0.31	0.23	0.20	0.30
|	0.16	0.63	0	
ISSR_22_4				

outcrossing parents, Grattapaglia and Sederoff (1994) proposed a so-called two-way pseudo-testcross strategy for linkage analysis and map construction with segregating dominant markers for which one parent is heterozygous, whereas the other is null. Using this strategy, two parent-specific linkage maps will be constructed. We will use the pseudo-testcross strategy to perform linkage analysis with markers from the full-sib family studied.

We selected three markers, F1_R3_1, F4_R1_5, and ISSR_22_4, which each are heterozygous for the pecan parent and null for the hickory parent. Both two- and three-point analyses were used to analyze the linkage among these three markers with results tabulated in Table 7. The pair-wise recombination fractions of these markers were estimated as 0.20, 0.31, and 0.16 for markers F1_R3_1 and F4_R1_5, markers F1_R3_1 and ISSR_22_4, and markers F4_R1_5, and ISSR_22_4, respectively. Each pair provides the estimation of apomixis rate, 0.30, 0.23, and 0.63. By significance tests, we found that each of these estimates of apomixis rate is significant at $P = 0.05$. Three-point analysis shows that these three markers are located in the same linkage group, with an optimal order F1_R3_1 – F4_R1_5 – ISSR_22_4. Three-point analysis is thought to provide a more precise estimation of recombination fractions than two-point analysis because more information is used in the former. According to three-point analysis, the recombination fractions are 0.2, 0, and 0.2 between markers F1_R3_1 and F4_R1_5, markers F1_R3_1 and ISSR_22_4, and markers F4_R1_5 and ISSR_22_4, respectively. The rate of apomixis by three-point analysis is 0.30, which is significant at $P = 0.05$ by a significance test.

Extension to the F_2 design

Consider an F_2 population of size n in which there are three genotypes, i.e., homozygote for the capital allele (2), homozygote for the small allele (0), and heterozygote (1). Table 8 gives the distribution of nine genotypes at two markers arising from meiotic crossovers and apomixis in the F_2 population. The log-likelihood for genotypic observations at two markers (**M**) can be expressed as

$$\begin{aligned}\log L(\mathbf{\Theta}|\mathbf{M}) &= \text{constant} \\ &+ (n_{22}+n_{00})\log[(1-r)^2(1-\theta)] + (n_{20}+n_{02})\log(r^2(1-\theta)) \\ &+ (n_{21}+n_{12}+n_{10}+n_{01})\log[r(1-r)(1-\theta)] \\ &+ n_{11}\log[\theta+\frac{1}{2}r^2(1-\theta)+\frac{1}{2}(1-r)^2(1-\theta)] \end{aligned} \quad (12)$$

where $\mathbf{\Theta}$ is the unknown parameters containing the apomixis proportion (θ) and recombination fraction (r). The EM algorithm is implemented to estimate the unknown parameters. In the E step, we calculate the proportion of the progeny arising from apomixis (ψ) and the number of crossovers (ϕ) within double heterozygote *AaBb* using

$$\psi = \frac{\theta}{\frac{1}{2}r^2(1-\theta)+\frac{1}{2}(1-r)^2(1-\theta)+\theta} \quad (13)$$

$$\phi = \frac{r^2}{r^2+(1-r)^2} \quad (14)$$

In the M step, these parameters are estimated using

$$\hat{\theta} = \frac{n_{11}\psi}{n} \quad (15)$$

$$\hat{r} = \frac{(n_{21}+n_{12}+n_{10}+n_{01})+2(n_{20}+n_{02})+2\phi n_{11}\bar{\psi}}{n-n_{11}\psi} \quad (16)$$

A similar procedure is derived for three-point analysis (Table 9). The log-likelihood of genotypic observations for three markers can be expressed as

Table 8 The frequencies of four genotypes at two markers in an F_2 with progeny arising from a mix of meiotic crossovers and apomixis

