

Examining the cause of high inbreeding depression: analysis of whole-genome sequence data in 28 selfed progeny of *Eucalyptus grandis*

Philip W. Hedrick¹, Uffe Hellsten² and Dario Grattapaglia^{3,4}

¹School of Life Sciences, Arizona State University, Tempe, AZ 85287, USA; ²US Department of Energy Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA;

³Laboratório de Genética Vegetal, EMBRAPA Recursos Genéticos e Biotecnologia, PqEB, Brasília, DF 70770-970, Brasil; ⁴Programa de Ciências Genômicas e Biotecnologia, Universidade Católica de Brasília, SGAN 916, Brasília, DF 70790-160, Brasil

Author for correspondence:

Philip W. Hedrick

Tel: +1 520 357 7501

Email: philip.hedrick@asu.edu

Received: 17 June 2015

Accepted: 9 August 2015

New Phytologist (2016) **209**: 600–611

doi: 10.1111/nph.13639

Key words: *Eucalyptus grandis*, heterozygosity, heterozygote advantage, linkage disequilibrium, pseudo-overdominance, tree, viability.

Summary

- The genome-wide heterozygosity at 9590 genes, all heterozygous in a single *Eucalyptus grandis* parent tree, was examined in a group of 28 S₁ offspring. Heterozygosity ranged from 52–79%, averaging 65.5%, much higher than the 50% expected under random segregation, supporting the occurrence of strong (47%) selection against homozygosity.
- The expected pattern of heterozygosity from theoretical calculations and simulations for recessive detrimental (pseudo-overdominance) and intrinsic heterozygote advantage was examined and compared with that observed. The observed patterns are consistent with at least several detrimental loci with large effects on both parental chromosomes of the 11 pairs. It is likely that 100 or more genes, many with substantial effects on viability, are contributing to this inbreeding depression.
- Although our genome-wide analysis of nearly 10 000 genes strongly suggested that pseudo-overdominance was responsible for the observed high inbreeding depression, heterozygote advantage could not be excluded.
- Finding inconvertible evidence of the cause of inbreeding depression still presents a difficult challenge. This study is the first theoretical examination of the genomic effect of inbreeding in a forest tree and provides an approach to analyze these data to determine the extent and cause of inbreeding depression across other plant genomes.

Introduction

Inbreeding depression, reduced fitness because of reduced survival, mating, and/or reproduction in the progeny of related individuals, has been documented in many different organisms (Keller & Waller, 2002; Charlesworth & Willis, 2009). In fact, Darwin (1876) carried out experiments in which he self-fertilized, and outcrossed to unrelated individuals, 57 plant species and found substantial negative phenotypic impacts of inbreeding in most of these species. Inbreeding depression is of great importance because of its potential role in evolution, its impact on human health, relevance in crop improvement, and significance in conservation of endangered species (Hedrick & Kalinowski, 2000; Keller & Waller, 2002; Charlesworth & Willis, 2009).

As a result, it is fundamental that the genetic basis of inbreeding depression be understood. It is widely recognized that inbreeding increases the level of homozygosity throughout the genome. The two major hypothesized genetic bases for inbreeding depression are that it is either the result of higher homozygosity of partially recessive (or recessive) detrimental variants (pseudo-overdominance) or for variants maintained by intrinsic

heterozygote advantage at given genes (also called overdominance) (see later Table 2). From the present evidence, it appears that the major cause of inbreeding depression is the increased homozygosity from partially recessive detrimental variants (Charlesworth & Willis, 2009). In support of this conclusion, there is no evidence of large numbers of loci with intrinsic heterozygote advantage (Hedrick, 2012), particularly for loci with the level of selection necessary to generate the observed inbreeding depression.

The partially recessive detrimental variants which are the major cause of inbreeding depression are thought to be individually rare because they are maintained by mutation–selection balance (Charlesworth & Willis, 2009) but these rare variants are thought to be widespread in the genome. If there is a statistical association of these rare detrimental variants with neutral molecular markers, then in progeny segregating for molecular markers there might appear to be heterozygote advantage at a neutral marker locus, a phenomenon termed pseudo-overdominance or associative overdominance (Ohta, 1971). This heterozygote excess occurs because the alleles at the marker locus are in linkage disequilibrium with different partially recessive detrimental variants and in

heterozygotes these are covered up by wild-type alleles on the homologous chromosome. However, in homozygotes at the marker locus, the genotype is homozygous for associated detrimental variants and consequently they have a lowered fitness.

Recently, the genome of *Eucalyptus grandis* (commonly known as rose gum) has been sequenced and analyzed (Myburg *et al.*, 2014). Of particular evolutionary interest here, the genomic variation at 9590 genes that were all heterozygous in a single parent was examined in a progeny array of 28 individuals produced by self-fertilization. The expectation from one generation of selfing without selection, given that the parent is heterozygous, is that 50% of these loci in the progeny would be heterozygous, 50% of the progeny would be homozygous and identical-by-descent for one of the two alternative alleles, and the expected inbreeding coefficient would therefore be 0.5. Contrary to this expectation, 65.5% of these loci were found to be heterozygous, only 35.5% were found to be homozygous (Myburg *et al.*, 2014), and the estimated inbreeding coefficient was 0.355, consistent with strong selection against homozygotes that were identical-by-descent from the one generation of selfing.

In this study, we first examine the observed genomic pattern of heterozygosity in these 28 progeny. In order to comprehend the potential genetic basis of these patterns, we next generate the amount and pattern of heterozygosity expected from pseudo-overdominance given that there are linked recessive detrimental and from loci with a heterozygote advantage using both mathematical analysis and simulation. These theoretical predictions are then compared to the observed data for the 9590 genes in 28 progeny.

Materials and Methods

Background information

Sequencing and single nucleotide polymorphism (SNP) genotyping of the parent tree The parent *Eucalyptus grandis* W. Hill ex Maiden tree M35D2 genome was sequenced at high depth to generate the genome-wide information on heterozygous SNP sites to be later analyzed in the 28 S_1 siblings (Myburg *et al.*, 2014). In total, 204.7 million pairs of 2×100 Illumina reads were trimmed for poor quality and aligned to the *E. grandis* reference sequence using BWA (-q 15). From the perfectly mapped pairs with mapping quality q29 or better, a pileup was generated using samtools mpileup with BAQ scores disabled. An initial set of heterozygous SNPs was called at sites with 20–100 \times coverage of bases of quality Q30 or higher, with at least four bases confirming both variants. This set had 3.18 million SNPs (0.92%) out of >346 million sites with eligible coverage. A much more stringent subset of these was later selected for genotyping the selfed offspring. First, to reduce effects from tandem duplicated regions the analysis was restricted to sites with coverage in the relative narrow window between 40–65 \times (depth distribution peaks at 52 \times). Furthermore, at each site we required the count of the least observed variant be at least 35.8% of the total, which corresponds to the 95% confidence interval of binomially distributed variants with $P = 50\%$ and a total depth of 52. Finally, we chose

to genotype SNPs only at annotated genes (introns and exons) to avoid effects from mis-aligned intergenic repeat families. In total, 308 784 high-confidence heterozygous SNPs within 22 619 genes were selected for genotyping of the selfed offspring.

SNP genotyping of the selfed progeny DNA from the 28 S_1 siblings were barcoded, pooled and sequenced in 11 lanes of 2×100 paired-end Illumina reads. To mitigate any misgenotyping due to low coverage annotated genes were used as ‘natural bins’ and the genotype was based on a consensus of several SNPs present within each gene. More specifically, only SNPs within annotated gene boundaries (UTR + exons + introns) were used, and only genes with at least four SNPs displaying sequence coverage of at least 3 \times were included. The mean (median) number of SNPs per gene with $\geq 3 \times$ coverage was 17 (11), and only 10.4% of the genes had only four SNPs. This indicates that the possibility of wrongly declaring a heterozygous genotype as homozygous was low. Figure 1, data from Myburg *et al.* (2014), give the observed frequency of the heterozygote and the two homozygotes for the 11 chromosomes for 9590 genes at different Mb positions in the 28 S_1 individuals.

