Supplementary information of "Purging deleterious mutations in conservation programmes: combining optimal contributions with inbred matings"

María Ángeles Rodríguez de Cara[1], Beatriz Villanueva[1], Miguel Ángel Toro[2] and Jesús Fernández[1]

[1] Departamento de Mejora Genética Animal, INIA, Ctra. de La Coruña km. 7.5, 28040 Madrid (Spain), [2] Dept. Producción Animal, E.T.S.I. Agrónomos, Universidad Politécnica de Madrid, 28040 Madrid (Spain)

1 Simulated Annealing

Simulated annealing (Kirkpatrick et al. 1983) is a probabilistic method to locate a good approximation to the global optimum of a given function in a large search space. At each step of the annealing procedure the current solution is perturbed to produce a new candidate solution. This new solution will be accepted depending on the difference between the corresponding function values and also on a parameter termed the "synthetic temperature", T. Thus, if $\exp(-\Delta D/T) > r$, where ΔD is the difference between the function values, and r is a random number in $[0, 1]$, the new solution will be accepted. If the temperature is very large then almost all new solutions will be accepted; conversely when T is small then most solutions will be rejected, with only those that substantially improve the solution being accepted. Thus a major component of simulated annealing is to lower the temperature T during the procedure. Once ΔD starts to decline only slowly, the temperature T is lowered, thus limiting the chances and the distance of a worse solution. The temperature can be lowered several times, and the process may then be "quenched" by accepting only "good" changes to find the local minimum of the cost function.

We used an algorithm of simulated annealing to optimise the contributions, where the function to optimise is

$$\sum_{i=1}^{N_p} \sum_{j=1}^{N_p} \frac{c_i c_j f_{ij}}{T^2} \tag{1}$$

and the variables changed in the optimisation are the c's. We started from a random set of c's, constrained such that the sum of contributions from parents is equal to the sum of contributions from mothers and equal to $2N_p$. ΔD was the change in the value of the sum described in Eq. 1 due to changing to another set of c solutions.

We also used simulated annealing to arrange the matings between contributing individuals according to $\sum_{a,b} t_{a,b} f_{a,b}$, where $t_{a,b}$ is zero if contributing male a and contributing female b do not mate and one otherwise. As the number of female and male contributions are the same, t is a square matrix, and it has $2N_p \times 2N_p$ elements, and each row and column can only have one element equal to one. We started from a diagonal solution for t so that all its elements on the diagonal

were one, and changed rows at random until reaching the optimum. ΔD was then the change in $\sum_{a,b} t_{a,b} f_{a,b}$ between the candidate and the current t matrices.

2 Genetic architecture: number of selected loci

As mentioned in material and methods, we ran simulations to generate populations with genomes including 2000 or 20000 selected loci. Here we show the results for 2000 selected loci (i.e., 100 selected loci per chromosome), evenly distributed across chromosomes and within chromosomes.

With 100 selected loci per chromosome, the ancestral population had over 68% of all loci fixed, and on average, 9% and 85% of the selected loci are fixed in the Mukai and CGD scenarios, respectively. Mean values of mean fitness, mean number of loci per individual carrying deleterious alleles and other variables at $t = 0$ for both mutational scenarios for a population with 100 selected loci per chromosome and 20 chromosomes are given in Tab. 1.

Figures 1 and 2 show the results on fitness and diversity when there are only 100 selected loci per chromosome. We can see there that the differences between using molecular information and using genealogical information in maintaining diversity are much smaller than in scenarios with 1000 selected loci, particularly in the CGD scenario and for $K \geq 10$ (Figure 1).

The reason for such smaller differences is that while there are more markers per selected locus, they are not as tightly linked as they are in the scenario with 1000 selected loci per chromosome.

The initial fitness of the populations are 0.47 and 0.95 under the Mukai and CGD scenarios, respectively, and thus very similar to the populations with 1000 selected loci per chromosome. The rate of inbreeding depression are slightly closer for 100 selected loci, being δ 0.80 and 0.26 under the Mukai and CGD scenarios, respectively. The combination of these values, together with the lower linkage between markers and selected loci, explain why the differences between using molecular and genealogical information are smaller with 100 selected loci. While for 1000 selected loci, the decay in fitness under the Mukai scenario was much stronger for $K = 10$ than for $K = 5000$, for 100 selected loci the differences are noticeably smaller (central and right top panels

of Figure 1), most likely due to the smaller inbreeding load.

References

Kirkpatrick, S.; Gelatt, C. D.; and Vecchi, M. P. ”Optimization by Simulated Annealing.” Science 220, 671-680, 1983.