Genotype	Diplotype configuration	Frequency	Observation
AABB	*AB|AB*	$\frac{1}{4}(1-r)^2(1-\theta)$	n_{22}
AABb	*AB|Ab*	$\frac{1}{2}r(1-r)(1-\theta)$	n_{21}
AAbb	*Ab|Ab*	$\frac{1}{4}r^2(1-\theta)$	n_{20}
AaBB	*AB|aB*	$\frac{1}{2}r(1-r)(1-\theta)$	n_{12}
AaBb	*AaBb* *Ab|aB* *AB|ab*	θ $\frac{1}{2}r^2(1-\theta)$ $\frac{1}{2}(1-r)^2(1-\theta)$	n_{11}
Aabb	*Ab|ab*]	$\frac{1}{2}r(1-r)(1-\theta)$	n_{10}
aaBB	*aB|aB*	$\frac{1}{4}r^2(1-\theta)$	n_{02}
aaBb	*aB|ab*	$\frac{1}{2}r(1-r)(1-\theta)$	n_{01}
aabb	*ab|ab*	$\frac{1}{2}(1-r)^2(1-\theta)$	n_{00}

$$
\begin{aligned}
\log L(\boldsymbol{\Theta}|\mathbf{M}) &= \text{constant} \\
&+ (n_{222}+n_{000})\log[g_{00}^2(1-\theta)] \\
&+ (n_{220}+n_{002})\log[g_{01}^2(1-\theta)] \\
&+ (n_{200}+n_{022})\log[g_{10}^2(1-\theta)] \\
&+ (n_{202}+n_{020})\log[g_{11}^2(1-\theta)] \\
&+ (n_{221}+n_{001})\log[g_{00}g_{01}(1-\theta)] \\
&+ (n_{212}+n_{010})\log[g_{11}g_{00}(1-\theta)] \\
&+ (n_{210}+n_{012})\log[g_{01}g_{10}(1-\theta)] \\
&+ (n_{201}+n_{021})\log[g_{11}g_{10}(1-\theta)] \\
&+ (n_{122}+n_{100})\log[g_{00}g_{10}(1-\theta)] \\
&+ (n_{120}+n_{102})\log[g_{01}g_{11}(1-\theta)] \\
&+ (n_{211}+n_{011})\log[(g_{00}g_{10}+g_{01}g_{11})(1-\theta)] \\
&+ (n_{121}+n_{101})\log[(g_{00}g_{11}+g_{10}g_{01})(1-\theta)] \\
&+ (n_{121}+n_{110})\log[(g_{00}g_{01}+g_{11}g_{10})(1-\theta)] \\
&+ n_{111}\log[\theta+\frac{1}{2}g_{00}^2(1-\theta)+\frac{1}{2}g_{01}^2(1-\theta) \\
&+ \frac{1}{2}g_{10}^2(1-\theta)+\frac{1}{2}g_{11}^2(1-\theta)] \qquad (17)
\end{aligned}
$$

The EM algorithm is implemented to estimate the parameters. In the E step, we calculate the numbers of crossers and the proportions of the progeny arising from apomixis within heterozygotes using

