

Marker-based investigation of inbreeding depression in the endangered species *Brassica insularis*

S Glémin¹, L Vimond², J Ronfort³, T Bataillon^{3,4} and A Mignot²

¹UMR 'Génome, Populations, Interactions, Adaptations', Université Montpellier II, F-34095 Montpellier, Cedex 05, France; ²Laboratoire Génétique et Environnement CC065, Institut des Sciences de l'Évolution, Université Montpellier II, F-34095 Montpellier, France; ³UMR 'Diversité et Génomes des Plantes Cultivées' INRA, Domaine de Melgueil, F-34130 Mauguio, France; ⁴BiRC – Bioinformatics Research Center, University of Aarhus, Hoegh-Guldbergs Gade 10, Building 1090, DK-8000 Aarhus C, Denmark

Various methods have been proposed to estimate inbreeding depression and to assess its consequences for natural populations. As an alternative to controlled crosses, the use of molecular markers has allowed direct investigation of inbreeding depression in natural populations, but usually suffers from low statistical power. Here, we investigated the effect of inbreeding depression on survival in two populations of the rare species *Brassica insularis*, using both controlled crosses and a marker-based approach. We compare the respective merits of the two approaches for studying

inbreeding depression. We also use information from the molecular markers to dissect in detail patterns of inbreeding depression in this species. *A posteriori*, we find that combining the approaches was not necessary to obtain simple point estimates of inbreeding depression. However, using molecular markers may give insight into the genetic basis of inbreeding depression, such as the occurrence of epistatic interactions among deleterious alleles or purging. *Heredity* (2006) **97**, 304–311. doi:10.1038/sj.hdy.6800870; published online 19 July 2006

Keywords: inbreeding depression; lethal equivalent; molecular markers; inbreeding coefficients; *Brassica insularis*; epistasis

Introduction

Inbreeding depression, the decline in fitness of inbred individuals compared to outbred ones, is increasingly of interest to conservation biologists, as inbreeding effects can be important in wild populations (Keller and Waller, 2002). Several studies suggest that inbreeding depression is largely caused by recessive deleterious mutations segregating at low frequencies within populations (Charlesworth and Charlesworth, 1999). Consanguineous mating through selfing, or mating between close relatives, can make these mutations homozygous – thereby reducing fitness of individuals. If individuals vary in inbreeding levels within populations, inbreeding depression can be expressed in the wild and may affect population dynamics (Keller and Waller, 2002). Such variation can occur in partially selfing populations, or when a proportion of matings involves relatives. In populations with more outcrossing, deleterious mutations can also be transiently expressed owing to genetic drift, when the population size declines. For instance, highly recessive mutations can induce high genetic load after a severe bottleneck (Kirkpatrick and Jarne, 2000). Estimation of the magnitude of inbreeding depression thus gives clear insight into the potential genetic risks of population

size decline. Knowing the genetic basis of inbreeding depression (the distribution of fitness and dominance effects at loci underlying inbreeding depression and the contribution of epistasis) can also be crucial for assessing the genetic risk to natural populations. However, few attempts have been made to investigate the genetic basis of inbreeding depression in natural populations (review by Carr and Dudash, 2003).

The deleterious effect of mutations exposed after inbreeding can be summarised by the concept of lethal equivalents, defined as the number of genes that would cause on average one genetic death if made homozygous (Morton *et al.*, 1956). One lethal equivalent can be owing to a single lethal allele or a large number of mildly deleterious alleles. Morton *et al.* (1956) developed a regression technique to estimate the number of lethal equivalents per gamete as

$$n_1 = \sum_{i=1}^L q_i s_i$$

where L is the total number of loci at which deleterious alleles can occur, s_i is the homozygous detrimental effect of the i th allele, and q_i its frequency. The slope of the regression of the logarithm of survival (or other fitness components) on inbreeding coefficients (f), often referred as \bar{B} and called the inbreeding load, gives a good estimate of n_1 . Recently, Keller and Waller (2002) suggested, using this method more systematically, to estimate inbreeding depression, especially to compare studies spanning different levels of inbreeding coefficients.

Correspondence: S Glémin, UMR 'Génome, Populations, Interactions, Adaptations', Université Montpellier II, CC63 Bat 24, Place Eugène Bataillon, 34095 Montpellier, Cedex 05, France.

E-mail: glemin@univ-montp2.fr

Received 22 November 2005; accepted 2 June 2006; published online 19 July 2006

To obtain these estimates, different crosses are classically performed, using individuals from a base population, to obtain progeny with various degree of inbreeding. In self-compatible hermaphrodite species, most studies compare the effects of outcrossing with self-fertilisation. It is implicitly assumed that individuals in the base population are unrelated and not previously inbred, which can be untrue in small and subdivided populations (Waller, 1993). This false assumption can potentially lead to over- or underestimation of the inbreeding load. Moreover, such methods can be very labour intensive as several generations of crosses may be needed to obtain different inbreeding levels, especially with dioecious or self-incompatible species. Molecular markers can thus be useful because natural variation in inbreeding levels and relatedness can be detected and exploited to obtain a range of inbreeding levels within a single generation cross design. Marker-based methods to estimate inbreeding depression have not yet been used in this way, but have mostly been developed for wild populations, and especially for animal species, for which crosses cannot be easily performed (for review see the meta-analysis of Coltman and Slate, 2003). Such studies try to infer the effect of inbreeding on fitness components in natural conditions by using naturally occurring levels of inbreeding (Keller and Waller, 2002). These studies mainly seek correlations between multilocus heterozygosity measures and fitness components to infer inbreeding depression. Several empirical and theoretical analyses suggest that the power to detect such association will be low (Slate and Pemberton, 2002; Coltman and Slate, 2003; Slate *et al*, 2004), partly because of the low correlation expected between heterozygosity and inbreeding coefficients when the variance of inbreeding coefficients is low (Balloux *et al*, 2004; Slate *et al*, 2004). However, non-equilibrium demographic conditions, such as bottlenecks, can generate high variance of inbreeding coefficient and substantial associative overdominance over the whole genome and may explain the correlations found in several species (Bierne *et al*, 2000). These approaches do not allow precise characterisation of inbreeding depression, for instance in terms of inbreeding load. The use of inbreeding coefficients instead of mere heterozygosity would allow more direct estimation of inbreeding depression, but such an approach suffers from high sampling variance in the inbreeding coefficient estimates and has scarcely been used (Lens *et al*, 2000).