Besides the large average excess of heterozygotes and deficiency of homozygotes, there was substantial variation of these frequencies both within and between chromosomes. For six of these chromosomes (4, 6, 7, 8, 9 and 10), one of the homozygotes was missing for part of the chromosome in the 28 progeny. The most extreme homozygosity was found in a region 25 Mb long on chromosome 4 where one of homozygotes was not seen in the 28 progeny.

Table 1 gives a summary of these data for each of 11 chromosomes based on their sequences where the designation of homozygotes refers to the two known parental chromosomes. The chromosomal heterozygosity, which averaged 0.655, ranged from 0.574 for chromosome 5 to 0.710 for chromosome 10, all higher than the 0.5 expected without selection. The average chromosomal homozygosity was 0.345 (on average 0.172 for each homozygote) and ranged from 0.075 for homozygote A_2A_2 on chromosome 10 to 0.302 for homozygote A_2A_2 on chromosome 11. The missing homozygotes in regions of the six chromosomes pointed out earlier might very well indicate that there were one or more lethals were present in these regions when homozygous. Of course, sampling might also contribute to the low numbers of homozygotes in a given region because with 28 progeny, the expected number of progeny assuming Mendelian expectations without selection would be only seven for each homozygote (see later).

Theory and modeling

Before we discuss any theoretical predictions, let us clearly outline the basis of the genetic models for inbreeding depression for the situation in which the parent is heterozygous and progeny are produced by self-fertilization. When there is intrinsic heterozygote advantage (overdominance), then there is a higher fitness for heterozygotes than for homozygotes. In other words, the relative fitness of the progeny heterozygote A_1A_2 is 1 and that of progeny

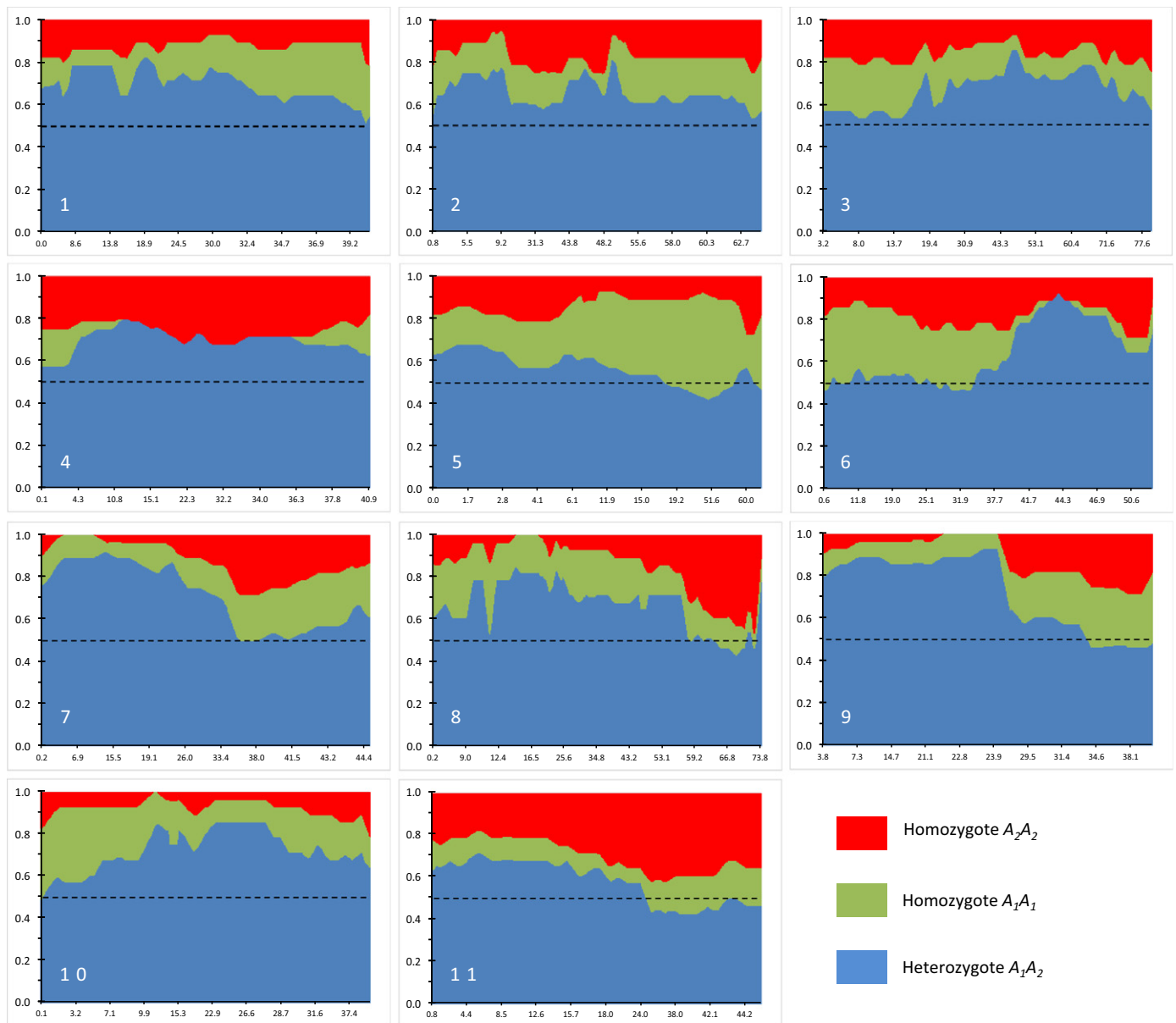


Fig. 1 Genome-wide proportions of heterozygotes (blue) and the two homozygous (green and red) single nucleotide polymorphism (SNP) genotypes at 9590 genes, heterozygous in parent MD35D2, in the 28 S_1 offspring. It is arbitrary which homozygote is green or red on different chromosomes. The panels show the genotypic frequencies for the 11 *Eucalyptus grandis* chromosomes for different Mb positions, data from Myburg *et al.* (2014).

homozygotes A_1A_1 and A_2A_2 are $1 - s_1$ and $1 - s_2$ where s_1 and s_2 are the selection coefficients against genotypes A_1A_1 and A_2A_2 (Table 2a).

When there is pseudo-overdominance, then the marker locus A is assumed to be neutral and in linkage disequilibrium with alleles at other loci with detrimental fitness effects (Hedrick, 2011). Table 2(b) gives the simplest scenario for pseudo-overdominance where allele A_1 at marker locus A is associated with a lethal (or detrimental) allele l_1 at the locus given to the right and a wild-type allele $+_1$ at another locus given to the left. Conversely, allele A_2 is associated with a wild-type $+_2$ allele at the locus to the right and a lethal (or detrimental) allele l_2 at the locus to the left. We will assume here that the $+$ alleles are completely dominant and the l alleles are completely recessive. If there is no breakdown

by recombination of these associations when progeny are produced, then one-quarter of the zygotes in the progeny are expected to be identical-by-descent for chromosome $+_1 A_1 l_1$, one-quarter identical-by-descent for chromosome $l_2 A_2 +_2$, and half are heterozygous for all three loci. As a result, genotype A_1A_1 has a lowered or zero fitness due to selection against the homozygous detrimental or lethal at the locus to the right and A_2A_2 has a lowered or zero fitness due to selection against the homozygous detrimental or lethal to the left.