$$
\begin{aligned}
\phi_1 &= \frac{g_{00}g_{10}}{g_{00}g_{10}+g_{11}g_{01}} \\
&\text{for genotypes } AABbCc, aaBbCc \\
\phi_2 &= \frac{g_{00}g_{11}}{g_{00}g_{11}+g_{10}g_{01}} \\
&\text{for genotypes } AaBBCc, AabbCc \\
\phi_3 &= \frac{g_{00}g_{01}}{g_{00}g_{01}+g_{11}g_{10}} \\
&\text{for genotypes } AaBbCC, AaBbcc \\
\psi &= \frac{\theta}{\theta+\frac{1}{2}(1-\theta)(g_{00}g_{00}+g_{01}g_{01}+g_{11}g_{11}+g_{10}g_{10})} \\
&\text{for genotype } AaBbCc \qquad (18) \\
\phi_4' &= \frac{g_{00}g_{00}}{g_{00}g_{00}+g_{01}g_{01}+g_{11}g_{11}+g_{10}g_{10}} \\
&\text{for genotype } AaBbCc \\
\phi_4'' &= \frac{g_{01}g_{01}}{g_{00}g_{00}+g_{01}g_{01}+g_{11}g_{11}+g_{10}g_{10}} \\
&\text{for genotype } AaBbCc \\
\phi_4''' &= \frac{g_{11}g_{11}}{g_{00}g_{00}+g_{01}g_{01}+g_{11}g_{11}+g_{10}g_{10}} \\
&\text{for genotype } AaBbCc \\
\phi_4'''' &= \frac{g_{10}g_{10}}{g_{00}g_{00}+g_{01}g_{01}+g_{11}g_{11}+g_{10}g_{10}} \\
&\text{for genotype } AaBbCc
\end{aligned}
$$

In the M step, the parameters are estimated by

$$
\begin{aligned}
\hat{\theta} &= \frac{n_{111}\psi}{n} \\
\hat{g_{00}} &= \frac{1}{n-n_{111}\psi}[2n_{222}+n_{221}+n_{212}+\phi_1 n_{211}+n_{122} \\
&\quad +\phi_2 n_{121}+\phi_3 n_{112}+2\phi_4' n_{111}(1-\psi) \\
&\quad +\phi_3 n_{110}+\phi_2 n_{101} \\
&\quad +n_{100}+\phi_1 n_{011}+n_{010}+n_{001}+2n_{000}] \\
\hat{g_{01}} &= \frac{1}{n-n_{111}\psi}[n_{221}+2n_{220}+(1-\phi_1)n_{211} \\
&\quad +n_{210}+(1-\phi_2)n_{121} \\
&\quad +n_{120}+\phi_3 n_{112}+2\phi_4'' n_{111}(1-\psi)+\phi_3 n_{110}+n_{102} \\
&\quad +(1-\phi_2)n_{101}+n_{012}+(1-\phi_1)n_{011}+2n_{002}+n_{001}] \\
\hat{g_{10}} &= \frac{1}{n-n_{111}\psi}[\phi_1 n_{211}+n_{210}+n_{201}+2n_{200}+n_{122} \\
&\quad +(1-\phi_2)n_{121}+(1-\phi_3)n_{112} \\
&\quad +2\phi_4''' n_{111}(1-\psi)+(1-\phi_3)n_{110}+(1-\phi_2)n_{101} \\
&\quad +n_{100}+2n_{022}+n_{021}+n_{012}+\phi_1 n_{011}] \\
\hat{g_{11}} &= 1-\hat{g_{00}}-\hat{g_{01}}-\hat{g_{10}} \qquad (19)
\end{aligned}
$$

Hypothesis tests for the linkage and apomixis proportion can be made similarly in the F_2. In addition, genetic interference can be estimated and tested, providing more information to understand the cytological properties of apomictic plants.

Discussion

Apomixis is a developmentally fascinating characteristic that occurs in over 400 species of angiosperms and plays an important role in agricultural production (Koltunow and Grossniklaus 2003; Bicknell and Koltunow 2004; Spillane et al. 2004; Ozias-Akins and van Dijk 2007; d'Erfurth et al. 2009). In contrast to sexual reproduction, apomixis produces seedlings without undergoing meiosis that retain the genotype of the female parent. Because of this, traditional models founded on meiotic crossovers and recombination events will not be appropriate for linkage analysis. In this article, we describe a new statistical model for linkage analysis when progeny genotypes are yielded through apomixis.