Here, we combine the information brought by controlled crosses as well as marker-based measures of kinship coefficients to estimate inbreeding depression on survival in two natural populations of the rare endangered species *Brassica insularis*. The other aims of this study were (i) to check the accuracy of the marker-based approach by 'benchmarking' it against the classical method, and (ii) to exploit information brought by molecular markers to test for the occurrence of epistasis and purging in *B. insularis* populations.

Materials and methods

Species and populations studied

B. insularis is a rare Mediterranean wild species of the *Brassica oleracea* group ($2n=18$ chromosomes). Its dis-

tribution is restricted to small and isolated populations in Corsica (nine populations), Sardinia (15–17), Tunisia (five), Algeria (one), and in Pantelleria island, near Sicily (one) (Snogerup *et al*, 1990). It lives mainly on limestone or schist cliffs between 300 and 1000 m high. It is a short-lived perennial with insect pollination, which flowers between March and May. Like other *Brassica* species, it has a sporophytic self-incompatibility system. We studied two Corsican populations 40 km distant from each other. The Caporalino population is located in central Corsica, near the town of Corte. The population is estimated to contain around 1500 plants. The Teghime population is located in the north of Corsica, near the town of Bastia, and is estimated to comprise around 2000 plants. These populations are genetically differentiated, as indicated by F_{ST} values estimated from allozyme ($F_{ST}=0.22$; Hurtrez-Bousses, 1996) and microsatellite markers ($F_{ST}=0.16$; Glémin *et al*, 2005). Both populations exhibit high and significant F_{IS} ($F_{IS}=0.22$ for Teghime and $F_{IS}=0.16$ for Caporalino), which suggests the existence of local substructure and biparental inbreeding in these self-incompatible species (Glémin *et al*, 2005). This type of population structure can also generate variance in inbreeding levels within populations.

Experimental design and measured characters

Seeds from these two populations were collected in 1992 and sown at the Conservatoire Botanique National Méditerranéen located on the Porquerolles Island. These plants were used to perform crosses between March and April 2000. Fifteen plants were used for Caporalino and 17 for Teghime (as both maternal and paternal plants). Inbreeding depression and the numbers of lethal equivalents were estimated by comparing offspring of self-crosses with outcrosses using randomly chosen plants within the same population. The kinship coefficients between the parents were then estimated using microsatellite markers (see below). All crosses were performed in insect-proof greenhouses and self-incompatibility was bypassed by pollinating immature buds, before self-incompatibility is expressed. Each of the 32 plants, was crossed with five other plants from the same population. Some crosses yielded few seeds, particularly following self-pollination. In total, the progeny of 13 and 10 maternal plants (for Caporalino and Teghime populations, respectively) were used. For each of these mother plants, crosses with at least two other plants and self-crosses were available, yielding 68 and 54 crosses for each population.

At the end of October 2000, seeds were sown on wet Whatmann paper in Petri dishes. After 6 days, eight germinated seeds per cross were transferred to 'Jiffy' peat pots with basic garden compost and placed in eight randomised blocks in the greenhouse (964 individuals in total). At the beginning of February 2001, surviving plants were transferred into larger pots outside (diameter = 25 cm), keeping the same design of eight randomised blocks. Plants were grown for 2 years until spring 2003. They flowered for the first time in spring 2002. For each plant, we recorded the survival rate at four stages: after the greenhouse period (February 2001, hereafter S1), after 1 year, before the first flowering (January 2002, hereafter S2), after the first flowering (June 2002, hereafter S3), and after the second year (April 2003, hereafter S4).

Estimation of kinship based on microsatellite markers

The 32 plants used for the crosses were genotyped for 11 microsatellite loci, 10 of which were developed in *B. oleracea* (38A, 59A (Szewc-McFaden *et al*, 1996); 10-B01, 10-B06, 10-D08, 10-E06, 10-F07, 11-H09, 12-D02 from UK CropNet Brassica Database: <http://ukcrop.net/perl/ace/search/BrassicaDB?class=Microsatellite>) and one in *Arabidopsis thaliana* (AT129; Callum and Ecker, 1994). We estimated kinship coefficients between mates and inbreeding coefficients of each individual according to Ritland (1996) or Loiselle *et al* (1995) using the SPAGED software (Hardy and Vekemans, 2002). These coefficients are computed relative to reference allele frequencies, which were estimated within each population separately. For outcrossed offspring, we assumed that the inbreeding coefficient, f , was the kinship coefficient between the two parents. Selfed offspring were assigned the inbreeding coefficient $(1+f_i)/2$, where f_i is the inbreeding coefficient of their maternal plant. Both estimation methods (Loiselle's and Ritland's) give similar results (except when indicated) and we present only those using Ritland's estimator.

