Theor Appl Genet (1999) 99: 140–146

H. Kuang · T. E. Richardson
S. D. Carson · B. C. Bongarten

Genetic analysis of inbreeding depression in plus tree 850.55 of *Pinus radiata* D. Don. II. Genetics of viability genes

Received: 13 December 1998 / Accepted: 28 December 1998

Abstract Most coniferous species exhibit severe inbreeding depression. Selfed individuals usually have decreased viability, reduced vigour and morphological defects. The number of filled seeds after selfing *Pinus radiata* plus tree 850.55 was 48% that of the outcrossing, and 26.1% of the selfed seedlings died at an early stage. The segregation of 172 markers (covering 56% of the genome) in selfed progenies of radiata pine plus tree 850.55 was studied. Based on the segregation ratio of the markers, genes associated with inbreeding depression on viability were identified ($P < 0.05$). Using the Expectation/Conditional Maximization (ECM) algorithm, we estimated the location, degree of dominance and selection coefficient of viability genes. Nine viability genes were discovered. Seven of them appeared to be dominant and one partially dominant (degree of dominance = 0.4). The other gene was overdominant or pseudo-overdominant, with selection coefficients for the two homozygotes of 0.4 and 0.42, respectively. Of the genes showing dominance or partial dominance, seven were sub-lethal with selection coefficients ranging from 0.55 to 0.79; one gene (*SDPr*), which was responsible for seedling death within the first month following germination, was lethal.

Key words Segregation distortion · Inbreeding depression · Viability gene · Lethal allele · *Pinus radiata*

Communicated by P. M. A. Tigerstedt

H. Kuang · B. C. Bongarten (✉)
Daniel B. Warnell School of Forest Resources, University of Georgia, Athens, Georgia, 30602-2152, USA
Fax: 706-542-8356
E-mail: bongarte@smokey.forestry.uga.edu

T. E. Richardson · S. D. Carson
New Zealand Forest Research Institute Ltd, Private Bag 3020, Rotorua, New Zealand

Introduction

Most coniferous species, like other outcrossing species, exhibit inbreeding depression (reviewed by Williams and Savolainen 1996). Inbreeding depression may be manifested in many ways. Upon selfing, tree species may have reduced seed yield, seeds may have **a** lower germination rate and seedlings may have a higher percentage of abnormality, lower survival rate and reduced growth. Reduced yield of filled seeds is one of the most consistent and severe results of inbreeding in conifers (Bingham and Squillace 1955; Snyder 1968; Sorensen 1971; Sorensen and Miles 1974; Fowler and Park 1983; Griffin and Lindgren 1985).

It is commonly believed that homozygous lethal or sublethal alleles are responsible for producing empty seeds upon selfing (Orr-Ewing 1957; Sarvas 1962; Hagman and Mikkola 1963; Mergen et al. 1965), but direct evidence is rare (Fowler 1964, 1965; Hedrick and Muona 1990). Based on the magnitude of inbreeding depression on viability (seed production, in most cases), the number of lethal equivalents has been estimated for many coniferous species (Williams and Savolainen 1996). However, how many genes are involved is unknown. The number of lethal equivalents is a close estimate of the number of genes involved only if the recessive homozygotes are completely lethal. Otherwise, many more genes may be operating.

By means of molecular markers, both fully lethal and sublethal alleles can be identified, and their genetic properties can be studied. Following natural selection, the detrimental genotypes in a family should be present at lower frequencies than those expected from Mendelian ratios. Tightly linked molecular markers should also show distorted Mendelian ratios. Segregation distortion of markers provides information on the location and genetic effects of the linked deleterious allele. Hedrick and Muona (1990) used the maximum likelihood method on segregation data of one isozyme

marker (fluorescent esterasase) in a selfed family to identify a deleterious allele in Scots pine. More sophisticated methods which utilize segregation data of two flanking codominant markers have been developed recently. Cheng et al. (1996) used the Expectation/Conditional Maximization (ECM) method to study viability genes in rice.

Radiata pine (*Pinus radiata*) has become one of the most economically important tree species in the world. Plus tree 850.55 has been shown to be the best of all *P. radiata* selections in terms of growth rate. Unfortunately, radiata pine, like most conifers, exhibits significant inbreeding depression (Wilcox 1983; Griffin and Lindgren 1985), which is expected to be manifested as recurrent selection continues. The number of filled seeds per cone after selfing plus tree 850.55 was 48% that of the outcrossing; the number of filled seeds in S_2 was 48% that of S_1 (Griffin and Lindgren 1985).