The new model embeds apomixis into a general linkage analysis setting within the mixture model framework. Its advantage is displayed not only by the correct estimate of the linkage, but also with the estimate of apomixis incidence. Because apomixis confounds the count of the progeny genotype identical to the female parent formed by meiosis, this phenomenon obstructs the precise estimate of the recombination. Simulation studies with two- and three-point analyses have demonstrated the usefulness and

Table 9 The frequencies of four genotypes at three markers in an F_2 with progeny arising from a mix of meiotic crossovers and apomixis

Genotype	Diplotype configuration	Diplotype frequency	Relative frequency	Observation
AABBCC	*ABC|ABC*	$\frac{1}{4}g_{00}^2(1-\theta)$	1	n_{222}
AABBCc	*ABC|ABc*	$\frac{1}{2}g_{00}g_{01}(1-\theta)$	1	n_{221}
AABBcc	*ABc|ABc*	$\frac{1}{4}g_{01}^2(1-\theta)$	1	n_{220}
AABbCC	*ABC|AbC*	$\frac{1}{2}g_{00}g_{11}(1-\theta)$	1	n_{212}
AABbCc	$\begin{cases} ABC\|Abc \\ ABc\|AbC \end{cases}$	$\begin{cases} \frac{1}{2}g_{00}g_{10}(1-\theta) \\ \frac{1}{2}g_{11}g_{01}(1-\theta) \end{cases}$	$\begin{cases} \phi_1 \\ 1-\phi_1 \end{cases}$	n_{211}
AABbcc	*ABc|Abc*	$\frac{1}{2}g_{01}g_{10}(1-\theta)$	1	n_{210}
AAbbCC	*AbC|AbC*	$\frac{1}{4}g_{11}^2(1-\theta)$	1	n_{202}
AAbbCc	*AbC|Abc*	$\frac{1}{2}g_{11}g_{10}(1-\theta)$	1	n_{201}
AAbbcc	*Abc|Abc*	$\frac{1}{4}g_{10}^2(1-\theta)$	1	n_{200}
AaBBCC	*ABC|aBC*	$\frac{1}{2}g_{00}g_{10}(1-\theta)$	1	n_{122}
AaBBCc	$\begin{cases} ABC\|aBc \\ Abb\|aBC \end{cases}$	$\begin{cases} \frac{1}{2}g_{00}g_{11}(1-\theta) \\ \frac{1}{2}g_{10}g_{01}(1-\theta) \end{cases}$	$\begin{cases} \phi_2 \\ 1-\phi_2 \end{cases}$	n_{121}
AaBBcc	[*Abc|aBc*	$\frac{1}{2}g_{01}g_{11}(1-\theta)$	1	n_{120}
AaBbCC	$\begin{cases} ABC\|abC \\ AbC\|aBC \end{cases}$	$\begin{cases} \frac{1}{2}g_{00}g_{01}(1-\theta) \\ \frac{1}{2}g_{11}g_{10}(1-\theta) \end{cases}$	$\begin{cases} \phi_3 \\ 1-\phi_3 \end{cases}$	n_{112}
AaBbCc	$\begin{cases} AaBbCc \\ ABC\|abc \\ Abc\|abC \\ AbC\|aBc \\ Abb\|aBC \end{cases}$	$\begin{cases} \theta \\ \frac{1}{2}g_{00}g_{00}(1-\theta) \\ \frac{1}{2}g_{01}g_{01}(1-\theta) \\ \frac{1}{2}g_{11}g_{11}(1-\theta) \\ \frac{1}{2}g_{10}g_{10}(1-\theta) \end{cases}$	$\begin{cases} \psi \\ \phi_4'(1-\psi) \\ \phi_4''(1-\psi) \\ \phi_4'''(1-\psi) \\ \phi_4''''(1-\psi) \end{cases}$	n_{111}
AaBbcc	$\begin{cases} ABc\|abc \\ Abc\|aBc \end{cases}$	$\begin{cases} \frac{1}{2}g_{01}g_{00}(1-\theta) \\ \frac{1}{2}g_{11}g_{10}(1-\theta) \end{cases}$	$\begin{cases} \phi_3 \\ 1-\phi_3 \end{cases}$	n_{110}
AabbCC	*AbC|abC*	$\frac{1}{2}g_{11}g_{01}(1-\theta)$	1	n_{102}
AabbCc	$\begin{cases} AbC\|abc \\ Abc\|abC \end{cases}$	$\begin{cases} \frac{1}{2}g_{11}g_{00}(1-\theta) \\ \frac{1}{2}g_{10}g_{01}(1-\theta) \end{cases}$	$\begin{cases} \phi_2 \\ 1-\phi_2 \end{cases}$	n_{101}
Aabbcc	*Abc|abc*	$\frac{1}{2}g_{10}g_{00}(1-\theta)$	1	n_{100}
aaBBCC	*aBC|aBC*	$\frac{1}{4}g_{10}^2(1-\theta)$	1	n_{022}
aaBBCc	*aBC|aBc*	$\frac{1}{2}g_{10}g_{11}(1-\theta)$	1	n_{021}
aaBBcc	*aBc|aBc*	$\frac{1}{4}g_{11}^2(1-\theta)$	1	n_{020}
aaBbCC	*aBC|abC*	$\frac{1}{2}g_{10}g_{01}(1-\theta)$	1	n_{012}
aaBbCc	$\begin{cases} aBC\|abc \\ aBc\|abC \end{cases}$	$\begin{cases} \frac{1}{2}g_{10}g_{00}(1-\theta) \\ \frac{1}{2}g_{11}g_{01}(1-\theta) \end{cases}$	$\begin{cases} \phi_1 \\ 1-\phi_1 \end{cases}$	n_{011}
aaBbcc	*aBc|abc*	$\frac{1}{2}g_{11}g_{00}(1-\theta)$	1	n_{010}
aabbCC	*abC|abC*	$\frac{1}{4}g_{01}^2(1-\theta)$	1	n_{002}
aabbCc	*abC|abc*	$\frac{1}{2}g_{01}g_{00}(1-\theta)$	1	n_{001}
aabbcc	*abc|abc*	$\frac{1}{4}g_{00}^2(1-\theta)$	1	n_{000}