Statistical analysis

General framework: The inbreeding load was analysed following Morton *et al* (1956). The logarithm of the survival rates was regressed against the estimated inbreeding coefficients. Following Kalinowski and Hedrick (1998), we used a maximum-likelihood approach for parameter estimation and hypothesis testing. The number of surviving individuals over n trials follows a binomial distribution with parameters n and p , where p is the probability of individual survival modelled as $p = e^{-A-Bf}$, where B is the inbreeding load and e^{-A} corresponds to the probability of survival of non-inbred individuals (Morton *et al*, 1956). Inbreeding depression, defined as the decline in fitness of selfed individuals relative to outcrossed ones, is thus estimated by

$$\hat{\delta} = 1 - e^{-\hat{B}/2}$$

The number of lethal equivalents per gamete (n_1) lies between B and $A+B$. If A is small, as usually found, B provides a good estimate of n_1 (Lynch and Walsh, 1997). Extensions to this model can be made to include other factors such as population effects – see, for example, Kruuk *et al* (2002), who considered year effects in their analysis of inbreeding depression in a bird population – or quadratic f terms to test for epistatic interactions across loci (see below).

A convenient framework for analysing such models is the generalised linear model (GLM) (McCullagh and Nelder, 1989). We performed GLM analyses with a binomial error, the natural logarithm as the link function, and a systematic linear component (η) depending on the model tested (see below). Survival rates were estimated as the number of live progeny among the eight full-sibs. As these individuals are genetically correlated, overdispersion can occur (McCullagh and Nelder, 1989, pp 124–128). This does not affect point estimation, but tends to increase the variance estimator by a factor Φ . If $\Phi > 1$, the data are overdispersed (if $\Phi < 1$, they are underdispersed). Φ was incorporated in the model and estimated from the data, and was used to perform F

tests, which are more accurate than the usual χ^2 tests (J-D Lebreton, personal communication; SAS Institute (1999) user guide, version 8). Confidence intervals also depend on Φ , being larger for overdispersed data. For each f value, survival rates were estimated from eight individuals only. In order to avoid having too few individuals in each f category (McCullagh and Nelder, 1989, pp 120–122), we pooled crosses with similar f and assigned the average f to these groups. After pooling, 17 and 13 inbreeding levels were retained for the Caporalino and Teghime populations, respectively. All analyses performed without this pooling procedure gave similar results, but P -values were lower and accordingly confidence intervals narrower. The different models were analysed using the procedure GENMOD of the SAS software (SAS Institute, version 8) using type 1 analysis and the option 'dscale' to estimate the overdispersion parameter Φ and to compute F tests, and the option 'lrci' to compute likelihood-based confidence intervals. Because of the pooling procedure described above, we could not test for random maternal effects. However, a previous analysis on ungrouped data detected neither significant maternal effects nor significant maternal \times inbreeding interactions (results not shown). This result is conservative, as analyses on ungrouped data tend to increase type I error.

Model section: We first tested for populations, inbreeding, and populations \times inbreeding interactions effects. Following the type 1 analysis of deviance method, these factors were tested successively. The corresponding systematic component of the GLM is thus $\eta = -A_i - B_i f$, where $i = \text{Caporalino or Teghime}$. Non-significant factors were removed to obtain a simpler model. Epistatic interactions across loci can be inferred from deviations from the linear model (Lynch and Walsh, 1997). We thus added a quadratic term in f to the model previously retained. To get sufficient power, cumulative mortalities were analysed instead of stage-by-stage mortalities. However, as B is additive, the point estimates of inbreeding loads and inbreeding depression stage-by-stage can easily be obtained by subtracting successive cumulative B estimates.

We then estimated the inbreeding load for each maternal plant separately, without testing for significance. To test for the occurrence of potential purging of this load, we correlated individual inbreeding load with the mother's inbreeding coefficient, using the non-parametric Spearman's correlation. A negative correlation would indicate the occurrence of purging. We performed the test stage-by-stage (using the procedure described above to obtain the B estimates) and on the total survival (corresponding to S4).

Accuracy of the marker-based approach: We compared A and B estimates with those obtained using the classical procedure, which simply compares selfed *vs* outcrossed progeny, assuming $f=1/2$ and $f=0$ for all selfed and outcrossed progenies, respectively. As an intermediate procedure, we also compared self and outcrosses assuming that each category had the mean inbreeding coefficient estimated through microsatellites markers. Overdispersion was taken into account as described above. As only two categories were compared, we did not group the data.

We tested for the effect of the number of loci used to infer kinship coefficients on the B estimates. We re-estimated f coefficients for all possible combinations of 1, 2, ... up to 10 loci and then re-estimated B for each of these 2046 combinations. To automate the procedure, we calculated the inbreeding and kinship coefficients from the per locus values given by SPAGeDI (Hardy and Vekemans, 2002), using the Mathematica Software (version 4.1; Wolfram, 1996). The maximum likelihood estimation of the model parameters was also implemented in Mathematica instead of SAS, using the FindMinimum function to maximise likelihood functions.