In the study presented here, the deleterious alleles in the genome of *P. radiata* plus tree 850.55 were studied based on the segregation data of random amplified polymorphic DNA (RAPD) and microsatellite markers. ECM was performed to estimate the location, degree of dominance and the selection coefficients of the corresponding viability genes.

Materials and methods

Plant materials, DNA isolation, PCR (polymerase chain reaction)

Three hundred and seventy-eight selfed seeds of plus tree 850.55 were germinated singly in containers. Eighty-seven of the sown seeds failed to germinate, leaving a total of 291 seeds that germinated successfully. Haploid megagametophytes from germinated seeds were collected and stored at -20°C until DNA extraction, but megagametophytes of seeds that failed to germinate were not collected. Of the 291 seedlings 76 died within 1 month after germination. Needles of the seedlings that died were not collected. The remaining 215 seedlings survived for more than 1 year. One hundred and ninety-eight of them were chosen for this study based on the quality of DNA from megagametophytes.

Using a Fast Prep FP 120 machine (Savant) with Bio 101 Kit H according to the manufacturer's instructions, we isolated genomic DNA from megagametophytes and needles of the 215 surviving individuals and from megagametophytes of 76 dead individuals. PCR was performed as previously described (Kuang et al. 1998).

Genotyping, linkage analysis, segregation distortion and viability genes

The DNA from 198 megagametophytes of the surviving seedlings had been screened with 202 markers. A linkage map with 19 linkage groups and 172 markers had been constructed previously (Kuang et al. 1999). Segregation distortion, as measured by departure from the expected 1:1 ratio for presence or absence of a band, was tested by chi-square analysis for each marker.

DNA from needles of surviving seedlings were genotyped with 54 markers, representing all previously identified distorted regions and covering well the whole genome. If the megagametophyte produced a band for a RAPD marker, no further information about the diploid genotype could be obtained. If the megagametophyte did not produce a band, the diploid genotype could be determined. For a dominant marker *A*, the diploid genotype was recorded as *A*- if megagametophyte had the dominant allele (band allele *A*); as *aA* if its megagametophyte did not produce a band (null allele *a*) and the diploid produced a band; and as *aa* if neither the megagametophyte nor the diploid produced a band. Segregation distortion, as measured by departure from the expected 2:1:1 ratio for genotype *A*-:*aA*:*aa*, was tested by chi-square analysis.

A viability gene was believed to be present and therefore further investigated if: (1) at least 2 continuous markers in a linkage group were distorted, and at least 1 of them was distorted with high statistical significance ($P < 0.01$); (2) these markers did not detect more double recombinants than expected; and (3) at least 1 of the distorted markers was a good marker (Kuang et al. 1998).

Expectation/Conditional Maximization (ECM)

Each distorted region was assumed to be controlled by a single viability gene. The Expectation/Conditional Maximization (ECM) algorithm was used to estimate the position and genetic parameters of the viability genes.

The approach is analogous to that used by Cheng et al. (1996), but the population differs and a dominance factor is included. Assuming that the linkage order of a viability gene *L* and two co-dominant markers *A* and *B* is *A* r_1 *B* r_2 *L* (r_1 and r_2 are recombination frequencies), that *ABL*/*abl* is in coupling phase, that there is no selection and that there is no chiasma interference, the expected frequencies F_{ij} ($i = 1 \ldots 9$; j = 1, 2, 3) of the 27 genotype classes are listed in Table 1. If the viability of genotypes *LL*, *Ll*, *ll* is 1, $1 - hs$, $1 - s$ respectively, the expected genotype frequencies, then, are: $Z_{ij} = F_{ij} \times G_j$. For genotypes with *LL*, j = 1, and $G_1 = 4/(4 - 2hs - s)$; for genotypes with *Ll*, j = 2, and $G_2 = 4(1 - hs)/(4 - 2hs - s)$; for genotype with *ll*, j = 3, and $G_3 = 4(1 - s)/(4 - 2hs - s)$.

Notice that the genotype of the viability gene L can not be observed directly. Only the nine marker genotype classes (a_i, i = 1 … 9) can be scored. However, counts of the nine marker genotypes can be used to calculate the expected counts (a_{ij}) for each of the 27 genotypes that include the *L* locus as follows:

$$a_{ij} = a_i \times \frac{F_{ij} \times G_j}{\sum_{k=1}^{3} [F_{ik} \times G_k]} \tag{1}$$

Then the likelihood can be expressed as:

$$e^L \propto \prod_{i=1}^{9} \prod_{j=1}^{3} [Z_{if}]^{a_{if}} = \prod_{i=1}^{9} \prod_{j=1}^{3} [F_{ij}(r_1, r_2) \times G_j(h, s)]^{a_{ij}}$$

and log-likelihood as:

$$L \propto \sum_{i=1}^{9} \sum_{j=1}^{3} [a_{ij} \log F_{ij} + a_{ij} \log G_j]$$