At this point, we will not assume a particular model of inbreeding depression (later on we will examine both pseudo-overdominance and heterozygote advantage) but just assume that the parent individual was heterozygous A_1A_2 for a given gene A . If the parent is self-fertilized, then given no selection, the

Table 1 The observed heterozygosity and homozygosity and the maximum heterozygosity and minimum homozygosity (for either parental homozygote for a region) on each of the 11 chromosomes for 28 progeny from an individual *Eucalyptus grandis* parent heterozygous for 9590 genes (number of genes on each chromosome is also given) (Myburg *et al.*, 2014)

Chromosome	Number of genes	Heterozygosity		Homozygosity		
		Mean	Maximum	Mean A_1A_1	Mean A_2A_2	Minimum
1	1019	0.697	0.821	0.181	0.121	0.071
2	1327	0.656	0.821	0.159	0.186	0.036
3	1035	0.670	0.857	0.170	0.160	0.071
4	560	0.690	0.786	0.199	0.111	0
5	603	0.574	0.714	0.280	0.147	0.071
6	1203	0.607	0.893	0.206	0.187	0
7	546	0.679	0.893	0.177	0.144	0
8	1185	0.659	0.857	0.092	0.249	0
9	585	0.690	0.893	0.135	0.175	0
10	840	0.710	0.857	0.215	0.075	0
11	687	0.576	0.679	0.122	0.302	0.036
Mean (total)	9590	0.655	0.825	0.176	0.169	0.026

Table 2 The parent and progeny genotypes and fitnesses for the heterozygote advantage (overdominance) and the pseudo-overdominance models. Gene A is the locus with (a) heterozygote advantage or the (b) observed marker locus where / and + indicate lethal (or detrimental) and wild-type alleles at loci in linkage disequilibrium with the A locus for the pseudo-overdominance model. s_1 and s_2 are the selection coefficients against the homozygotes A_1A_1 and A_2A_2 , respectively

	Parent genotype	Progeny genotypes and their fitnesses		
(a) Heterozygote advantage	A_1A_2	A_1A_1 $1 - s_1$	A_1A_2 1	A_2A_2 $1 - s_2$
(b) Pseudo-overdominance	$\frac{+1A_1l_1}{l_2A_2+2}$	$\frac{+1A_1l_1}{+1A_1l_1}$ $1 - s_1$	$\frac{+1A_1l_1}{l_2A_2+2}$ 1	$\frac{l_2A_2+2}{l_2A_2+2}$ $1 - s_2$

expected proportions of genotypes A_1A_1 , A_1A_2 , and A_2A_2 in the progeny are $\frac{1}{4}$, $\frac{1}{2}$, and $\frac{1}{4}$, respectively, where the two homozygotes are identical-by-descent. However, if the homozygotes have lower fitness, then the relative fitnesses of progeny genotypes A_1A_1 , A_1A_2 , and A_2A_2 can be designated as $1 - s_1$, 1, and $1 - s_2$, respectively. If $s_1 = s_2$, then selection against homozygotes is equivalent, or symmetrical, and if s_1 is not equal to s_2 , then selection against the two homozygotes is different, or asymmetrical. Given these relative fitnesses, then the proportions of the three genotypes A_1A_1 , A_1A_2 , and A_2A_2 in the progeny are P , H , and Q , respectively and are equal to:

$$P = \frac{1}{4}(1 - s_1)/\bar{w} \quad \text{Eqn 1(a)}$$

$$H = \frac{1}{2}/\bar{w}$$

$$Q = \frac{1}{4}(1 - s_2)/\bar{w} \quad \text{Eqn 1(b)}$$

where the mean fitness of the progeny is:

$$\bar{w} = 1 - \frac{1}{4}(s_1 + s_2) \quad \text{Eqn 1(c)}$$

Eqn 1(a) for P can be solved for s_1 as:

$$s_1 = \frac{1 - P(4 - s_2)}{1 - P} \quad \text{Eqn 2(a)}$$

and Eqn 1(b) for Q can be solved for s_2 as:

$$s_2 = \frac{1 - Q(4 - s_1)}{1 - Q} \quad \text{Eqn 2(b)}$$

By substitution, these equations can be solved for s_1 and s_2 as:

$$s_1 = \frac{1 - Q - 3P}{(1 - P)(1 - Q) - PQ} \quad \text{Eqn 3(a)}$$

and:

$$s_2 = \frac{1 - P - 3Q}{(1 - P)(1 - Q) - PQ} \quad \text{Eqn 3(b)}$$

If the selection against the two homozygotes is equal, $s_1 = s_2 = s$, then:

$$P = Q = \frac{1}{4}(1 - s)/\bar{w} \quad \text{Eqn 4(a)}$$

and

$$\bar{w} = 1 - \frac{1}{2}s$$

This expression can be solved for s as:

$$s = \frac{1 - 4P}{1 - 2P} \quad \text{Eqn 4(b)}$$

If one of the homozygous classes is missing, say $Q=0$, then Eqn 3(a) can be solved as:

$$s_1 = \frac{1 - 3P}{1 - P} \quad \text{Eqn 4(c)}$$

and from Eqn 3(b), $s_2 = 1$.

In addition, this information can be used to obtain an estimate of the number of lethal equivalents, a measure of the mortality caused by the cumulative effects of all the detrimental alleles when homozygous (Hedrick & Kalinowski, 2000). Assuming the relative mean fitnesses with and without inbreeding are \bar{w}_f and $\bar{w}_{f=0}$, then:

$$B = -\frac{1}{f} \ln \left(\frac{\bar{w}_f}{\bar{w}_{f=0}} \right) \quad \text{Eqn 5}$$

and $2B$ is the expected number of lethal equivalents in a diploid individual (Morton *et al.*, 1956), and is a measure of the inbreeding genetic load.

Let us examine the situation outlined in Table 2 but allow recombination between the locus to the left of the marker and the marker locus. In this case, given that the rate of recombination in this region is c , the different types of gametes produced and their probabilities are given in Table 3. For example, the probability that gamete $+_1 A_1 l_1$ is produced is $(1 - c)/2$ because half of the non-recombinant gametes are of this type. The probability that gamete $l_2 A_1 l_1$ is produced is $c/2$ because half of the recombinant gametes are of this type. In Table 3, the relative fitnesses of the genotypes produced are also given and the fitness of the genotype homozygous for two detrimentals is assumed to be the product of their separate fitnesses (multiplicative fitness values). Multiplying these genotypic fitnesses by the frequencies of the different genotypes, summing them up, and simplifying gives the following expressions for the frequencies of the progeny genotypes $A_1 A_1$, $A_1 A_2$, and $A_2 A_2$:

$$\begin{aligned} P &= \frac{1}{4}(1 - s_1)(1 - c^2 s_2) / \bar{w} \\ H &= \frac{1}{2}[1 - s_2 c(1 - c)] / \bar{w} \\ Q &= \frac{1}{4}[1 - s_2(1 - c)^2] / \bar{w} \end{aligned} \quad \text{Eqn 6}$$

Table 3 The four gametes produced by self-fertilization by a parent with genotype $+_1 A_1 l_1 / l_2 A_2 +_2$ when the rate of recombination between the marker locus and the gene to its left is c , the frequency of these gametes, and the relative fitness of progeny genotypes where s_1 and s_2 are the selection against $l_1 l_1$ and $l_2 l_2$ homozygotes and the fitness is multiplicative over the loci. There is no recombination between the marker locus and the gene to its right

Gamete	Frequency	$+_1 A_1 l_1$ (1 - c)/2	$+_1 A_2 +_2$ c/2	$l_2 A_1 l_1$ c/2	$l_2 A_2 +_2$ (1 - c)/2
$+_1 A_1 l_1$	(1 - c)/2	$1 - s_1$	1	$1 - s_1$	1
$+_1 A_2 +_2$	c/2	1	1	1	1
$l_2 A_1 l_1$	c/2	$1 - s_1$	1	$(1 - s_1)(1 - s_2)$	$1 - s_2$
$l_2 A_2 +_2$	(1 - c)/2	1	1	$1 - s_2$	$1 - s_2$

where

$$\bar{w} = 1 - \frac{1}{4}(s_1 + s_2 - s_1 s_2 c^2)$$

Note that if $c = 0$, no recombination, these expressions become those in Eqn 1.

To examine whether the overall patterns observed for particular chromosomes (Fig. 1) can be generated by pseudo-overdominance or heterozygote advantage, computer simulation of linked loci on a chromosome was carried out (Fortran code for these simulations is available from PWH). In general, either 100 or 1000 loci were assumed to be equally spaced along a chromosome. The two gametes for each individual were randomly generated from the parent. For each gamete, one recombination was generated randomly over the length of the chromosome. This is close to the average number of recombinations per chromosome expected from detailed mapping of 1310 cM for the total genome of 11 chromosomes for *E. grandis* (Petroli *et al.*, 2012). In other words, this assumes that recombination is spread evenly over the 11 chromosomes of 100 cM and there is c . 119 cM per chromosome in the linkage map.

From the cytogenetics standpoint, there is little information on the location of centromeres and potentially associated cold spots of recombination in *Eucalyptus*. However, recent analysis has demonstrated a quite homogeneous rate of recombination across chromosomes with little if any plateaus or restrictions in recombination across the 11 chromosomes (Silva-Junior & Gratapaglia, 2015). In other words, our simulation assumptions of randomly generated recombination along the chromosomes are consistent with the known patterns of recombination in *E. grandis*.