Because of the sampling variance in marker-based f coefficient estimates, testing for epistasis through a quadratic regression in f may be questionable. To investigate the robustness of our analysis, we performed two additional tests. In order to simulate the uncertainty of f coefficients, we fitted the quadratic model using f coefficients estimated with all combinations of 7–11 markers (in the same way as we did for B estimates). We also tested for the robustness by repeating the same analysis (with all 11 markers), removing alternatively one of the 23 mothers from the design. Maximum likelihood estimates of the model parameters were also implemented in Mathematica.

Results

Inbreeding and kinship coefficients

For the whole sample, 2–14 alleles per locus were detected, the mean being 6.36. In the Teghime population, one locus was monomorphic. The distribution of offspring individual inbreeding coefficients is presented in Figure 1. Maternal inbreeding coefficients show a large range, from -0.166 to 0.453 , with a mean of 0.110 , leading to inbreeding coefficients for the selfed offspring ranging from 0.417 to 0.726 , with a mean of 0.555 . The distribution is somewhat bimodal, with some individuals much more strongly inbred than the others. This finding suggests that some biparental inbreeding does occur in these populations. Kinship coefficients are also highly dispersed, ranging from -0.146 to 0.536 , with a mean of 0.044 , indicating that some outcrosses can

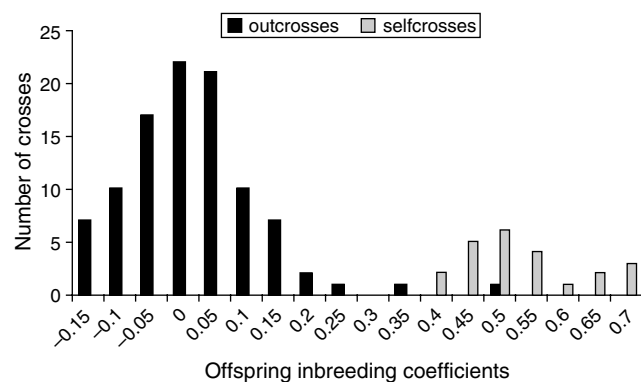


Figure 1 Distribution of the inbreeding coefficients of offspring from outcrosses (black bars) and selfing (grey bars), estimated from Ritland's (1996) method using 11 microsatellite markers. For outcrossed offspring, f is the kinship coefficients between the two parents. For selfed offspring, $f = (1 + f_i)/2$, where f_i is the inbreeding coefficient of the maternal plant.

sometimes lead to offspring as inbred as selfed ones. Despite this continuous distribution, there is a clear separation between outcrossed and selfed offspring (t -test: $t = 21.32$, $P < 0.0001$). These patterns are similar in both populations.

Inbreeding load estimates

For the four survival stages, a significant or marginally significant (P -value of about 0.1) population effect was detected. Mortality was higher in the Caporalino population. However, we did not detect any significant interaction between population and inbreeding levels, suggesting that the two populations do not differ in their inbreeding load. In what follows, the population factor was kept in all analyses for homogeneity and the interaction factor was removed. A common inbreeding load is thus estimated for both populations. Because some parents – those which did not produce seed by selfing – were excluded from the analyses, our estimates might be biased. To mimic the effect of parent exclusion, we re-analysed our data by removing each parent successively. We found that this only very slightly changed the estimate and the significance of population \times inbreeding effects (data not shown and see below for epistasis).

Estimates of the A and B parameters obtained for each stage, using three different methods, are given in Table 1. As general mortality differs between Caporalino and Teghime, we give the two A estimates. Based on the full-regression method, survival values at stages 2 (S2) and 3 (S3) are very similar, and mortality occurred mostly after the two winter seasons (S4). Both the mortality not attributable to inbreeding (given by A) and the inbreeding load (estimated by B) increased through time. After the second winter, the total mortality was very high (about 75%) in the Caporalino population, but only 40% in the Teghime population. The inbreeding load (averaged over the two populations) also increased through time and reached the value of about 1, which corresponds to an inbreeding depression of 40%, a high value given that survival is the only fitness component considered here. As A is high, B underestimates the number of lethal equivalents, which is estimated to be between one and two.

The two other estimation procedures give very similar results (Table 1). Correcting for the mean inbreeding levels of outcrossed and selfed offspring ('out vs self controlled' in Table 1) does not greatly affect the estimates. This result is not particularly surprising, as the mean inbreeding levels are close to the theoretical expectation of 0 and 0.5: 0.071 and 0.54 for Caporalino and 0.04 and 0.52 for Teghime. When all inbreeding coefficients are taken into account (full regression in Table 1), the estimated inbreeding load is lower, although confidence intervals are overlapping with intervals obtained using different methods and coefficients of variation are similar.

Influence of the number of loci

To explore the influence of the number of marker loci used to estimate f on the estimates of inbreeding depression, we used S3 survival because the population effect was not significant ($P = 0.11$). We fitted the simplest model, assuming the same parameters for the two

Table 1 Parameter estimates of survival and inbreeding depression at four stages (S1–S4) using three methods

