Genet. Sel. Evol. 31 (1999) 255–261

Note

Optimal use of genetic markers in conservation programmes

Miguel Toro*, Luis Silió, Jaime Rodrigañez, Carmen Rodriguez, Jesús Fernández

Departamento de Mejora Genética y Biotecnología, INIA, Ctra. de La Coruña km 7, 28040 Madrid, Spain

(Received 24 November 1998; accepted 26 March 1999)

Abstract – Monte Carlo simulations were carried out in order to study the benefits of using molecular markers to minimize the homozygosity by descent in a conservation scheme of the Iberian pig. A selection criterion is introduced: the overall expected heterozygosity of the group of selected individuals. The method to implement this criterion depends on the type of information available. In the absence of molecular information breeding animals are chosen that minimize the average group coancestry calculated from pedigree. If complete molecular information is known the average group coancestry is calculated either from markers alone or by combining pedigree and genotypes with the markers. When a limited number of markers and alleles per marker are considered, the optimal criterion is the average group coancestry based on markers. Other alternatives, such as optimal within-family selection and frequency-dependent selection, are also analysed. © Inra/Elsevier, Paris

conservation genetics / molecular markers / average coancestry

Résumé – Utilisation optimale des marqueurs génétiques dans les programmes de conservation. Des simulations Monte Carlo ont été effectuées pour étudier l'intérêt de l'utilisation des marqueurs moléculaires pour minimiser le taux d'homozygotie par réplication mendélienne dans un schéma de conservation du porc ibérique. Un critère de sélection a été introduit : le taux global espéré d'hétérozygotie du groupe des individus sélectionnés. La méthode pour appliquer ce critère dépend du type d'information disponible. À défaut d'information moléculaire, on choisit les animaux reproducteurs qui minimisent la parenté moyenne du groupe calculée d'après les généalogies. En cas d'information moléculaire, la parenté moyenne est calculée soit d'après les marqueurs seuls, soit en combinant généalogies et génotypes aux marqueurs. Lorsque l'on considère un nombre limité de marqueurs et d'allèles par marqueur, le critère optimal est la parenté moyenne du groupe conditionnée aux

* Correspondence and reprints
E-mail: toro@inia.es

marqueurs. D'autres alternatives, telles que la sélection intra-familiale et la sélection dépendant des fréquences alléliques, ont été également analysées. © Inra/Elsevier, Paris

génétique de conservation / marqueurs génétiques / parenté moyenne

1. INTRODUCTION

Molecular markers are being advocated as a powerful tool for paternity exclusion and for the identification of distinct populations that need to be conserved [1]. Here we focused on a different application, namely the use of markers to delay the loss of genetic variability in a population of limited size. In a previous paper [12] we conclude that a conventional tactic, such as the restriction of the variance of family sizes, is the most important tool for maintaining genetic variability. In this context, frequency-dependent selection seems to be a more efficient criterion than selection for heterozygosity, but an expensive strategy with respect to the number of genotyped candidates and markers is required in order to obtain substantial benefits.

For this reason, we have considered a new criterion of selection: the overall expected heterozygosity of the group of selected individuals. The implementation of this criterion depends on the type of information available, either from pedigree or from molecular markers. A new type of conventional tactics, optimal within-family selection (OWFS) recently proposed by Wang [14], is also considered.

2. SIMULATION

The breeding population consisted of $N_s = 8$ sires and $N_d = 24$ dams. Each dam produced three progeny of each sex. These 72 offspring of each sex were candidates for selection to breeding of the next generation. This nucleus mimicked the conservation programme carried out in the Guadyerbas strain of the Iberian pig [11].

The techniques of simulation of the genome, marker loci and frequency-dependent selection have been previously described [12]. Here, we introduced a new criterion, the average expected heterozygosity of the group of selected individuals, implemented by three different methods depending on the type of information available: a) average coancestry, including reciprocal and self-coancestries, calculated from pedigree (GCP); b) average coancestry for the markers (GCM), which can be calculated using $1 - \sum_{k}^{L} \sum_{i}^{n} p_{ik}^2$, where p_{ik} is the average frequency in the selected population, of allele i of locus k, n the number of alleles and L the number of marker loci; and c) the average coancestry calculated by combining information given by pedigree and by molecular markers (GCPM). The calculation of coancestry, based on marker information, has been made possible via Monte Carlo Markov chain methods, with the help of a computer program kindly provided by L. Varona [13].

The implementation of this selection criterion would require the examination of $\binom{3N_d}{N_s}\binom{3N_d}{N_d}N_s!$ at all possible combinations and this would be cumbersome even for a small nucleus. It can be solved using integer mathematical programming techniques, whose computational cost would be feasible in most practical situations but not for simulation work, where the algorithm should be used repeatedly. For this reason we used a simulated annealing algorithm [10] that, although not assuring the optimal solution, was generally shown to exhibit a very good behaviour when dealing with similar problems [5, 8].

