Benefits Of Using Genome-Wide Information For Managing Conservation Programs

J. Fernández[*], M.A. Toro[†] and B. Villanueva[*]

Introduction

It is generally accepted that the best strategy to manage populations under conservation programs is to determine the contribution (i.e. the number of offspring yield to the next generation) of each potential parent by minimising the global coancestry between individuals weighted by those contributions (e.g., Fernández *et al.*, 2003). This approach maintains the highest levels of genetic diversity and limits the rise of inbreeding. The development of this methodology assumed the use of genealogical information to calculate coancestry. However, this information is often not available.

Fernández *et al.* (2005) studied the ability of molecular markers to replace or complement the pedigree and concluded that the exclusive use of marker information was of very limited value for maintaining genetic diversity. However, this study modelled microsatellite-like markers, relatively scarce, and in linkage equilibrium with the loci on which measurements of diversity were performed. The development of high-throughput genotyping techniques is making a large amount of single-nucleotide polymorphisms (SNPs) available. It is therefore necessary to reanalyse the efficiency of molecular information under this new scenario. The objective of this study was to compare the ability of methods using information from a large panel of SNPs and those relying on pedigree for the management of a conservation program.

Material and methods

Population and genetic model. A population at mutation-drift equilibrium was generated by simulating 5,000 generations of random mating with mutation and recombination and constant effective population size (N_e = 100). The number of chromosomes (c) in the genome was 1 or 20 of 1 Morgan each. At generation 0, 10,000 biallelic genomic non-markers loci evenly spaced and 1,000 SNP markers, also interspersed, were generated per chromosome. The number of SNP and non-marker loci per chromosome still segregating in generation 5,001 was about 850 and 8,500, respectively. Ten males and ten females were randomly sampled from the population at generation 5,001 and they formed the base generation where management started.

Management strategies. Methods implemented can be classified according to two criteria: i) the type of information (pedigree or SNP genotypes) used for computing the coancestry matrix used in the optimisation; and ii) the group of individuals where decisions are taken.

[*] INIA, Departamento de Mejora Genética Animal, Carretera de La Coruña km 7,5, 28040 Madrid, Spain
[†] Universidad Politécnica de Madrid, ETS Ingenieros Agrónomos, Ciudad Universitaria, 28040 Madrid, Spain

Regarding this second classification, we could act on the potential parents, optimising the number of offspring to be obtained from each of them (summing up to a total of 20). This strategy aimed at minimising the expected average coancestry in the offspring by minimising the average coancestry in the parents weighted by their contributions. In a second scenario, potential parents were mated at random and eight offspring (four males and four females) were generated per mating (i.e. a total of 80). Then, the average coancestry of offspring was actually minimised by selecting the group of 20 offspring (10 males and 10 females) which showed the lowest average coancestry between them. Notice that, in the first strategy we calculate expected coancestries of the offspring whereas in the second strategy we calculate observed coancestries.

Summarising, four management strategies were implemented:

- Pedigree: Use genealogical coancestry matrix of parents.
- Molecular: Use molecular coancestry matrix of parents.
- Pedigree (off): Use genealogical coancestry matrix of offspring.
- Molecular (off): Use molecular coancestry of offspring.

In all cases, optimisations were performed via a *simulated annealing* algorithm. Each of the four strategies was applied to the population during ten discrete generations with constant census size. Pedigree and molecular information were recorded every generation from the base population in all scenarios, irrespective of being used or not for the management. Parameters calculated for comparing strategies were the genealogical inbreeding (*F*), the expected heterozygosity or gene diversity (*GD*), allelic diversity (*AD*) and observed heterozygosity (*OH*). The last three parameters were averages across all non-marker loci.

The efficiency of the optimisation using SNP or genealogical information was also compared for larger populations where all the 100 individuals in generation 5,000 constituted the base population. Census size was kept constant across generations. The number of offspring per couple in the strategy relying on offspring information was four males and four females, as before.

Results and discussion

Table 1 shows results for a population of 20 individuals after ten generations of management under different management strategies, assuming small (c = 1) and large (c = 20) genomes. As expected, genome length did not affect the performance of methods based on pedigree information as genealogical coancestry is the probability of identity by descent at any locus in the genome (i.e. an expectation for infinite unlinked loci). The possibility of taking decisions on offspring instead of parents did not change performance either when using pedigree information. It must be noticed that, once the matings are decided, genealogical coancestry is constant, because the relationship with any other individual in the population is the same for all the full-sibs. Contrarily, for a constant marker density, methods using molecular data were more efficient in scenarios with small genomes. With large genomes, the results of the molecular strategy approached the performance of the genealogies strategy when optimisation was based on parents' genotypes.