The maximum-likelihood equations for scores are:

$$S_{r_1} = \frac{dL}{dr_1} = \sum_{i=1}^{9} \sum_{j=1}^{3} \left[\frac{a_{ij}}{F_{ij}} \times \frac{dF_{if}}{dr_1} \right] = 0 \tag{2}$$

$$S_{r_2} = \frac{dL}{dr_2} = \sum_{i=1}^{9} \sum_{j=1}^{3} \left[\frac{a_{ij}}{F_{ij}} \times \frac{dF_{if}}{dr_2} \right] = 0 \tag{3}$$

$$S_h = \frac{dL}{dh} = \sum_{j=1}^{3} \left[\frac{\sum_{i=1}^{9} a_{ij}}{G_j} \times \frac{dG_j}{dh} \right] = 0 \tag{4}$$

$$S_s = \frac{dL}{ds} = \sum_{j=1}^{3} \left[\frac{\sum_{i=1}^{9} a_{ij}}{G_j} \times \frac{dG_j}{ds} \right] = 0 \tag{5}$$

Table 1 Expected genotype frequencies F_{ij} in S_1 population if linkage order $A\ r_1 B\ r_2\ L$[a, b]

Genotype	LL (F_{i1})[c]	Ll (F_{i2})	ll (F_{i3})
AABB	$(1-r_1)^2\ (1-r_2)^2$	$2(1-r_1)^2 r_2(1-r_2)$	$(1-r_1)^2 r_2^2$
AABb	$2r_1(1-r_1)r_2(1-r_2)$	$2r_1(1-r_1)[(1-r_2)^2+r_2^2]$	$2r_1(1-r_1)r_2(1-r_2)$
AAbb	$r_1^2 r_2^2$	$2r_1^2 r_2(1-r_2)$	$r_1^2(1-r_2)^2$
AaBB	$2r_1(1-r_1)(1-r_2)^2$	$4r_1(1-r_1)r_2(1-r_2)$	$2r_1(1-r_1)r_2^2$
AaBb	$2[r_1^2+(1-r_1)^2]r_2(1-r_2)$	$2[r_1^2+(1-r_1)^2][(1-r_2)^2+r_2^2$	$2[r_1^2+(1-r_1)]r_2(1-r_2)$
Aabb	$2r_1(1-r_1)r_2^2$	$4r_1(1-r_1)r_2(1-r_2)$	$2r_1(1-r_1)(1-r_2)^2$
aaBB	$r_1^2(1-r_2)^2$	$2r_1^2 r_2(1-r_2)^2$	$r_1^2 r_2^2$
aaBb	$2r_1(1-r_1)r_2(1-r_2)$	$2r_1(1-r_1)[(1-r_2)^2+r_2^2]$	$2r_1(1-r_1)r_2(1-r_2)$
aabb	$(1-r_1)^2 r_2^2$	$2(1-r_1)^2 r_2(1-r_2)$	$(1-r_1)^2(1-r_2)^2$

[a] r_1 and r_2 are recombination frequencies
[b] Marker A and B are codominant markers
[c] 1/4 is omitted for all frequencies

Because the data in a_{ij} are not directly observed, the Expectation Maximization (EM) algorithm has to be applied. The M-step of conventional EM algorithm is very complicated here, therefore, Expectation/Conditional Maximization (ECM) is used instead (Meng and Rubin 1993). The ECM algorithm consists of one E (Expectation) step and two CM (Conditional Maximization) steps. We start with r_1^0, r_2^0, h^0, s^0 which are initial guesses of r_1, r_2, h, and s, respectively. In the first E-step of the ECM algorithm, a_{ij}^1 is calculated by Eq. 1 using the observed a_i, and r_1^0, r_2^0, h^0, and s^0. In the first CM-step of the algorithm: (1) r_1^1 is estimated by solving Eq. (2) using a_{ij}^1 and r_2^0; and h^1 is estimated by solving Eq. 4 using a_{ij}^1 and s^0, (2) then, r_2^1 is estimated by solving Eq. 3 using a_{ij}^1 and r_1^1, and s^1 is estimated by solving Eq. 5 using a_{ij}^1 and h^1. In the (k + 1)st E-step of the ECM algorithm, a_{ij}^{k+1} is calculated by Eq. 1 using the observed a_i, and r_1^k, r_2^k, h^k, and s^k. In the (k + 1)st CM-step of the algorithm: (1) r_1^{k+1} is estimated by solving Eq. 2 using a_{ij}^{k+1} and r_2^k, and h^{k+1} is estimated by solving Eq. 4 using a_{ij}^{k+1} and s^k. (2) r_2^{k+1} is estimated by solving Eq. 3 using a_{ij}^{k+1} and r_1^{k+1}, and s^{k+1} is estimated by solving Eq. 5 using a_{ij}^{k+1} and h^{k+1}. The ECM is repeated until the difference between the estimates in the current cycle and those in the previous cycle is smaller than a pre-determined threshold (0.001).