utilization of the new model for linkage analysis. Generally speaking, three-point analysis is more conservative in apomixis and linkage detection than two-point analysis when genetic interference exists. In practice, if the apomixis rate is estimated to be low (say 0.1 or lower), three-point analysis performs better because it can well control the FPR. When the apomixis rate is estimated to be high, two-point analysis should be used to test the apomixis significance because it is more powerful.

The model was used to analyze a practical marker data set collected in a full-sib family derived from interspecific hybridization in *Carya* spp. hybridization in hickory, leading to the detection of about 30–60% apomixis in progeny production, depending on markers. Previous studies have shown that the occurrence of apomixis may be loci-specific, i.e., apomixis may be predominant in particular chromosomal regions (Ozias-Akins et al. 1998; Spillane et al. 2001; Ozias-Akins and van Dijk 2007). To

the best of our knowledge, these findings are the first report on the estimation of apomixis in plants using genetic markers. For outcrossing *Carya* spp. different types of markers, testcross markers (i.e. those that are segregating in one parent but null in the other), and intercross markers (i.e., those that are segregating in both parents) (Lu et al. 2004) should be used simultaneously to better cover its genome. Then a more precise estimate of apomixis can be obtained by analyzing all these markers genotyped throughout the genome. It should be pointed out that, if apomixis is uniformly distributed on the genome, the apomixis rate can be approximately observed by identifying the fully heterozygous genotypes for a large number of markers. In this case, almost all of the fully heterozygous genotypes are derived from apomixis.