One or more loci for each parental gamete, with a random position on the chromosome, were assumed to carry a recessive detrimental with selection of s in homozygous progeny. Or one of the loci for each parental gamete, with a random position on the chromosome, was assumed to carry variants for heterozygote advantage so that both homozygotes for both parental alleles generally had selection against them of s . When there were multiple selected loci, we assumed fitness to be the product of the fitness of individual loci (multiplicativity). In addition, we examined asymmetrical heterozygote advantage where the equilibrium frequency of allele A_2 is:

$$q_e = \frac{s_1}{s_1 + s_2}$$

Presented later is the case where $s_1 = 0.3$ and $s_2 = 0.9$ so that $s_1 + s_2 = 1.2$, the same sum as the symmetrical case discussed in simulation with $s_1 = s_2 = 0.6$. The equilibrium frequency of A_2 in this case is $q_e = 0.25$, halfway between the equilibrium of $q_e = 0.5$ for $s_1 = s_2 = 0.6$ and that for a recessive lethal of c . 0. Each progeny was randomly determined to die or survive based on its relative fitness, and if it died, another individual was randomly generated to replace it until there were 28 selfed progeny.

Results

Analytical results

First, given an excess of heterozygotes, let us determine the expected level of selection necessary against the homozygotes. Figure 2 gives the selection for two different situations, that is, when the selection is the same against both homozygotes, $P=Q$ or $s_1 = s_2$, using Eqn 4(b), and when one of the homozygotes is missing because of a lethal, $Q=0$ or $s_2 = 1$, using Eqn 4(c). For both situations, the excess of heterozygosity is often the result of quite strong selection. For example, the average heterozygosity observed in the progeny array of 28 individuals was $H=0.655$. If it assumed that the frequency of each of the two homozygotes was reduced by half the amount the heterozygosity was increased, then $P=Q=0.1725$ and using Eqn 3(b), $s=0.473$ as indicated by the closed circle in Fig. 2. In other words, the average relative fitness of a homozygote was only 0.527, slightly above one-half that of heterozygotes. When there is a lethal on one chromosome ($Q=0$), then an excess of heterozygosity can result in a large estimated selection level. For example, if $H=0.75$ and $Q=0$, then, $s_1 = 0.333$ and $s_2 = 1$, compared to $H=0.667$ and $Q=0$ where $s_1 = 0$ and $s_2 = 1$)

Using the genotypic frequencies in Table 1 for the different chromosomes, the estimated level of selection against homozygotes can be estimated using Eqns 3(a) and (b). Table 4 gives these values, which range from a high of 0.789 for homozygote A_2A_2 on chromosome 10 to -0.049 for homozygote A_2A_2 on chromosome 11. These translate into a relative fitness of only 0.211 for homozygote A_2A_2 on chromosome 10 to a relative

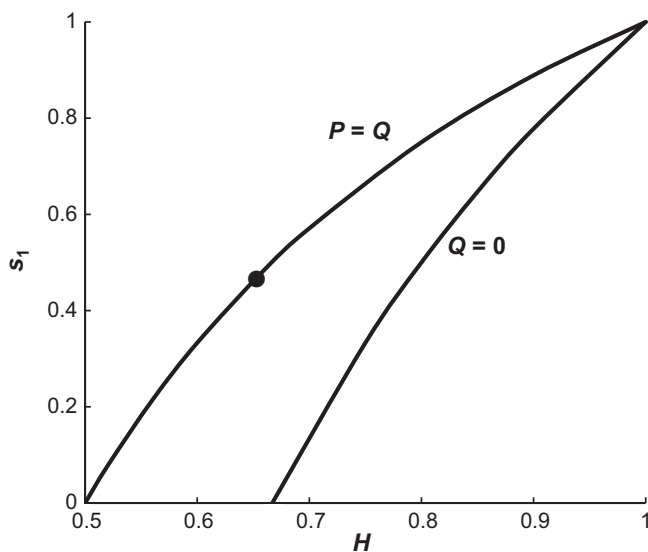


Fig. 2 The estimated selection against one of the homozygotes (s_1) for different levels of heterozygosity (H). When $P=Q$, then $s_1 = s_2$ and the selection against the two homozygotes is the same. When $Q=0$, then $s_2 = 1$ and A_2A_2 is lethal. The closed circle indicates the mean heterozygosity observed (0.655) in the *Eucalyptus grandis* progeny and the consequent estimated selection against the homozygotes, $s_1 = s_2 = 0.473$.

Table 4 The estimated level of selection against the two parental chromosomes (s_1 and s_2) from their frequencies in the 28 *Eucalyptus grandis* progeny using expressions (3a) and (3b); using these values in expression (1c), the mean fitness \bar{w} was calculated

Chromosome	s_1	s_2	\bar{w}
1	0.481	0.653	0.716
2	0.514	0.432	0.763
3	0.492	0.522	0.746
4	0.432	0.678	0.725
5	0.023	0.487	0.873
6	0.321	0.384	0.824
7	0.479	0.576	0.736
8	0.721	0.244	0.759
9	0.609	0.493	0.725
10	0.394	0.789	0.704
11	0.576	-0.049	0.868
Mean or product	0.458	0.482	0.053

fitness of 1.049 for homozygote A_2A_2 on chromosome 11. This later value is slightly higher than the fitness of 1 for the heterozygote, suggesting that there was not a major detrimental effect on this parental chromosome.

The rightmost column of Table 4 gives the relative mean fitness for each chromosome based on Eqn 1(c) and the average estimated values of s_1 and s_2 for that chromosome. These values range from 0.873 for chromosome 5 to 0.704 for chromosome 10. Because these values are really relative survival values and they can be assumed to be multiplicative in their effects, the expected overall relative survival is the product of the mean relative survivals for the 11 chromosomes, or only 0.053. In other words, the relative fitness from one generation of self-fertilization is only 5.3% that when there is no inbreeding. Therefore, the number of lethal equivalents can be estimated from Eqn 5 assuming $\bar{w}_f = 0.053$, $\bar{w}_{f=0} = 1$, and $f=0.5$ as $2B=11.75$, a substantial number of lethal equivalents and a high level of detrimental variation.

Generally, to estimate the number of lethal equivalents, the fitnesses for the inbred and non-inbred groups are known. In this study we are using a different approach and inferring the relative fitness of the selfed group from estimates of the amount of selection necessary to result in the observed deviation from the expected proportions of genotypes in the selfed progeny. This estimate is a lower bound because one generation of recombination has occurred in the production of the selfed progeny which on average should act to reduce the apparent selection against the marker loci. Let us examine the impact of recombination on the frequencies of the three progeny genotypes using Eqn 6. In this study we assume that there is no recombination between the marker locus and the gene with the detrimental allele to the right, in other words the alleles A_1 and l_1 remain completely associated. Figure 3 gives the expected heterozygosity at the neutral marker locus when the selection against homozygotes $l_1 l_1$ and $l_2 l_2$ at the two linked loci is equal and is 1, 0.75, or 0.5.

First, the heterozygosity when there are recessive lethals ($s = 1$) at both loci is at least as high as that observed even when $c = 0.5$. In this case, if the marker locus and completely linked l_1 are in

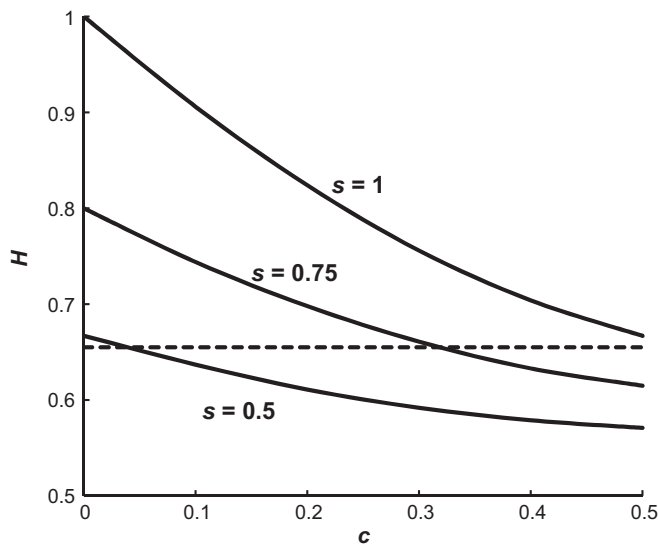


Fig. 3 The expected heterozygosity (H) when there are different rates of recombination (c) between the marker locus and gene to its left for three levels of selection, 1, 0.75, and 0.5, against homozygotes at the detrimental loci. The dashed line indicates the average heterozygosity observed in the 28 *Eucalyptus grandis* progeny from self-fertilization.

the middle of a chromosome 100 map units long, then the whole chromosome is expected to have a heterozygosity as high as that observed. In this case, there is a very low frequency of one of the homozygotes with the expected frequencies of the genotypes A_1A_1 , A_1A_2 , and A_2A_2 when $c=0.5$ are $P=0$, $H=0.667$, and $Q=0.333$.