	Mukai	CGD
$\bar{w}$	0.48 (0.05)	0.96 (0.04)
$\bar{n_a}$	177.70 (9.62)	2.18 (1.49)
$\bar{n_{seg}}$	2607 (58)	2339 (51)
$\bar{\delta}$	0.88 (0.04)	0.26 (0.03)
$\bar{r^2}$	0.04	0.04

Table 1: Mean fitness ($\bar{w}$), mean number of loci carrying a deleterious allele a per individual ($\bar{n_a}$), mean number of segregating loci averaged over individuals ($\bar{n_{seg}}$), average inbreeding load (δ) averaged over the last 1000 generations prior to reaching equilibrium, and average r^2 over the first five positions away from each locus in the ancestral populations in the Mukai and the CGD scenarios with 1000 selected loci, 1000 markers and 2000 neutral loci per chromosome. Standard deviations for each mean are given in brackets.

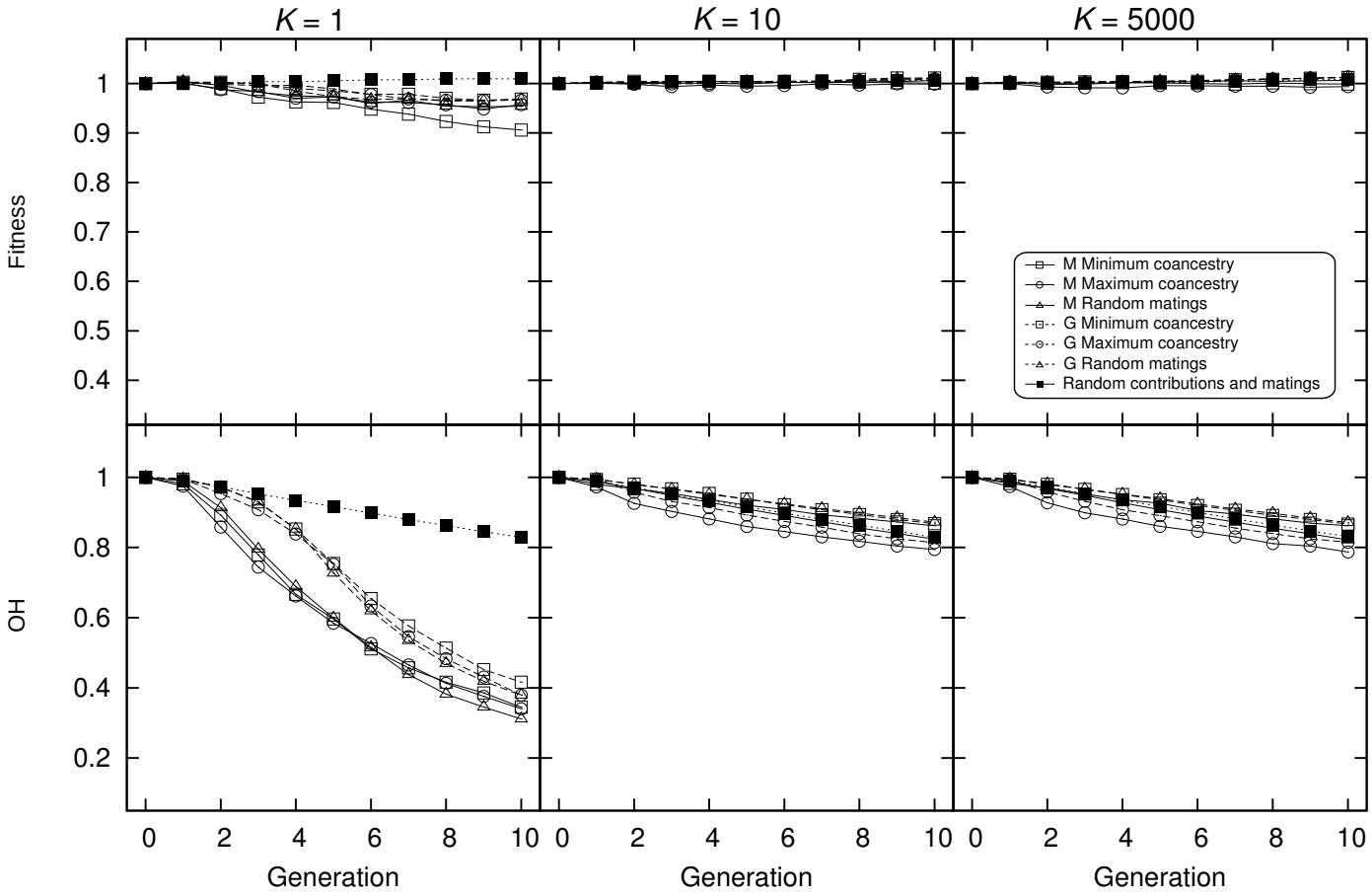

Figure 1: Observed heterozygosity (bottom panels) averaged over all neutral loci that are initially polymorphic and mean fitness (top panels) versus management generations using different management strategies, under the CGD scenario with 100 selected loci per chromosome, using molecular (M) or genealogical (G) coancestry and different mating strategies. The results show averages over 100 replicates, and the maximum standard deviation in mean fitness is 0.003 at $t = 0$ and 0.03 at $t = 10$, and in observed heterozygosity at neutral sites initially segregating is 0.001 at $t = 0$ and 0.006 at $t = 10$.