	Full regression			Out vs self <i>f</i> controlled			Out vs self		
	A_{Capo}	A_{Tegh}	B	A_{Capo}	A_{Tegh}	B	A_{Capo}	A_{Tegh}	B
S1	0.115**** (0.081–0.155)	0.028** (0.015–0.051)	0.144** (0.053–0.251)	0.102**** (0.068–0.141)	0.015 (–0.001–0.038)	0.223**** (0.102–0.379)	0.117**** (0.08–0.154)	0.024*** (0.011–0.045)	0.212*** (0.097–0.360)
S2	0.217**** (0.165–0.279)	0.156**** (0.113–0.212)	0.671**** (0.449–0.912)	0.152**** (0.100–0.208)	0.108**** (0.067–0.156)	1.048**** (0.763–1.373)	0.226**** (0.183–0.275)	0.076**** (0.113–0.196)	0.994**** (0.723–1.302)
S3	0.254**** (0.190–0.332)	0.178**** (0.127–0.247)	0.660**** (0.399–0.947)	0.183**** (0.120–0.251)	0.127**** (0.077–0.185)	1.083**** (0.75–1.468)	0.259**** (0.207–0.319)	0.170**** (0.125–0.225)	1.027**** (0.711–1.392)
S4	1.277*** (1.111–1.459)	0.424*** (0.345–0.516)	0.960**** (0.540–1.424)	1.206**** (1.039–1.387)	0.371**** (0.286–0.466)	1.361**** (0.802–2.012)	1.302**** (1.148–1.470)	0.426**** (0.349–0.513)	1.299*** (0.766–1.919)

** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$.

B is the estimated inbreeding load, assumed to be identical in the two populations as no interaction population $\times f$ was significant; $e^{-A_{Capo}}$ and $e^{-A_{Tegh}}$ are the probability of survival without inbreeding in each population. Full regression: Regression of survival against individuals' f estimated using genetic markers. Out vs self f controlled: Only two categories (selfed and outcrossed) are used, the mean f of each category is estimated using genetic markers. Out vs self: Identical to the second method without controlling for the mean f (ie assuming $f = 0$ vs $f = 0.5$). Below each estimate, a 95% confidence interval is given in parentheses.

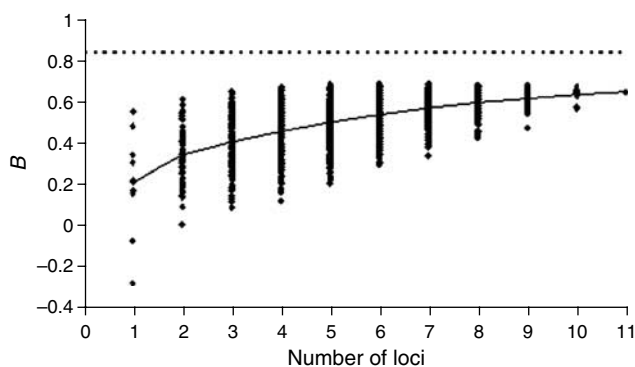


Figure 2 Estimates of the inbreeding load, B , as a function of the number of microsatellite loci used to assess inbreeding levels. Each point corresponds to a particular combination of n loci drawn among 11 available. The plain line corresponds to the mean of B estimates for a given number of loci. The dotted line corresponds to the B extrapolated for an infinite number of loci, $B_{limit} = 0.84$.

populations. The B estimates are plotted in Figure 2. As the number of marker loci used to estimate f increases, the mean value of B increases and the sampling variance around B estimates decreases. Excluding the estimates obtained with only one or two loci, the mean follows a hyperbolic curve almost perfectly. We thus determined the limit that would be reached with an infinite number of loci by regressing $1/B$ on $1/\text{number of loci}$. This value, $B_{limit} = 0.84$, is given by the dotted line in Figure 2.

Testing for epistasis

Table 2 summarises the analysis of deviance results for the quadratic model. As the population effect (A_i) was significant but not the interaction between population and inbreeding coefficient, the systematic component of the model tested was $\eta = -A_i - Bf - Cf^2$. For all stages, the quadratic inbreeding term has a negative effect on survival, but it is significant only for S2 and S3. This indicates that for these two stages the deleterious effect of inbreeding increases more than log-linearly with f . When one mother is removed at random from the design, we still detect a significant effect of the quadratic

Table 2 Analysis of deviance of the quadratic model in f at the four stages

	Estimate	Deviance	$F_{1,26}$	P-value
S1				
Intercept	–0.027	68.20		
pop (Capo)	–0.088	47.08	16.39	0.0004
f	–0.115	33.61	10.45	0.0033
f^2	–0.066	33.51	0.07	0.7890
Scale	1.135			
S2				
Intercept	–0.142	108.52		
pop (Capo)	–0.073	104.42	3.27	0.0816
f	–0.128	40.27	51.18	<0.0001
f^2	–1.259	32.59	6.13	0.0202
Scale	1.120			
S3				
Intercept	–0.163	113.28		
pop (Capo)	–0.090	107.93	3.33	0.0796
f	0.017	51.66	35.00	<0.0001
f^2	–1.535	41.80	6.14	0.0201
Scale	1.268			
S4				
Intercept	–0.414	188.40		
pop (Capo)	–0.869	65.62	96.11	<0.0001
f	–0.392	35.17	23.84	<0.0001
f^2	–1.217	33.21	1.53	0.2270
Scale	1.130			

Significant quadratic terms are in bold.

term (Figure 3a for S2). When f coefficients are estimated with less than 11 loci, numerous tests still are significant at the 5% level (43–63%, from 7 to 10 loci) or at least yield P -values $< 10\%$ (62–100%, from 7 to 10 loci) (Figure 3b for S2). As expected, P -values increase as the number of loci decreases, as tests become less powerful. If the low P -values associated with the quadratic regression term were merely owing to chance under a null hypothesis of no epistasis, the distribution of P -values when repeating the analysis should be flat. This is clearly not the case (Figure 3b). These analyses suggest that we did not detect a significant effect of quadratic term by chance alone.