Second, when there is somewhat less selection against the homozygotes ($s=0.75$), then the heterozygosity is above that observed up to a recombination level c of about 0.3. In this case, if the marker locus is in the middle of a chromosome, then 30 map units on either side is expected to have a heterozygosity as high as that observed. For $c=0.3$, the expected frequencies of the genotypes A_1A_1 , A_1A_2 , and A_2A_2 are $P=0.091$, $H=0.661$, and $Q=0.248$, quite different homozygote frequencies but not unlike some of the mean observed chromosomal values in Table 1. Another example that is informative is if $s=0.5$ for the detrimental on one chromosome and $s=1$ on the other chromosome (not given in Fig. 3). For example, when $c=0.2$, $P=0.096$, $H=0.698$, and $Q=0.206$, again similar to that seen for some chromosomes in Table 1. Overall, for this case, the heterozygosity is similar to that for $s=0.75$ on both chromosomes.

Simulation results

Let us first examine the frequency pattern of the three genotypes on a simulated chromosome which has selection against a detrimental allele at one locus on each parental chromosome. Figure 4 gives the pattern observed from a chromosome with 100 genes when there is one recessive lethal on chromosome copy 1 randomly chosen at gene position 17 and one recessive lethal on chromosome copy 2 randomly chosen at gene position 74 (these positions are indicated by closed circles on the x -axis). First, the overall heterozygosity is 0.635 for 100 genes for 28 simulated

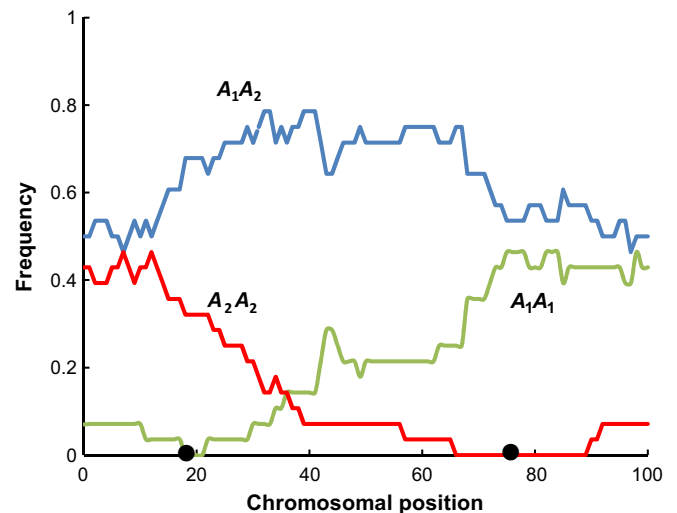


Fig. 4 The frequency pattern of the genotypes over a chromosome with 100 genes when there is a single randomly selected lethal on chromosome 1 at gene position 19 and a single one on chromosome 2 at gene position 74 (closed circles). The average heterozygosity for all 100 loci in this example is 0.635.

progeny, very similar to that in the 28 observed progeny. Second, in the regions near the position of the lethals, the homozygotes are missing from the chromosome with the lethal. Finally, the overall heterozygosity is highest in the center of the chromosome which in this case is between the two lethals.

Next let us examine the frequency pattern of the three genotypes over a simulated chromosome with one locus randomly positioned on each parental chromosome that gives an advantage in heterozygotes. Figure 5 gives the pattern observed when there are loci at positions 26 and 87 that give an advantage in heterozygotes and $s_1 = s_2 = 0.5$. The overall simulated heterozygosity is 0.646, again very similar to that observed in the 28 progeny. Second, in the regions near the selected loci, the heterozygosity is somewhat elevated. Finally, the pattern of frequencies over the chromosome appears more even than that seen for the recessive lethal example in Fig. 4.

To determine the overall effect of having pseudo-overdominance or heterozygote advantage, 1000 samples of 28 individuals with 11 chromosomes with 1000 genes were generated by simulation and Fig. 6 gives the heterozygosity observed for different levels of selection (s) against the homozygote recessive on each chromosome (pseudo-overdominance) or against the homozygotes when there is heterozygote advantage. For pseudo-overdominance, the average heterozygosity ranges from 0.5 when it is assumed that there is no selection operating to 0.633 when it is assumed that there is a recessive lethal on each chromosome. The heterozygosity for these simulated values is similar to the 0.655 in the 28 observed progeny. Using the observed heterozygosity from these simulations, the average selection using Eqn 4(b) was estimated. For example, when s was assumed to 1 in these simulations, it resulted in an average $H=0.633$ and consequently an average $s_1 = s_2 = 0.420$.

For heterozygote advantage, the average heterozygosity ranges from 0.5 when there is no selection operating to 0.753 when the

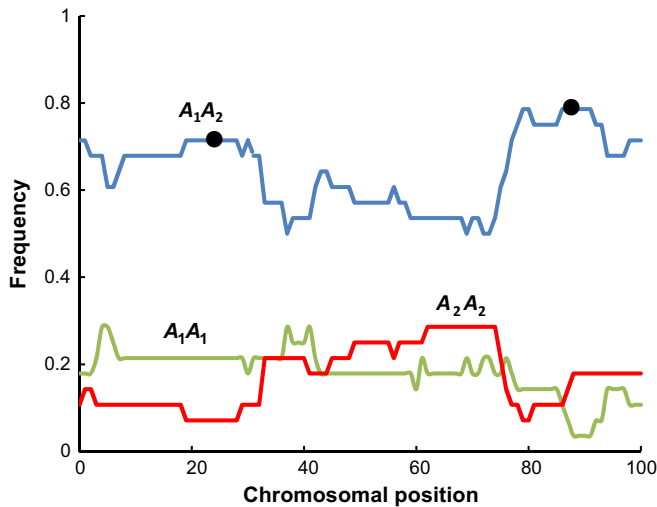


Fig. 5 The frequency pattern of the genotypes over a chromosome when there is a single randomly selected locus at position 87 on chromosome 1 which has a heterozygote advantage when combined with the other chromosome and another similar locus on chromosome 2 at position 26 (indicated by closed circles on the A_1A_2 frequency). The average heterozygosity for all 100 loci in this example is 0.646.

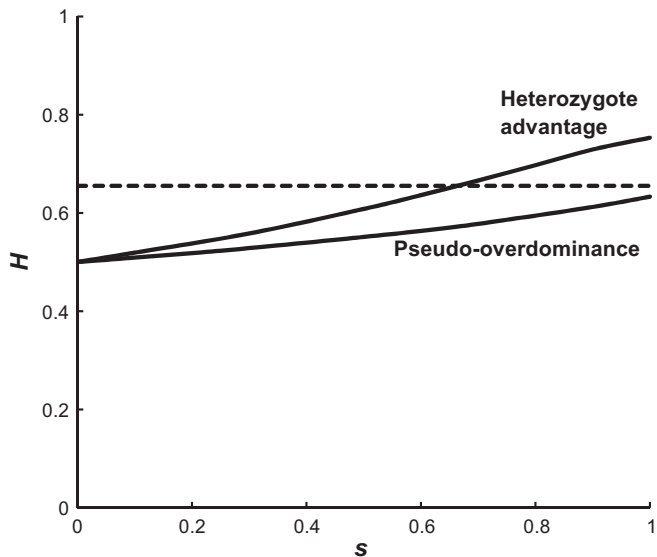


Fig. 6 The heterozygosity (H) estimated from different levels of selection against homozygotes (s) when there is either a single randomly selected recessive lethal locus (pseudo-overdominance) or a single randomly selected locus with an advantage in heterozygotes on each parental chromosome where $s = s_1 = s_2$. These values are the result of 1000 simulations of 11 chromosomes each with 1000 genes each, only two of which are undergoing selection. The broken line indicates the heterozygosity observed in the 28 *Eucalyptus grandis* selfed progeny.

homozygotes at the selected locus are completely selected against, $s = s_1 = s_2 = 1$. Again if we use Eqn 4(b) to estimate the observed level of selection on these chromosomes, $s_1 = s_2 = 0.672$.