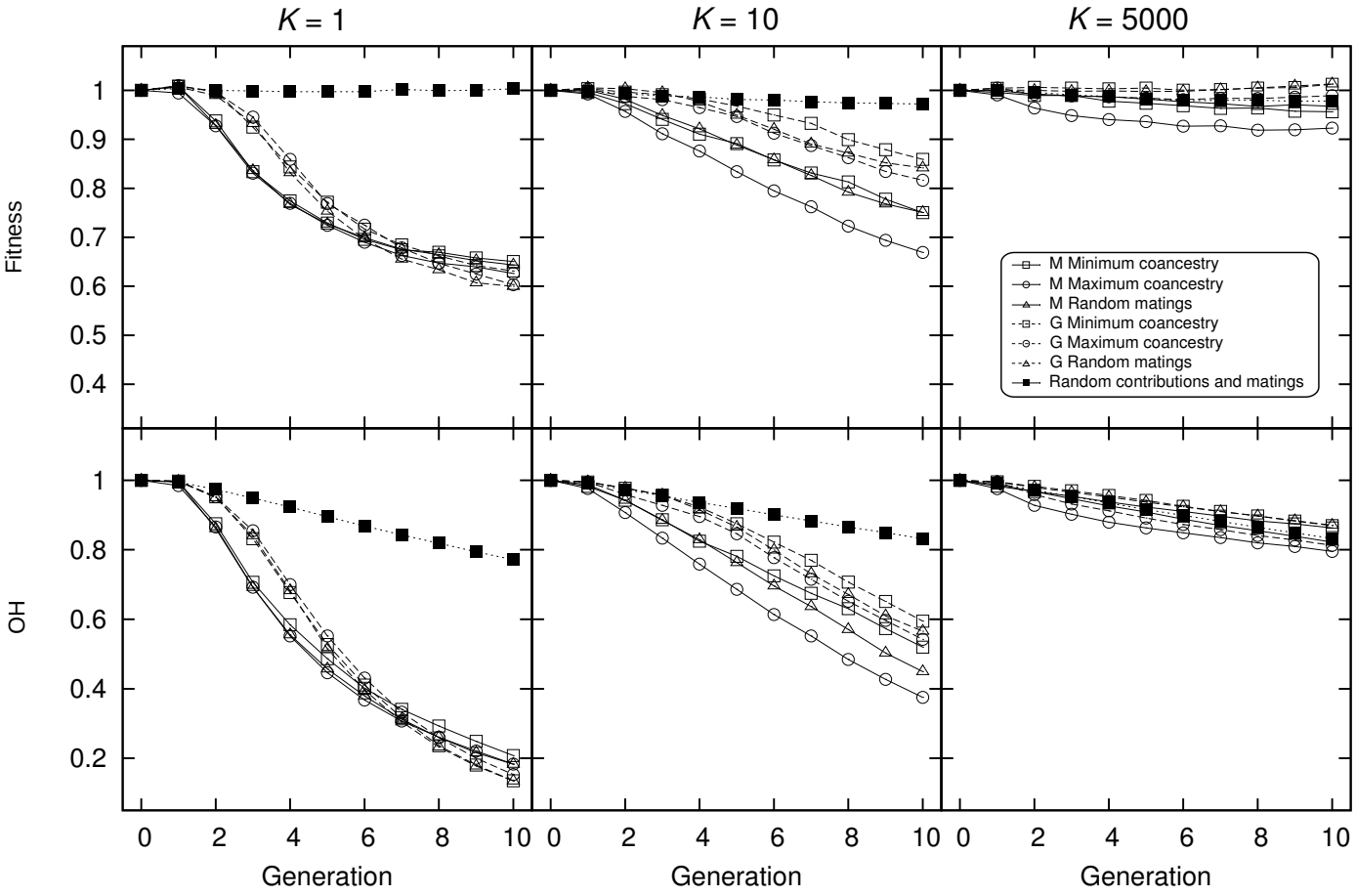

Figure 2: Observed heterozygosity (bottom panels) averaged over all neutral loci that are initially polymorphic and mean fitness (top panels) plotted against management generations using different management strategies, under the Mukai scenario with 100 selected loci per chromosome. using molecular (M) or genealogical (G) coancestry and different mating strategies. The results show averages over 100 replicates, and the maximum standard deviation in mean fitness is 0.005 at $t = 0$ and 0.03 at $t = 10$, and in observed heterozygosity at neutral sites initially segregating is 0.001 at $t = 0$ and 0.006 at $t = 10$.