Maternal inbreeding and test for purging

Because of the large (binomial) variance of the survival data, we did not detect any maternal variation in the inbreeding load. However, point estimates of inbreeding load are quite variable among maternal plants, ranging from -0.10 to 3.5 for S3. Across the two populations, no relationship between these point estimates and the maternal inbreeding levels was detected. However, for the Caporalino population, we detected a significant negative relationship between inbreeding loads and inbreeding levels for the first stage (S1) and over the 2 years (S4) (Table 3). This negative relationship suggests that the most inbred individuals may have purged a part of their load. However, this result is not very robust because the P -values of the same correlations are slightly

higher than the 5% level when using Loiselle's estimator. Evidence of purging must, consequently, be viewed with caution.

Discussion

To use or not to use markers?

We estimated inbreeding depression for survival in the rare species *B. insularis* using three different methods relying on different amounts of marker information. Methods used no marker (self *vs* outcrossed), used markers only to 'control' for the mean inbreeding coefficients of groups of 'selfed' *vs* 'outcrossed' individuals, or used markers to infer individual inbreeding coefficients ('full regression'). The confidence intervals for the results obtained using the three methods are widely overlapping and of similar width (see Table 1). However, the 'full regression' gives systematically lower estimates of inbreeding depression than those obtained by merely comparing selfed and outcrossed progenies. One reason could be that individual kinship coefficients estimated using markers cannot be directly equated to identity by descent coefficients. For example, some individuals have negative estimated f values just because their parents were slightly less related than random individuals. A wider range of f values will decrease the slope of the regression thereby yielding lower estimates of inbreeding depression. Similarly, the variance associated with f estimates will also automatically decrease the slope of the regression. This is clearly illustrated in Figure 2: as the number of loci used to estimate f increases, the variance decreases and the estimated mean B increases. We therefore extrapolated our results to an infinite number of loci. For survival at stage 3, the extrapolated result, B_{limit} is closer to the two other estimates, but still lower (0.84 *vs* 1.08 or 1.02). One can ask how many loci should have been used for making correct inferences. In our data set, the mean B increases slowly with the number of loci. Assuming the extrapolation converges to the true value, 29 loci would have been necessary to reach 90% of B_{limit} and 61 to reach 95%. In accordance with this observation, in studies based on heterozygosity–fitness correlations, more than 50 loci are also typically required to detect inbreeding depression with sufficient power (Slate and Pemberton, 2002; Colman and Slate, 2003).

Why should one then use markers? Molecular markers are often used when crosses cannot be performed, or pedigrees easily reconstructed. If the sole aim is to estimate inbreeding depression from two or more crosses, the regression method does not yield more

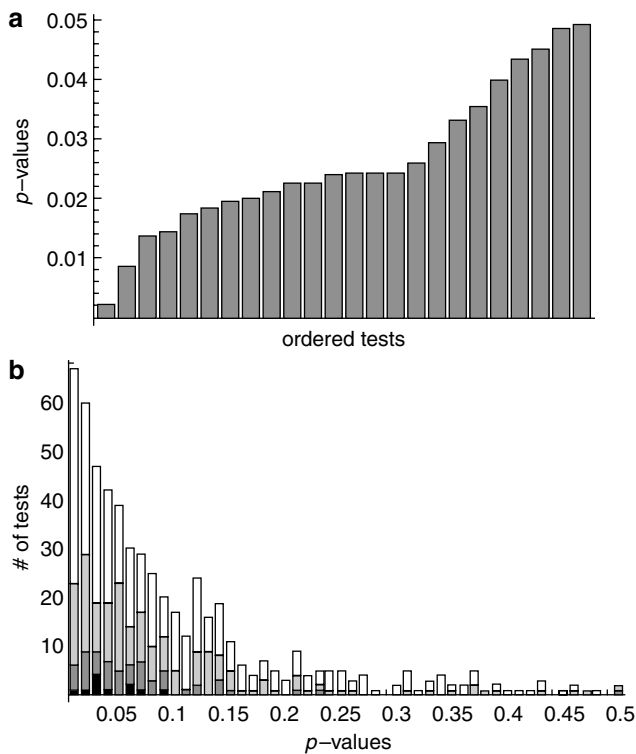


Figure 3 Robustness of the epistasis analysis. (a) Sorted P -values associated with the quadratic inbreeding term (test for epistasis) when each mother is removed successively from the analysis. (b) Distribution of the P -values associated with the quadratic inbreeding term when the analysis is performed with all combinations of 10 loci (black), nine loci (dark grey), eight loci (light grey), and seven loci (white).

Table 3 Stage-by-stage correlation between maternal inbreeding loads and inbreeding coefficients

	Caporalino			Teghime		
	n	r_{Spearman}	P -value	n	r_{Spearman}	P -value
1st stage (S1)	13	-0.577	0.039	10	0.135	0.711
2nd stage (S2–S1)	13	0.093	0.762	10	0.515	0.128
3rd stage (S3–S2)	13	-0.040	0.898	10	-0.450	0.192
4th stage (S4–S3)	9	-0.117	0.766	9	0.283	0.460
Cumulated (S4)	9	-0.783	0.013	9	0.130	0.739

Significant correlations are in bold.

precise results (in terms of confidence intervals) and seems to be biased downwards, at least with the number of loci we used. In addition, using marker information to control for the mean inbreeding levels in outcrossed and selfed progenies yielded estimates very similar to the estimates obtained by assuming that all maternal plants are unrelated and non-inbred (that is assuming that $f=0$ and $f=0.5$ for outcrossed and selfed offspring, respectively). Given the large SE typically associated with inbreeding load estimates, controlling for inbreeding and relatedness was not necessary in our two populations, but it would be useful for populations with strong substructure. Molecular markers are needed to detect hidden population substructure, so that we cannot know *a priori* if they will be useful or not. Controlling for relatedness, even if not always useful, is thus a conservative procedure.