The patterns of variation across the chromosomes, as exemplified by the examples in Figs 4 and 5, appear different for pseudo-overdominance and heterozygote advantage. For example, the amount of variation in both homozygosity and heterozygosity

appear less for heterozygote advantage than pseudo-overdominance, the correlation between the two homozygosity values at a given gene appears higher for heterozygote advantage than for pseudo-overdominance, and the frequency of chromosomes with a missing homozygote appears higher for pseudo-overdominance than heterozygote advantage.

These patterns are consistent with intuitive predictions about selection for heterozygote advantage, which would potentially maintain similar homozygosity and heterozygosity levels and lower variation across the chromosome, and would maintain chromosomes with both alleles. By contrast, pseudo-overdominance, which would potentially decrease homozygosity for alleles linked to one parental chromosome that has a lethal or detrimental allele and not for alleles homozygous for the other parental chromosome that did not have a lethal or detrimental, could result in higher variance over the chromosome, a lower (maybe negative) correlation between homozygosity from alleles on the two parental chromosomes, and loss of some homozygous segments along a chromosome.

Table 5 gives the proportion of chromosomes missing a homozygote, the correlation of the frequencies of the two homozygotes and the variances for the frequencies of homozygotes and heterozygotes determined by simulation for no selection, three examples of pseudo-overdominance, and two examples of heterozygote advantage as well as that observed for the 11 chromosomes. On average, for the 11 observed chromosomes 0.273 had a missing homozygous segment, the observed correlation between homozygous frequencies was 0.092, and the variances for homozygous and heterozygous frequencies were 0.0060 and 0.0115. Interestingly, the variance of heterozygosity (Franklin, 1977) from simulations with no selection was only 0.0060 while the average observed heterozygosity for the 11 chromosomes was 0.0115, about twice as large, suggesting that the selection present resulted in an increase in the variance of heterozygosity.

When there is a recessive lethal on each parental chromosome, every chromosome had a segment where the homozygote was missing, the variance of homozygous and heterozygous frequencies are 0.0148 and 0.0152, higher than that observed, and the correlation between the frequencies of the two homozygotes was negative, -0.326 , much lower than that observed. Conversely, for the two other pseudo-overdominance examples, where the levels of selection were picked to result in heterozygosity close to that observed, had fewer missing homozygous segments similar to that observed, somewhat higher correlations of homozygosities but not as high as that observed, and variances similar to that observed. By contrast, when there was a locus with symmetrical heterozygote advantage on each parental chromosome, the frequency of missing homozygous segments was less than that observed, the correlation was even more positive than that observed, and the variances were similar to that observed. For the example of asymmetrical heterozygote advantage on each parental chromosome, the values were closer to that observed than for symmetrical heterozygote advantage. However, even with very strong selection ($s_1 = 0.3, s_2 = 0.9$), the level of heterozygosity was not as high as that observed.

Table 5 The mean heterozygosity (H), frequency of missing homozygous segments (Hom = 0.0), correlation between the two homozygotes (r), and the variance of homozygotes (Var (hom)) and heterozygotes (Var (het)) for 1000 simulations of 11 chromosomes each with 1000 equally spaced genes (see the Materials and Methods section for details) with no selection, detrimental selection with a recessive lethal or multiple recessive deleterious loci, or a heterozygous advantage locus on each parental chromosome. The amount observed for each chromosome in the *Eucalyptus grandis* progeny and the mean over the 11 chromosomes (SE in parentheses) is given at the bottom

		H	Hom = 0.0	r	Var (hom)	Var (het)
No selection	$s = 0$	0.500	0.006	-0.242	0.0045	0.0060
Detrimental selection	$s = 1$, 1 locus	0.632	1.000	-0.326	0.0148	0.0152
	$s = 0.7$, 2 loci	0.648	0.335	-0.286	0.0092	0.0104
	$s = 0.4$, 4 loci	0.645	0.162	-0.270	0.0063	0.0077
Heterozygote advantage	$s = 0.6$	0.635	0.117	0.165	0.0053	0.0124
	$s_1 = 0.3$, $s_2 = 0.9$	0.612	0.167	-0.058	0.0058	0.0104
Chromosome	1	0.697	0.0	-0.34	0.0031	0.0047
	2	0.656	0.0	0.01	0.0025	0.0050
	3	0.670	0.0	0.45	0.0028	0.0078
	4	0.690	0.5	-0.87	0.0121	0.0032
	5	0.574	0.0	-0.85	0.0089	0.0052
	6	0.607	0.5	-0.22	0.0090	0.0148
	7	0.679	0.5	-0.80	0.0064	0.0218
	8	0.659	0.5	0.45	0.0055	0.0150
	9	0.690	0.5	0.90	0.0076	0.0284
	10	0.710	0.5	0.25	0.0041	0.0099
	11	0.576	0.0	0.43	0.0038	0.0103
	Mean	0.655 (0.014)	0.273 (0.075)	-0.054 (0.175)	0.0060 (0.0009)	0.0115 (0.0023)

The values for these measures observed for the 11 individual chromosomes are quite variable and the SEs over the 11 chromosomes are quite high. For example, the correlation ranges from -0.87 for chromosome 4 to 0.90 for chromosome 9 with a SE of 0.175. Chromosome 4 appears to be a good example of a recessive lethal on one of the parental chromosomes because of the missing homozygous region for a large part of the chromosome. Chromosome 11 appears to be consistent with a heterozygous advantage locus with no homozygous segments, a positive correlation, and variances somewhat below that obtained from simulation although a similar patterns are possible with other combinations of multiple recessive alleles with smaller effects and partial recessivity (data not shown). However, the high amount of differences in the individual chromosomes does not allow easy attribution to the effects to pseudo-overdominance or heterozygote advantage in most cases. In other words, the patterns observed did not appear to be completely explained by any of the models examined but the pseudo-overdominance model with several loci per parental chromosome appears generally most consistent with the observed data.

In an effort to understand how much variation might be expected in these measures, given different causes of inbreeding depression, Fig. 7 gives the expected distribution of the correlation between the two homozygous frequencies for the two types of selection for simulations with 1000 replicates of 1000 genes for 11 different chromosomes. Even though the means for pseudo-overdominance and heterozygous advantage are substantially different (as indicated by the closed circles), the distributions for both types of selection range from very negative to very positive values. For example, when there is a recessive lethal, a substantial proportion of the time the correlation is very negative as seen for chromosome 4 but also a substantial proportion of the

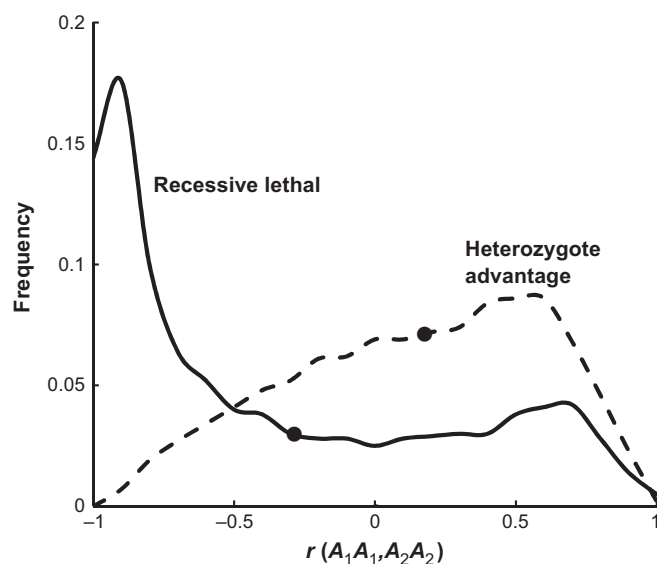


Fig. 7 The expected distribution of the correlation between the two homozygous frequencies $r(A_1A_1, A_2A_2)$ for pseudo-overdominance ($s = 1$) and heterozygote advantage ($s = 0.6$) from 1000 simulated sets of 11 chromosomes, each with 1000 genes. The means are indicated by the closed circles on the two curves.

time the correlation is above 0.5. Similarly, for heterozygote advantage, although there is a mode of this measure $c.$ 0.5, a substantial proportion of the time, the correlation is negative. In other words, it does not appear that the very high expected variance in these data over chromosomes allows for easy differentiation of the cause of inbreeding depression between pseudo-overdominance and heterozygote advantage with the measures suggested here.