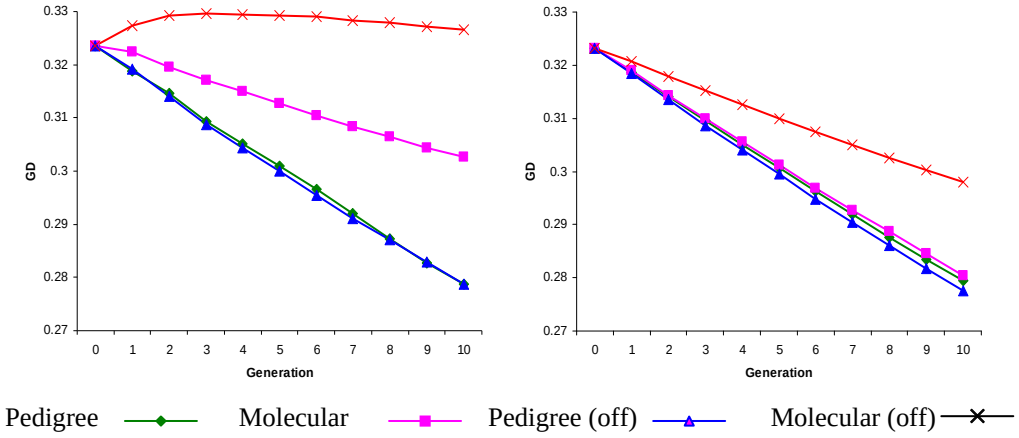
Table 1: Average values (in %) for different diversity parameters after ten generations of management with different strategies, assuming genomes with different number of chromosomes (c) when the census size was 20 individuals*

c	Strategy	*F*	*DG*	*DA*	*OH*
1	Pedigree	12.12	27.87	89.97	29.24
	Molecular	21.62	30.26	90.83	32.23
	Pedigree (off)	12.53	27.87	89.83	29.14
	Molecular (off)	16.07	32.65	93.81	34.18
20	Pedigree	11.98	27.95	90.03	29.31
	Molecular	13.13	28.05	89.84	29.49
	Pedigree (off)	12.31	27.75	89.72	29.20
	Molecular (off)	13.22	29.80	91.35	31.27

*F: genealogical inbreeding; GD: expected heterozygosity or gene diversity; AD: allelic diversity; OH: observed heterozygosity

Methods relying on molecular information yielded higher genealogical inbreeding than those relying on pedigree since the former refer to identity in state instead of identity by descent. However, for all the parameters calculated with non-markers loci genotypes (DG, DA and OH) the use of molecular information gave better results than the use of genealogical information. This superiority was clear for small genomes but it only held for large genomes when offspring information is available. This pattern was observed during the whole period (ten generations) as shown in Figure 1 for the expected heterozygosity. Observed heterozygosity was slightly higher than expected for all scenarios reflecting the effect of minimum coancestry contributions which avoided, indirectly, matings between close relatives (Sánchez *et al.*, 2003). In the scenario where decisions were taken on genotypes of a group of offspring the GD even increased in the first generations because allelic frequencies were moved to intermediate values yielding higher expected heterozygosities.

Figure 1: Gene diversity (GD) across generations when managing a population of 20 individuals under different strategies. Genome length was c = 1 a) or c = 20 b).



Results for larger populations followed a similar trend as those for smaller populations but differences between methods were reduced (Table 2).

Table 2: Average values (in %) for different diversity parameters after ten generations of management with different strategies, assuming genomes with different number of chromosomes (c) when the census size was 100 individuals*

c	Strategy	*F*	*DG*	*DA*	*OH*
1	Pedigree	2.15	32.22	97.75	32.45
	Molecular	10.58	33.67	95.8	34.34
	Pedigree (off)	2.29	32.19	97.63	32.48
	Molecular (off)	7.33	33.98	96.28	34.46
20	Pedigree	2.17	32.27	97.74	32.53
	Molecular	4.16	32.69	97.00	33.07
	Pedigree (off)	2.31	32.23	97.66	32.51
	Molecular (off)	3.18	33.59	97.61	33.89

*F: genealogical inbreeding; GD: expected heterozygosity or gene diversity; AD: allelic diversity; OH: observed heterozygosity

Conclusion

This study shows that the use of large numbers of SNPs for minimizing coancestry leads to higher levels of diversity than using genealogies, especially when several offspring per mating can be genotyped. Notwithstanding, pedigree still appears to be quite powerful for managing conserved populations, particularly for genome sizes typical of those found in livestock species. Other possibilities open by the new massive genotyping techniques such as the possibility of maintaining diversity in specific regions of the genome should be investigated.

Acknowledgements

This work was funded by the Ministerio de Ciencia e Innovación (CGL2009-13278-C02-02).

References

Fernández, J., Toro, M.A. and Caballero, A. (2003). *Genetics*, 165: 885-894.

Fernández, J., Villanueva, B., Pong-Wong, R. and Toro, M.A. (2005). *Genetics*, 170: 1313-1321.

Sánchez, L., Bijma, P. and Woolliams, J.A. (2003). *Genetics*, 164: 1589–1595.