Variances of the final estimates are determined using Fisher's information matrix I. Let $\theta_1 = r_1$, $\theta_2 = r_2$, $\theta_3 = h$, $\theta_4 = s$, then Fishes's information matrix I is:

$$I = \begin{bmatrix} I_{11} & I_{12} & 0 & 0 \\ I_{21} & I_{22} & 0 & 0 \\ 0 & 0 & I_{33} & I_{34} \\ 0 & 0 & I_{43} & I_{44} \end{bmatrix}$$

According to Bailey (1961) and Cheng et al. (1996):

$$I_{11} = I_{r_1} = -E\frac{d^2L}{dr_1^2} = n\sum_{i=1}^{9}\sum_{j=1}^{3}\left[\frac{G_i}{F_{ij}}\times\left(\frac{dF_{ij}}{dr_1}\right)^2\right]$$

$$I_{22} = I_{r_1} = -E\frac{d^2L}{dr_2^2} = n\sum_{i=1}^{9}\sum_{j=1}^{3}\left[\frac{G_i}{F_{ij}}\times\left(\frac{dF_{ij}}{dr_2}\right)^2\right]$$

$$I_{33} = I_h = -E\frac{d^2L}{dh^2} = n\sum_{j=1}^{3}\left[\frac{\sum_{i=1}^{9}F_{ij}}{G_j}\times\left(\frac{dG_j}{dh}\right)^2\right]$$

$$I_{44} = I_s = -E\frac{d^2L}{ds^2} = n\sum_{j=1}^{3}\left[\frac{\sum_{i=1}^{9}F_{ij}}{G_j}\times\left(\frac{dG_j}{ds}\right)^2\right]$$

$$I_{12} = I_{21} = -E\frac{d^2L}{dr_1dr_2} = n\sum_{i=1}^{9}\sum_{j=1}^{3}\left[\frac{G_i}{F_{ij}}\left(\frac{dF_{ij}}{dr_1}\right)\left(\frac{dF_{ij}}{dr_2}\right)\right]$$

$$I_{34} = I_{43} = -E\frac{d^2L}{dhds} = n\sum_{i=1}^{9}\sum_{j=1}^{3}\left[\frac{F_{ij}}{G_j}\times\left(\frac{dG_j}{dh}\right)\left(\frac{dG_j}{ds}\right)\right]$$

The variance of the four estimates can be obtained from inversion of the information matrix:

$$V_{(r_1)} = I^{11} = \frac{I_{22}}{I_{11}I_{22} - I_{12}^2}$$

$$V_{(r_2)} = I^{22} = \frac{I_{11}}{I_{11}I_{22} - I_{12}^2}$$

$$V_{(h)} = I^{33} = \frac{I_{44}}{I_{33}I_{44} - I_{34}^2}$$

$$V_{(s)} = I^{44} = \frac{I_{33}}{I_{33}I_{44} - I_{34}^2}$$

The viability gene can also be on the other end of 2 markers or between the 2 markers. All three possibilities are considered for each viability gene. The estimation procedure is the same but F_{ij} changed. When the viability gene is located between the 2 markers, the expected F_{ij} is shown in Table 1 of Cheng et al. (1996). The chi-square

$$x^2 = \sum_{i=1}^{9}\frac{(a_i - n\sum_{j=1}^{3}F_{ij}G_j)^2}{n\sum_{j=1}^{3}F_{ij}G_j}$$

is calculated for all three possibilities. The linkage order with the smallest x^2 is accepted. A linkage order is rejected if it leads to a negative estimate of recombination frequency.

The expected frequencies, F_{ij}, also vary with marker type (dominant or codominant marker) and linkage phase (coupling or repulsion). The above method can be applied in all cases as long as the corresponding F_{ij} is used. When both megagametophytes and needles are genotyped with dominant markers, nine genotype classes can be distinguished. Table 2 shows the expected genotype frequencies of 2 dominant markers.