Discussion

With genome-wide data from nearly 10 000 genes across 11 chromosomes and a high selection against homozygotes estimated as 47%, we attempted to provide insights into the cause of inbreeding depression, including whether it was from pseudo-overdominance or heterozygote advantage, a question that has been difficult to answer (Charlesworth & Willis, 2009). First, there was a widespread genomic excess of heterozygotes and deficiency of homozygotes on 20 of 22 parental chromosomes and only one parental copy for each of chromosomes 5 and 11 did not show this effect. In addition, for six of the 22 (27%) parental chromosomes, there were regions in which one of the homozygotes was actually missing in the 28 S_1 progeny. Although chance might have some impact on these genotypic proportions, in simulation with no selection, only 0.6% of the time was a homozygote missing (Table 5).

Although chromosomal regions that are missing homozygotes suggest that recessive lethals were present on some of these chromosome, simulations with a recessive lethal on each chromosome always resulted in a region of missing homozygotes for that chromosome. As a result, it appears that some chromosomes did not have a single recessive lethal but several detrimental alleles. In fact, simulations with two loci with $s=0.7$ resulted in 34% of the chromosomes having missing homozygous regions, not greatly different from the proportion observed. For the six regions with missing homozygotes (assuming $Q=0$), the average value of P was 0.149 and using expression (4c), $s_1=0.650$. This suggests that the six chromosomal regions that have missing homozygotes for one parental chromosome also have strong selection against the other chromosome when it is homozygous. In other words, the missing homozygote is in, or close to, a region in which the other parental chromosome is likely to have recessive detrimental alleles. It is possible that a heterozygote advantage locus could cause such a pattern but for symmetrical, and particularly for asymmetrical, heterozygote advantage, the frequency of missing homozygotes on the other parental chromosome was low.

Overall, from these simulations it appears that a large number of loci contribute to inbreeding depression in this progeny group. For example, even with lethals at 22 different genes (two on each of the 11 chromosomes), the expected increase in heterozygosity was still slightly less than that observed. Further, the overall pattern of the proportion of missing homozygotes, correlation of homozygosities, and variances of genotype frequencies was better approximated when there were detrimental alleles at 44 or 88 different genes. Given that there would be variation in the detrimental effects over loci, it is likely that 100 or more genes, many with substantial effects on viability, are contributing to the inbreeding depression in this tree. For heterozygote advantage, 22 different genes with a selective disadvantage in homozygotes of 0.6 are necessary to generate the excess of heterozygosity. This substantial number of genes with large effects suggests that heterozygote advantage is probably not a major contributor to inbreeding depression because very few genes with large heterozygote advantage effects have been identified (Hedrick, 2012).

Charlesworth & Willis (2009) reviewed a number approaches that have been used to evaluate the relative importance in inbreeding depression of pseudo-overdominance and heterozygote advantage (and other hypotheses). One approach that might be useful to distinguish between these alternatives in *E. grandis* is to compare genotypic ratios of other matings between related individuals that have had the potential for recombination, for example, S_2 or full-sib matings. In this case, if pseudo-overdominance is the major cause of inbreeding depression, then the effect would be less while if heterozygote advantage is the major cause, then the effect would be retained. In addition, if heterozygote advantage were important, then in the population causing the effect the variants would be polymorphic and shared over individuals while the variants causing pseudo-overdominance would rarely be shared between unrelated individuals. As a result, crosses of unrelated heterozygotes would have a heterozygosity excess if heterozygote advantage were important while crosses of unrelated heterozygotes would have no heterozygosity deficiency if pseudo-overdominance were important because there would be no identity-by-descent in these progeny. In fact, in the several *Eucalyptus* linkage mapping studies published to date involving segregating populations derived from crossing unrelated trees (reviewed in Grattapaglia *et al.*, 2012), no reports can be found of excess heterozygous individuals, providing further evidence in favor of the pseudo-overdominance hypothesis.

Self-fertilization in *E. grandis* is generally very low with outcrossing rates of 96.7% (self-fertilization of 3.3%) reported in seed orchard studies (Chaix *et al.*, 2003). Burgess *et al.* (1996) also found high outcrossing but also high variation in outcrossing rates for 40 different parental *E. grandis* trees. The seed output per *Eucalyptus* tree is generally large although there is significant annual variation. Natural populations of mixed *Eucalyptus* species were reported to produce at least 40 000 seeds per tree yr^{-1} (Burrows & Burrows, 1992), while managed *E. grandis* seed orchards averaged 20 000–32 000 seeds per tree yr^{-1} (Eldridge *et al.*, 1993). As a result, because of this high effective population size and high potential for selection against low fitness genotypes, the expected level of detrimental genetic variation in *E. grandis* tends to be large. This is also consistent with a relatively high mutation rate recently estimated at the genome-wide level in *E. grandis* ($4.9 \times 10^{-8} \text{ bp}^{-1}$ per generation) (Silva-Junior & Grattapaglia, 2015). In other words, the estimated high number of lethal equivalents in the parental plant and the excess of heterozygosity throughout the genome for the 28 selfed progeny are not unexpected.

Variable self-incompatibility and high genetic load in *Eucalyptus* has been documented, by the observation of widespread and frequently late-acting inbreeding depression in selfed progenies (Griffin & Cotterill, 1988; Hardner & Potts, 1995; McGowen *et al.*, 2010). In the S_1 progeny analyzed in our study, only 18 out of 120 S_1 trees originally planted were still alive at age 7, making the seedling and post-seedling survival 15% and the mortality 85%. The sample of 28 S_1 trees analyzed contained 17 live and 11 dead trees. The mean heterozygosity of the living trees was slightly higher at 0.671, but not significantly different than the heterozygosity of the dead trees at 0.631.

However, the four trees with the highest heterozygosity were all alive in 2011 (mean of 0.765 heterozygosity) and the four trees with lowest heterozygosity were all dead (mean of 0.549 heterozygosity). The fitness value estimated in Table 4 from genotypic frequencies over the 11 chromosomes was 0.053, indicating that survival to the seedling stage was 35.4%. These results taken together suggest that strong selection was also taking place at an earlier developmental stage, such as in the gametophyte or embryo.

An estimate of the impact of inbreeding can be derived from the data of McGowen *et al.* (2010) on the comparative survival of viable seeds in outcrossed vs selfed progenies in *E. globulus*. Based on these data from 105 different selfed mother trees, they report that outcrossing produced on average 20.9 viable seeds per capsule while selfing produced only 3.2 viable seeds per capsule, that is, only 15% of the outcrossed value. Given the wide variation observed among mother trees in the tolerance to selfing and final seed production and the potential impact of inbreeding on other fitness components, our estimate of about 5% survival for this particular *E. grandis* tree M35D2 is well within the possible range. Costa e Silva *et al.* (2011) also examined four levels of inbreeding in *E. globulus* and observed both strong age and environmental effects on inbreeding depression of survival and size but observed overall less inbreeding depression than estimated here.

In Myburg *et al.* (2014) a detailed molecular analysis was carried out to provide biological insights regarding the genetic load that may be segregating in the selfed offspring. The potential functional effects of 308 784 SNPs heterozygous in 22 619 protein-coding genes in the selfed parent M35D2 were investigated. Of the 81 061 SNPs in coding sequence, 38 849 (47.9%) were predicted to be nonsynonymous mutations and of these, 618 were predicted loss of function (LOF) mutations, potentially impacting the function of 584 genes. Such LOF mutations in protein-coding genes likely contribute to the observed inbreeding depression in M35D2. However, when a genomic segment of conserved heterozygosity detected on chromosome 6 was analyzed (26 out of 28 siblings were found heterozygous in this segment), no significant difference was seen in the rate of silent, missense, nonsense and predicted LOF SNPs when compared with the rest of the genome. This result suggests that the molecular nature of the genetic load and resulting inbreeding depression in the selfed progeny cannot be explained by a direct molecular analysis of predicted functional effects.