Our model is simply founded on the phenomenon of apomixis. It is of great importance to implement the cytological mechanisms of apomixis into the linkage analysis model. Apomixis may occur sporophytically or gametophytically (Ozias-Akins 2006). Depending on the origin of the unreduced megagametophytes, gametophytic apomixis is formed through two different mechanisms, namely diplosporous and aposporous. Diplospory stems from meiotic failure in megasporocytes that develop into unreduced female gametophytes. In aposporous apomicts, one or more unreduced female gametophytes result mitotically from somatic nucellar cells, while the legitimate sexual line generally aborts (Grimanelli et al. 1998). By collecting more embryological data for the formation of apomixis, these different mechanisms can be integrated to better study the developmental and reproductive machinery of flowering plants (Grimanelli et al. 2001; Koltunow and Grossniklaus 2003). Finally, as a characteristic, apomixis is thought to have a genetic component (Grimanelli et al. 1998, 2001; Richards 2003). Specific genes, or quantitative trait loci (QTLs), that control apomixis can be identified by developing a new mapping methodology that integrates apomixis.

Acknowledgments This work is partially supported by an Open Fund 200602 from former Zhejiang Forestry University, a grant 30771752 from the NSF of China, NSF/IOS-0923975, Changjiang Scholars Award, and "Thousand-person Plan" Award.

References

Bicknell RA, Koltunow AM (2004) Understanding apomixis: recent advances and remaining conundrums. Plant Cell 16:S228–S245

d'Erfurth I, Jolivet S, Froger N, Catrice O, Novatchkova M et al (2009) Turning meiosis into mitosis. PLoS Biol 7(6):e1000124

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

Grimanelli D, Leblanc O, Espinosa E, Perotti E, Gonzlez de Leon D, Savidan Y (1998) Mapping diplosporous apomixis in tetraploid Tripsacum: one gene or several genes? Heredity 80:33–39

Grimanelli D, Leblanc O, Perotti E, Grossniklaus U (2001) Developmental genetics of gametophytic apomixis. Trends Genet 17:597–604

Hanna WW (1995) Use of apomixis in cultivar development. Adv Agron 54:333–354

Koltunow AM (1993) Apomixis: embryo sacs and embryos formed without meiosis or fertilization in ovules. Plant Cell 5:1425–1437

Koltunow AM, Grossniklaus U (2003) Apomixis: a developmental perspective. Annu Rev Plant Biol 54:547–574

Lathrop GM, Lalouel JM, Julier C, Ott J (1984) Strategies for multilocus linkage analysis in humans. Proc Natl Acad Sci USA 81:3443–3446

Lu Q, Cui YH, Wu RL (2004) A multilocus likelihood approach to joint modeling of linkage, parental diplotype and gene order in a full-sib family. BMC Genetics 5:20

Ozias-Akins P (2006) Apomixis: developmental characteristics and genetics. Crit Rev Plant Sci 25:199–214

Ozias-Akins P, van Dijk PJ (2007) Mendelian genetics of apomixis in plants. Annu Rev Genet 41:509–537

Ozias-Akins P, Roche D, Hanna WW (1998) Tight clustering and hemizygosity of apomixis-linked molecular markers in Pennisetum squamulatum implies genetic control of apospory by a divergent locus that may have no allelic form in sexual genotypes. Proc Natl Acad Sci USA 95:5127–32

Richards AJ (2003) Apomixis in flowering plants: an overview. Philos Trans R Soc Lond B Biol Sci 358:1085–1093

Spielman M, Vinkenoog R, Scott RJ (2003) Genetic mechanisms of apomixis. Philos Trans R Soc Lond B Biol Sci 358:1095–1103

Spillane C, Steimer A, Grossniklaus U (2001) Apomixis in agriculture: the quest for clonal seeds. Sex Plant Reprod 14:179–187

Spillane C, Curtis MD, Grossniklaus U (2004) Apomixis technology development-virgin births in farmers' fields? Nat Biotechnol 22:687–691

Wu RL, Ma CX, Casella G (2007) Statistical genetics of quantitative traits: linkage, maps, and QTL. Springer, New York