Table 2 Expected genotype frequencies F_{ij} in S_1 population if linkage order is $A\ r_1\ Lr_2B$[a, b]

Genotype	LL (F_{i1})[c]	Ll (F_{i2})	ll (F_{i3})
A-b-	$(1-r_1)r_2$	$r_1+r_2-2r_1r_2$	$r_1(1-r_2)$
A-Bb	$(1-r_1)r_2(1-r_2)$	$(1-r_1)(1-r_2)^2+r_1r_2^2$	$r_1r_2(1-r_2)$
A-BB	$(1-r_1)(1-r_2)^2$	$r_2(1-r_2)$	$r_1r_2^2$
aAb-	$r_1(1-r_1)r_2$	$(1-r_1)^2(1-r_2)+r_1^2r_2$	$(r_1(1-r_1)(1-r_2)$
aab-	$r_1^2r_2$	$r_1(1-r_1)$	$(1-r_1)^2(1-r_2)$
aABb	$r_1(1-r_1)r_2(1-r_2)$	$r_1^2(1-r_2)^2+r_2^2(1-r_1)^2$	$r_1(1-r_1)r_2(1-r_2)$
aABB	$r_1(1-r_1)(1-r_2)^2$	$[r_1^2+(1-r_1)^2]r_2(1-r_2)$	$r_1(1-r_1)r_2^2$
aaBb	$r_1^2r_2(1-r_2)$	$r_1(1-r_1)[r_2^2+(1-r_2)^2$	$(1-r_1)^2r_2(1-r_2)$
aaBB	$r_1^2(1-r_2)^2$	$2r_1(1-r_1)r_2(1-r_2)$	$(1-r_1)^2r_2^2$

[a] r_1 and r_2 are recombination frequencies
[b] Allele A and b are dominant to allele a and B, respectively
[c] 1/4 is omitted for all frequencies

When there is an excess of heterozygotes, overdominance is considered. The above method can also be used for overdominant genes. Assuming that alleles A and b are dominant (i.e. producing a band) to alleles a and B, respectively, and that the viability gene L is between the 2 markers, the genotypes and their expected frequencies F_{ij} are identical to those in Table 2. When the viability of LL, Ll and ll is $1-s1$, 1, $1-s2$ respectively, G_j ($j=1,2,3$), in this case, becomes

$$\frac{4(1-s_1)}{4-s_1-s_2}, \frac{4}{4-s_1-s_2}, \frac{4(1-s_2)}{4-s_1-s_2}$$

respectively.

A single viability gene does not change the estimate of the recombination frequency when megagametophytes of selfed seeds are used (Kuang et al. 1999). Therefore, the LOD for a viability gene is:

$$LOD=\sum_{i=1}^{9}\left[a_i\times\log_{10}\left(\sum_{j=1}^{3}F_{ij}G_i\right)\right]-\sum_{i=1}^{9}\left[a_i\times\log_{10}\left(\sum_{j=1}^{3}F_{ij}\right)\right].$$

And $4.6\times$ LOD is asymptotically distributed as x^2 with 4 df (Wilks 1938).

Results

Of the 172 markers, 59 (34%) were distorted ($P<0.05$) from the expected 1:1 ratio in megagametophytes of the S_1 progeny (Kuang et al. 1999). Markers *AD02500A* and *u507900a*, which had normal segregation ratios in megagametophytes of S_1 seeds, showed highly distorted segregation in S_1 diploids. Nine distorted regions, each controlled by a single viability gene (see Methods and materials section), were discovered. Together, they accounted for 48 distorted markers. The 2 best markers in each distorted region were chosen to screen diploid genotypes. All of these markers were significantly distorted from Mendelian ratios in diploid genotypes.

Different models (i.e. different F_{ij}) were used to estimate the properties of the deleterious genes, depending on marker type, linkage phase and linkage order (e.g. Tables 1 and 2). The estimates of recombination frequencies, degree of dominance, selection coefficient of viability gene and their standard errors are shown in Table 3.

Two viability genes were linked with microsatellites. Viability gene 1 (VG_1) was linked with *Mic9.3* and a co-dominant RAPD marker *B13*, the two alleles of which are *B13750A* and *B13800A*. The deleterious allele of VG_1 was in coupling phase with allele 9 of *Mic9.3* and allele *B13800A* of *B13*. The recombination frequency between VG_1 and *Mic9.3* was 0.096. VG_6 was linked with *Mic5.1B* by a recombination frequency of 0.084. Its deleterious allele was in coupling phase with allele X (81 bp) of *Mic5.1B*. All other viability genes were linked only with RAPD markers. Recall, by our definition, the deleterious alleles were in repulsion phase with alleles A and B. In Table 3, allele A is always the band allele of marker A.