A further analysis in the segment of conserved heterozygosity on chromosome 6 focused on potential enrichment of genes with tandem duplicates in the 584 genes with predicted LOF SNPs, based on the hypothesis that inactivation of redundant genes can be tolerated by organisms. Although 45.4% of the LOF genes had tandem duplicates, higher than the genomic average of 34.6%, only five of the 16 genes with LOF SNPs within the conserved heterozygosity region had tandem duplicates. This result suggests that one or more of the 11 LOF genes with no tandem duplicates could explain the extreme levels of heterozygosity in this region if any of

them have essential functions in *Eucalyptus*. For example, the *Arabidopsis* homolog of one of these genes, Eucgr.F03739 (At3G19770.1 encoding VACUOLAR PROTEIN SORTING 9A, VPS9A) was previously shown to be essential for embryo development (Goh *et al.*, 2007). Embryogenesis was arrested at torpedo stage in an *Arabidopsis* mutant for the gene. An LOF mutation in the *Eucalyptus* ortholog could result in strong selection against homozygotes. Local accumulation of such recessive mutations on both parental homologs would explain the near absence of both homozygous classes in the selfed progeny of M35D2.

In conclusion, detailed genomic studies of inbred groups, such as the one of present study, will provide substantial insight into the impact of inbreeding on genetic variation and the underlying genetic basis of inbreeding depression. Although our analysis of nearly 10 000 genes in 28 S₁ progeny strongly suggested that pseudo-overdominance was responsible for the observed high inbreeding depression, finding inconvertible evidence of the cause of inbreeding depression still presents a difficult challenge.

Acknowledgements

P.W.H. thanks the Ullman Professorship for partial support. Sequencing and sequence analysis were performed under the auspices of the US Department of Energy's Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Berkeley National Laboratory under contract no. DE-AC02-05CH11231, Lawrence Livermore National Laboratory under contract no. DE-AC52-07NA27344, and Los Alamos National Laboratory under contract no. DE-AC02-06NA25396. DG thanks the Brazilian Ministry of Science, Technology and Innovation through its agency CNPq and the Brazilian Federal District Research Foundation (FAP-DF) for support. We appreciate the comments of M. Bruford, D. Charlesworth, A. Garcia-Dorado, D. Hedgecock, O. Savolainen, J. Wang and several anonymous reviewers on earlier versions of the manuscript and the effort by D. Hedgecock to identify vQTLs.

Author contributions

P.W.H., U.H. and D.G. planned and designed the research, carried out the analysis, and wrote the manuscript.

References

- Burgess IP, Williams ER, Bell JC, Harwood CE, Owen JV. 1996. The effect of outcrossing rate on the growth of selected families of *Eucalyptus grandis*. *Silvae Genetica* 45: 97–100.
- Burrows DM, Burrows WH. 1992. Seed production and litter fall in some eucalypt communities in central Queensland. *Australian Journal of Botany* 40: 389–403.
- Chaix G, Gerber S, Razafimaharo V, Vigneron P, Verhaegen D, Hamon S. 2003. Gene flow estimation with microsatellites in a Malagasy seed orchard of *Eucalyptus grandis*. *Theoretical and Applied Genetics* 107: 705–712.
- Charlesworth D, Willis JH. 2009. The genetics of inbreeding depression. *Nature Genetics* 10: 783–798.

- Costa e Silva J, Hardner C, Tilyard P, Potts BM. 2011. The effects of age and environment on the expression of inbreeding depression in *Eucalyptus globulus*. *Heredity* 107: 50–60.
- Darwin CR. 1876. *The effects of cross and self fertilization in the vegetable kingdom*. London, UK: John Murray.
- Eldridge K, Davidson J, Harwood C, van Wyk G. 1993. *Eucalypt domestication and breeding*. Oxford, UK: Clarendon Press.
- Franklin I. 1977. The distribution of the proportion of the genome which is homozygous by descent in inbred individuals. *Theoretical Population Biology* 11: 60–80.
- Goh T, Uchida W, Arakawa S, Ito E, Dainobu T, Ebine K, Takeuchi M, Sato K, Ueda T, Nakano A. 2007. VPS9a, the common activator for two distinct types of Rab5 GTPases, is essential for the development of *Arabidopsis thaliana*. *Plant Cell* 19: 3504–3515.
- Grattapaglia D, Vaillancourt RE, Shepherd M, Thumma BR, Foley W, Kulheim C, Potts BM, Myburg AA. 2012. Progress in Myrtaceae genetics and genomics: *Eucalyptus* as the pivotal genus. *Tree Genetics and Genomes* 8: 463–508.
- Griffin AR, Cotterill PP. 1988. Genetic-variation in growth of outcrossed, selfed and open-pollinated progenies of *Eucalyptus regnans* and some implications for breeding strategy. *Silvae Genetica* 37: 124–131.
- Hardner CM, Potts BM. 1995. Inbreeding depression and changes in variation after selfing in *Eucalyptus globulus* ssp. *Globulus*. *Silvae Genetica* 44: 46–54.
- Hedrick PW. 2011. *Genetics of populations*. Sudbury, MA, USA: Jones and Bartlett.
- Hedrick PW. 2012. What is the evidence for heterozygote advantage selection? *Trends in Ecology and Evolution* 27: 698–704.
- Hedrick PW, Kalinowski S. 2000. Inbreeding depression and conservation biology. *Annual Review of Ecology and Systematics* 31: 139–162.
- Keller LF, Waller DM. 2002. Inbreeding effects in wild populations. *Trends in Ecology and Evolution* 17: 230–241.
- McGowen MH, Vaillancourt RE, Pilbeam DJ, Potts BM. 2010. Sources of variation in self-incompatibility in the Australian forest tree, *Eucalyptus globulus*. *Annals of Botany* 105: 737–745.
- Morton NE, Crow JF, Muller HJ. 1956. An estimate of the mutational damage in man from data on consanguineous marriages. *Proceedings of the National Academy of Sciences, USA* 42: 855–863.
- Myburg AA, Grattapaglia D, Tuskan GA, Hellsten U, Hayes RD, Grimwood J, Jenkins J, Lindquist E, Tice H, Bauer D *et al.* 2014. The genome of *Eucalyptus grandis*. *Nature* 510: 356–362.
- Ohta T. 1971. Associative overdominance caused by linked detrimental mutations. *Genetical Research* 18: 277–286.
- Petroli CD, Sansaloni CP, Carling J, Steane DA, Vaillancourt RE, Myburg AA, da Silva OB, Pappas GJ, Kilian A, Grattapaglia D. 2012. Genomic characterization of dart markers based on high-density linkage analysis and physical mapping to the eucalyptus genome. *PLoS One* 7: e44684.
- Silva-Junior DA, Grattapaglia D. 2015. Genome-wide patterns of recombination, linkage disequilibrium, and nucleotide diversity from pooled resequencing and single nucleotide polymorphism genotyping unlock the evolutionary history of *Eucalyptus grandis*. *New Phytologist* 208: 830–845.



About New Phytologist

- *New Phytologist* is an electronic (online-only) journal owned by the New Phytologist Trust, a **not-for-profit organization** dedicated to the promotion of plant science, facilitating projects from symposia to free access for our Tansley reviews.
- Regular papers, Letters, Research reviews, Rapid reports and both Modelling/Theory and Methods papers are encouraged. We are committed to rapid processing, from online submission through to publication 'as ready' via *Early View* – our average time to decision is <27 days. There are **no page or colour charges** and a PDF version will be provided for each article.
- The journal is available online at Wiley Online Library. Visit **www.newphytologist.com** to search the articles and register for table of contents email alerts.
- If you have any questions, do get in touch with Central Office (np-centraloffice@lancaster.ac.uk) or, if it is more convenient, our USA Office (np-usaoffice@lancaster.ac.uk)
- For submission instructions, subscription and all the latest information visit **www.newphytologist.com**