The use of microsatellites also allowed us to obtain some additional insights into the pattern of inbreeding depression detected. With a range of f values (not just two values), we could fit non-linear models, and thereby test for epistatic interactions between deleterious alleles. Because of the large variance of the individual f estimates, the results should be viewed with caution, but additional analyses (Figure 3) suggest that our findings regarding the presence of epistasis are quite robust. If we fit the linear model alone, excluding selfed progenies, inbreeding no longer has a significant effect (for all four stages). This could be interpreted as a lack of power of such an approach (at natural inbreeding levels), leading artificially to a better quadratic fit in some cases. However, for some other traits we studied that are not presented here, we found no evidence of epistasis and significant slopes when excluding selfed progenies, which suggests that the variance in inbreeding levels among crossed progenies is not fully masked by noisy estimates (data not shown).

The occurrence of purging could also be tested thanks to marker information. In our case, given the precision we have in the estimation of individual inbreeding coefficients, the power to detect purging is *a priori* quite low. Indeed, we found no evidence for purging, except in two stages and in one population (Caporalino).

In conclusion, the joint use of controlled crosses and molecular markers is not mandatory to obtain a better estimate of inbreeding depression *per se*, but together they provide additional insights into the genetic basis of inbreeding depression. Such inferences can only be made if there is actually some variance in inbreeding levels. This approach would not, therefore, be very powerful in large panmictic populations, but would work well in populations that undergo some form of mating between close relatives or populations that have experienced a bottleneck in their recent history. This approach should be particularly well suited to populations of partial selfers that exhibit substantial variation in inbreeding levels and inbreeding history. In such natural populations, our approach is an interesting tool to dissect the architecture of inbreeding depression.

Magnitude and timing of inbreeding depression affecting survival in *B. insularis*

We found high and significant inbreeding depression affecting survival in both *B. insularis* populations.

Depending on the estimation approach, inbreeding depression ranged from 28 to 40% before first flowering, and from 38 to 49% at the end of the experiment. In Husband and Schemske's (1996) meta-analysis, the mean inbreeding depression for survival before first flowering was only 12% for allogamous species. Consequently, the number of lethal equivalents we estimated (about one for survival before flowering) is quite high for an angiosperm (most reported estimates are lower than 0.5 (Lynch and Walsh, 1997, p 232). However, these reviews mostly consider annual angiosperms (and perennial gymnosperms). A perennial life history could explain the high inbreeding depression we observed. Indeed, Morgan (2001) argued that, provided that somatic mutations occur, a 'selective sieve' can shift the structure of the genetic load towards highly recessive mutations that are mainly involved in inbreeding depression. This could explain the high numbers of lethal equivalents found in coniferous trees (cf review by Lynch and Walsh, 1997) and the fact that the value we found in our study is intermediate between annual angiosperms and long-lived perennial gymnosperms. Our results also suggest that deleterious alleles may interact synergistically to cause inbreeding depression, at least at some stages. This could also reinforce inbreeding depression as these alleles would be weakly selected as long as the population reproduces through outcrossing, but cause a severe decline in fitness upon inbreeding.

Our results also show that inbreeding depression increases through time, with a strong increase after the winter period. In perennial species, surveys of several life stages and, if possible, over several years are necessary to accurately estimate the magnitude of inbreeding depression. Life stages also differ in their pattern of inbreeding depression. For instance, in the Caporalino population but not in Teghime, we found some evidence for purging, as offspring of inbred plants had lower inbreeding depression (Table 3). Comparison of selfing and outcrossing species indicates that purging mostly occurs in early life stages (Husband and Schemske, 1996), suggesting that mutations affecting early stages have larger selection coefficients and are more recessive. Despite low statistical power (significant results were found with Ritland's estimator but not with Loiselle's one), we also found some evidence for purging affecting early survival (S1 in Table 3) but not in the following stages. However, no significant pattern was observed in the Teghime population. It remains unclear why the two populations differ in this respect and our results must be viewed with caution.

The deleterious consequences of inbreeding in small populations have been much debated with regard to conservation issues (Hedrick and Kalinowski, 2000; Amos and Balmford, 2001; Keller and Waller, 2002). A high genetic load *per se* does not threaten a population as long as it is not expressed, if naturally occurring inbreeding levels are low. This seems to be the case in the populations we studied, as we did not detect any inbreeding effect in outcrossed offspring. However, more consanguineous matings may occur in the wild as some mother plants show quite high inbreeding coefficients. Knowing both the detailed genetic basis of inbreeding depression and the mating pattern is thus required to help diagnose the risk.