Most of the viability genes appear to be completely recessive even though the estimates of degree of dominance varied from -0.33 to 0.41. Degrees of dominance with negative values may indicate overdominance, and positive values suggest partial dominance. However, the SEs of these estimates were very large, and the estimates did not significantly depart from 0, except in the case of gene VG_7. Based on the marker segregation data, there was no evidence of overdominance or partial dominance for genes VG_1-VG_6 and *SDPr*. VG_7 might be partially dominant. The degree of dominance and SE for VG_7 were 0.41 and 0.12, respectively.

Viability genes VG_1-VG_7 were sub-lethal. The selection coefficient of their recessive homozygotes ranged from 0.55 to 0.79. Homozygotes of *SDPr*, an allele responsible for seedling death shortly after germination (Kuang et al. 1998), were lethal, with an estimated selection coefficient 0.99 ± 0.01.

Segregation of markers *AD02500A* and *u507900a* was normal in megagametophytes. However, the diploid genotypes at these loci were highly distorted, with an excess of heterozygotes. Analysis using the overdominance model showed that a viability gene, VG_8,

Table 3 Location and genetic parameters of viability genes

Gene name	Linked markers	Linkage order	MEL[a]				
			h *SE*	*s* *SE*	r_1 *SE*	r_2 *SE*	LOD β
VG_1	A: *Mic93* B: *B13*	*A* r_1L r_2 B	− 0.1 0.24	0.71 0.08	0.1 0.02	0.14 0.02	4.98 < 0.005
VG_2	A: *AL19600A* B: *D12440B*	A r_1 L r_2 B	− 0.3 0.34	0.63 0.1	0.09 0.02	0.07 0.02	4.24 < 0.005
VG_3	A: *U427350A* B: *U564850A*	A r_1 L r_2 B	− 0.33 0.4	0.55 0.12	0.02 0	0.02 0	4.17 < 0.005
VG_4	A: *AC19700A* B: *ar01500a*	*L* r_1 A r_2 B	− 0.16 0.24	0.57 0.1	0.02 0	0.13 0.02	3.25 < 0.025
	A: *AC19700A* B: *u370600b*	A r_1 L r_2 B	− 0.23 0.3	0.67 0.09	0.149 0.02	0.01 0	3.86 < 0.005
VG_5	A: *AB16800A* B: *u564600a*	*L* r_1 A r_2 B	0.3 0.22	0.57 0.1	0.09 0.02	0.04 0.01	2.13 < 0.05
VG_6	A: *Mic5-1B* B: *AB19510A*	A r_1 L r_2 B	0.05 0.24	0.64 0.09	0.08 0.02	0.03 0.01	4.58 < 0.005
VG_7	A: *AH04900A* B: *u427490b*	*L* r_1 A r_2 B	0.41 0.12	0.79 0.06	0.1 0.02	0.1 0.02	5.4 < 0.001
			S_1^b	S_2			
VG_8	A: *AD02500A* B: *U507900*	*L* r_1 A r_2 B	0.4 0.11	0.42 0.11	0 0	0.03 0.01	2.47 < 0.05
SDPr	A: *AI05800A* b: *u327600b*	*L* r_1 A r_2 B	− 0.15 0.18	0.99 0.01	0.01 0	0.1 0.02	20.11 < 0.001

[a] *h*, Degree of dominance; *s*, selection coefficient; *SE*, standard error
[b] S_1 and S_2 are selection coefficients against homozygote *LL* and *ll*, respectively

was located between these 2 markers. The selection coefficients of two homozygotes were 0.4 and 0.42, respectively.

The LOD score for testing viability genes was mainly determined by the selection coefficient and map distance between the viability gene and the markers. Gene *SDPr* had the highest LOD = 20.11 ($P < 0.0001$). LOD scores for other viability genes were at least 2.13 ($P < 0.05$).

Discussion

Nine deleterious alleles were discovered in the genome of radiata pine plus tree 850.55. The location and genetic parameters of the viability genes were studied using an ECM algorithm. Eight of the deleterious genes appeared to be dominant or partially dominant and the other was overdominant.

Genetics of deleterious genes

Dominance and overdominance are two genetic models that have been used to explain inbreeding depression, with more evidence in support of the former (Charlesworth and Charlesworth 1987). It is commonly believed that homozygous recessive lethal or sub-lethal alleles cause the embryonic development failure which is responsible for producing an empty seed. In this study, eight out of nine viability genes were dominant. Of them, seven were close to being completely dominant. One gene, VG_7, appeared to be partially dominant. VG_8 appeared to be overdominant. However, our data could not distinguish overdominance from pseudo-overdominance (Stuber et al. 1992). Therefore, the data can be alternatively interpreted as that two deleterious genes with selection coefficients of 0.40 and 0.42, respectively, were highly linked in repulsion phase.