Acknowledgements

We thank D Tautzin, M Vabre, S Roussel, D Carbonnell, and C Durand for their great help with the plant experiment, and F Justy and E Paignon for their help on microsatellite genotyping. We also thank M Virevaire and the technical staff of the CBNMP. We are grateful to O Hardy for helpful advices on kinship analyses, and to J-D Lebreton, who suggested taking overdispersion into account in the GLM analyses. D Charlesworth, J Pemberton, I Olivieri, and two anonymous reviewers gave insightful comments on the manuscript. This work has been supported by the 'programme interdisciplinaire du CNRS: Impact des Bio-technologies dans les Agro-écosystèmes'. This is publication ISEM 2006-045 of the Institut des Sciences de l'Evolution.

References

- Amos W, Balmford A (2001). When does conservation genetics matter? *Heredity* **87**: 257–265.
- Balloux F, Amos W, Coulson T (2004). Does heterozygosity estimate inbreeding in real populations? *Mol Ecol* **13**: 3021–3031.
- Bierne N, Tsitrone A, David P (2000). An inbreeding model of associative overdominance during a population bottleneck. *Genetics* **155**: 1981–1990.
- Callum JB, Ecker JR (1994). Assignment of 30 microsatellite loci to the linkage map of *Arabidopsis*. *Genomics* **19**: 137–144.
- Carr DE, Dudash MR (2003). Recent approaches into the genetic basis of inbreeding depression in plants. *Philos Trans R Soc London B* **358**: 1071–1084.
- Charlesworth B, Charlesworth D (1999). The genetic basis of inbreeding depression. *Genet Res Cambridge* **74**: 329–340.
- Coltman DW, Slate J (2003). Microsatellite measures of inbreeding: a meta-analysis. *Evolution* **57**: 971–983.
- Glémin S, Gaude T, Guillemin ML, Lourmas M, Olivieri I, Mignot A (2005). Balancing selection in the wild: testing population genetics theory of self-incompatibility in the rare species *Brassica insularis*. *Genetics* **171**: 279–289.
- Hardy O, Vekemans X (2002). SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Mol Ecol Notes* **2**: 618–620.
- Hedrick PW, Kalinowski ST (2000). Inbreeding depression in conservation biology. *Annu Rev Ecol Syst* **31**: 139–162.
- Hurtrez-Bousses S (1996). Genetic differentiation among natural populations of the rare Corsican endemic *Brassica insularis* Moris: implications for conservation guidelines. *Biol Conserv* **76**: 25–30.
- Husband BC, Schemske DW (1996). Evolution of the magnitude and timing of inbreeding depression in plants. *Evolution* **50**: 54–70.
- Kalinowski S, Hedrick PW (1998). An improved method for estimating inbreeding depression in pedigrees. *Zoo Biol* **17**: 481–497.
- Keller LF, Waller DM (2002). Inbreeding effects in wild populations. *Trends Ecol Evol* **17**: 230–241.
- Kirkpatrick M, Jarne P (2000). The effect of a bottleneck on inbreeding depression and the genetic load. *Am Nat* **155**: 154–167.
- Kruuk LE, Sheldon BC, Merila J (2002). Severe inbreeding depression in collared flycatchers (*Ficedula albicollis*). *Proc R Soc London B* **269**: 1581–1589.
- Lens L, Van Dongen S, Galbusera P, Schenk T, Matthysen E, Van de Castele T (2000). Developmental instability and inbreeding in natural bird populations exposed to different levels of habitat disturbance. *J Evol Biol* **13**: 889–896.
- Loiselle BA, Sork VL, Nason J, Graham C (1995). Spatial genetic structure of a tropical understory shrub, *Psychotria officinalis* (Rubiaceae). *Am J Bot* **82**: 1420–1425.
- Lynch M, Walsh B (1997). *Genetics and Analysis of Quantitative Traits*. Sinauer Associates Inc.: Sunderland.
- McCullagh P, Nelder JA (1989). *Generalized Linear Models*. Chapman & Hall: London.
- Morgan MT (2001). Consequences of life history for inbreeding depression and mating system evolution in plants. *Proc R Soc London B* **268**: 1817–1824.
- Morton NE, Crow JF, Muller HJ (1956). An estimate of the mutational damage in man from data on consanguineous marriages. *Proc Natl Acad Sci USA* **12**: 851–863.
- Ritland K (1996). Estimators for pairwise relatedness and individual inbreeding coefficients. *Genet Res Cambridge* **67**: 175–185.
- SAS Institute (1999). *SAS/STAT User's Guide*. SAS: Cary, NC.
- Slate J, David P, Dodds KG, Veenvliet BA, Glass BC, Broad TE et al (2004). Understanding the relationship between the inbreeding coefficient and multilocus heterozygosity: theoretical expectations and empirical data. *Heredity* **93**: 255–265.
- Slate J, Pemberton JM (2002). Comparing molecular measures for detecting inbreeding depression. *J Evol Biol* **15**: 20–31.
- Snogerup S, Gustafsson M, Bothmer RV (1990). *Brassica* sect. *Brassica* (Brassicaceae). I. Taxonomy and variation. *Willdenowia* **19**: 271–365.
- Szewc-McFaden AK, Kresovich S, Bliet SM, Mitchell SE, McFerron JR (1996). Identification of polymorphic, conserved simple-sequence repeats (SSRs) in cultivated *Brassica* species. *Theor Appl Genet* **93**: 534–538.
- Waller DM (1993). The statics and dynamics of mating systems evolution. In: Thornhill NW (ed) *The Natural History of Inbreeding and Outbreeding*. The University of Chicago Press: Chicago and London. pp 97–117.
- Wolfram S (1996). *The Mathematica Book*. Cambridge University Press: Cambridge.