The viability genes were detected using the segregation of markers in surviving seedlings. They could be expressed anytime from fertilization to the seedling development stage. *SDPr*, the only completely lethal allele detected, was expressed soon after seed germination only (Kuang et al. 1998). Expression time of the other genes could be determined if genotype frequencies were obtained in early embryos or mature seeds. Nevertheless, the genes associated with inbreeding depression during seed production in plus tree 850.55 were sublethal.

Upon selfing, plus tree 850.55 produced 48% as many filled seeds as produced with outcrossing (Griffin and Lindgren 1985). Unfortunately, there was no record about seed abortion in the process of producing the materials used in this experiment. Of the 378 selfed seeds used in this study, 87 failed to germinate, while 76

seedlings died soon after germination. The total percentage of dead individuals, including aborted seeds [assumed to be 52% (Griffin and Lindgren 1985)], ungerminated seeds and dead seedlings, was 72.7%. Based on the estimated degrees of dominance and selection coefficients of the nine viability genes, 81.4% of the individuals should have died from the time of fertilization through the first year of growth. This percentage (81.4%) is close to the observed percentage (72.7%). Several factors may account for the difference between the two. First, the seed production of plus tree 850.55 upon selfing might have been different during the two seasons. Secondly, the estimates of the selection coefficients and other parameters may not be accurate due to sampling error. Thirdly, gametic selection could also be responsible. If gametic selection occurs, the losses predicted by parameters estimated by segregation distortion will be overestimated because gametic selection creates segregation distortion without seed or seedling death. For the gene that appeared to be partially dominant, gametic selection is an alternative explanation for the observed genotype ratios. However, for the seven viability genes that are almost completely recessive, gametic selection is not indicated. Gametic selection disturbs the segregation as $LL = Ll > lL = ll$; zygotic selection changes the ratio to $LL = Ll = lL > ll$; a combination of gametic and zygotic selection leads to $LL = Ll > lL > ll$. In this study, only zygotic selection was suggested.

Detection of viability genes

The power to detect viability genes is determined by the selection coefficient and location of the gene, marker type, linkage phase and sample size. Suppose that a deleterious gene is completely recessive and that the viability of the recessive homozygote is 0.4; then, the power to detect it by 2 markers with a recombination rate of r = 0.2 in a selfed family with $n = 200$ is shown in Fig. 1. The LOD score is calculated as:

$$LOD = \sum_{i=1}^{9}\left[a_i \times \log_{10}\left(\sum_{j=1}^{3} F_{ij}G_i\right)\right] - \sum_{i=1}^{9}\left[a_i \times \log_{10}\left(\sum_{j=1}^{3} F_{ij}\right)\right]$$

where a_i is the expected number of each genotype. Codominant markers (only diploid genotype available) are better than dominant markers (with both megagametophyte and diploid genotypes) in detecting viability genes. If the dominant alleles of markers are in coupling phase with the deleterious allele, the chance to detect the deleterious gene is the least (curve "a, b domin"). However, if the recessive alleles of markers are in coupling phase with the deleterious allele, the

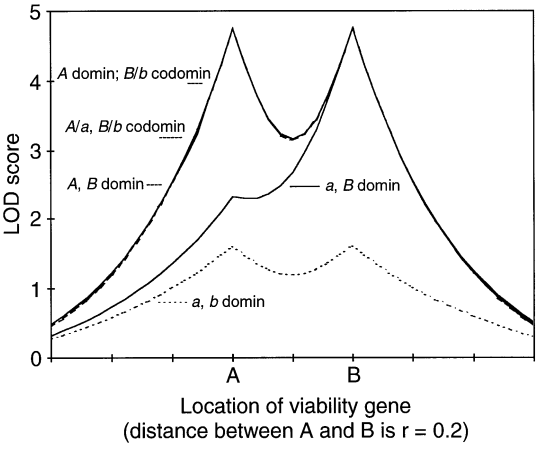

Fig. 1 The power to detect viability gene by different types of markers. Alleles *a* and *b* are in coupling phase with the deleterious allele of viability gene (*domin*–Dominant, *codomin*-codominant)

LOD score by dominant markers ("A, B domin") is almost the same as that by codominant markers ("A/a, B/b codomin"). At least 1 such type of marker was chosen in each case of this study. For non-conifer species, codominant markers must be used to keep enough degrees of freedom.

Purging lethal and sublethal alleles

Sub-lethal genes in trees have been suggested by many experiments, even though direct evidence has been rare (Fowler 1964; Hedrick and Muona 1990). That the second generation of selfing yields a lower percentage of filled seed than the first-generation of selfing (Andersson et al. 1974; Griffin and Lindgren 1985) might suggest that sub-lethal genes are involved. In this study, eight of the nine deleterious alleles detected were sublethal. Their selection coefficients ranged from 0.55 to 0.79.

Recurrent inbreeding is generally not a suitable method for purging deleterious genes from trees, because trees usually have large numbers of lethal equivalents, long juvenile stages and sub-lethal genes are involved; the breeding population may become extinct after a few generations of inbreeding when the number of lethal equivalents is very large. Sub-lethal alleles may be fixed in the population when recurrent inbreeding is practiced (Hedrick 1994). However, sub-lethal alleles can be easily purged by marker-assisted selection (MAS) after association of markers and deleterious alleles is detected. With fewer such deleterious genes in their genomes, trees can be inbred without heavy depression.

Declaration This experiment complies with the current laws of USA and New Zealand, where it was performed.

References

Andersson E, Jansson R, Lindgren D (1974) Some results from second generation crossings involving inbreeding in Norway spruce (*Picea abies*). Silvae Genet 23:34–43

Bailey NTJ (1961) Mathematical theory of genetic linkage. Oxford University Press, Amen House, London

Bingham RT, Squillace AE (1955) Self-compatibility and effects of self-fertility in western white pine. For Sci 1:121–129

Charlesworth D, Charlesworth B (1987) Inbreeding depression and its evolutionary consequences. Annu Rev Ecol Syst 18:237–268

Cheng R, Saito A, Takano Y, Ukai Y (1996) Estimation of the position and effect of a lethal factor locus on a molecular marker linkage map. Theor Appl Genet 93:494–502

Fowler DP (1964) Pre-germination selection against a deleterious mutant in red pine. For Sci 10:335–336

Fowler DP (1965) Natural self-fertilization in three jack pines and its implications in seed orchard management. For Sci 11:55–58

Fowler DP, Park TW (1983) Population studies of white spruce. I. Effects of self-pollination. Can J For Res 13:1133–1138

Griffin AR, Lindgren D (1985) Effect of inbreeding on production of filled seed in *Pinus radiata*-experimental results and a model of gene action. Theor Appl Genet 71:334–343

Hagman M, Mikkola L (1963) Observations on cross-, self- and interspecific pollinations in *Pinus peuce* Griseb. Silvae Genet 12:73–79

Hedrick PW (1994) Purging inbreeding depression and the probability of extinction: full-sib mating. Heredity 73:63–372

Hedrick PW, Muona O (1990) Linkage of viability genes to marker loci in selfing organisms. Heredity 64:67–72

Kuang H, Richardson TE, Carson SD, Bongarten BC (1998) An allele responsible for seeding death in *Pinus radiata* D. Don. Theor Appl Genet 96:640–644

Kuang H, Richardson TE, Carson SD, Wilcox P, Bongarten BC (1999) Genetic analysis of inbreeding depression in plus tree 850.55 of *Pinus radiata* D. Don. I. Linkage map with distorted markers. Theor Appl Genet 95:697–703

Meng XL, Rubin DB (1993) Maximum likelihood estimation via the ECM algorithm: general framework. Biometrika 80:267–278

Mergen F, Burley J, Furnival GM (1965) Embryo and seedling development in *Picea glauca* (Moench) Voss after self-, cross-, and wind- pollination. Silvae Genet 14:188–194

Orr-Ewing AL (1957) A cytological study of the effects of self-pollination on *Pseudosuga menziesii* (Mirb.) Franco. Silvae Genet 6:179–185

Sarvas R (1962) Investigation of the flowering and seed crop of *Picea abies*. Commun Inst For Fenn 67.5

Snyder EB (1968) Seed yield and nursery performance of self-pollinated slash pines. For Sci 14:68–74

Sorensen FC (1971) Estimate of self fertility in coastal Douglas-fir from inbreeding studies. Silvae Genet 20:11–120

Sorensen FC, Miles RS (1974) Self-pollination effects on Douglas-fir and ponderosa pine seeds and seedlings. Silvae Genet 23:135–138

Stuber CW, Lincoln LE, Wolff DW, Helentjaris T, Lander ES (1992) Identification of genetic factors contributing to heterosis in a hybrid from two elite maize inbred lines using molecular markers. Genetics 132:823–839

Wilcox MD (1983) Inbreeding depression and genetic variances estimated from self- and cross-pollinated famalies of *Pinus radiata*. Silvae Genet 32:89–96

Wilks SS (1938) The large sample distribution of the likelihood ratio for testing composite hypothesis. Ann Math Stat 9:60–62

Williams CG, Savolainen O (1996) Inbreeding depression in conifers: implications for breeding strategy. For Sci 42:102–117