Besides the basic situation of no restriction on the family sizes, two types of restrictions were considered: a) within-family selection (WFS) where each dam family contributes one dam and each sire family contributes one sire to the next generation; and b) optimal within-family selection (OWFS): among the $\frac{N_d}{N_s}$ dams mated with each sire, one is selected at random to contribute one son, another one to contribute two daughters and the remaining $\left(\frac{N_d}{N_s}\right)-2$ contribute one daughter each [14].

The values of true genomic homozygosity by descent and inbreeding of evaluated individuals at each generation were calculated together with the expected genomic homozygosity of individuals selected from the previous generation and averaged over 100 replicates. The various situations analysed were also compared according to their rate of homozygosity per generation calculated from generation 6 to generation 15 as $\Delta Ho = \frac{Ho(t) - Ho(t-1)}{1 - Ho(t-1)}$, where $Ho(t)$ is the average homozygosity by descent of individuals in generation. The rate of inbreeding was calculated in a similar way.

3. RESULTS AND DISCUSSION

3.1. No molecular information or complete molecular information

Several cases were considered for two extreme situations: the absence of molecular information or the complete knowledge of the genome. The relative ranking of the methods was maintained for all generations and the results of generation 15 are shown in *table I*. With no molecular information, the true homozygosity values were almost identical to those calculated from pedigrees. Optimal within-family selection [14] was substantially (about 15 %) more efficient than classical within-family selection. The restrictions on family size distribution are unnecessary if the method of minimum average group coancestry of selected individuals (GCP) is used. The commonly accepted measure of genetic variability of a population is the expected heterozygosity [9] under the Hardy-Weinberg equilibrium $\left(1 - \sum p_i^2\right)$. In the absence of molecular information the average group coancestry measures the expected homozygosity by descent [4] and therefore the best method for choosing breeding animals should minimize the average group coancestry calculated from pedigree [2–4, 7]. If only full and half-sib relationships are considered, the criterion would lead to the optimal within-family selection method proposed by Wang [14].

Table I. Rate of genome homozygosity (ΔHo, %) and inbreeding (ΔF, %) from generations 6–15. Homozygosity (Ho^{15}, %) and expected genome homozygosity (EHo^{15}, %) at generation 15.

Molecular information	Selection criterion	Restriction in family sizes	ΔHo	ΔF	Ho^{15}	EHo^{15}
None	none	none	1.87	1.89	25.14	27.85
		WFS	1.32	1.34	18.96	20.98
		OWFS	1.17	1.17	16.30	18.75
	group coancestry (GCP)	none	1.12	1.14	15.96	18.55
Complete	group coancestry (GCM)	none	0.71	1.19	11.08	13.36
		WFS	0.69	1.24	11.26	13.00
		OWFS	0.69	1.21	11.58	13.38
	frequency-dependent	none	1.86	1.88	24.80	27.74
		WFS	0.72	1.22	12.25	13.98
		OWFS	0.68	1.14	11.09	13.12

WFS, within-family selection; OWFS, optimal within-family selection; GCP, coancestry calculated from pedigree; GCM, calculated from markers. Standard errors of Ho^{15} range from 0.06 to 0.13.

When using complete molecular information for selection, the best method was still the same although now the true coancestry for all of the genome was known. In this case, the inbreeding coefficient did not reflect the true homozygosity, and the discrepancy could have been considerable. Furthermore, the rate of advance in the true homozygosity, unlike the rate of inbreeding, does not attain an asymptotic value after a short number of generations but decreases continuously.

The method of minimum average group coancestry using all the molecular information (GCM) reduced the rate of homozygosity by almost a half, although the algorithm utilized did not warrant the attainment of the optimal solutions. The impact of imposing additional restriction on family size was negligible. In a balanced structure, the minimization of average coancestry is mainly attained, as previously explained, by selecting individuals from different families.

Frequency-dependent selection, very easy to apply, can also be efficient as a conventional tactic, although not being theoretically justified and therefore lacking generality. The results of frequency-dependent selection depended on family size restrictions. Without restrictions, the results were almost as bad as when the molecular information was ignored owing to an increasing tendency to co-select sibs [12]. But, after optimal family size restrictions were imposed, the method was as good as the group coancestry method, since the differences were not significant.

3.2. Limited number of markers and alleles per marker

The relative utility of the number of markers and alleles per marker is presented in *table II*, where values of the true genomic homozygosity and inbreeding are given for three situations: average group coancestry criterion (GCM), used either without restriction or with optimal family size restrictions, and frequency-dependent selection with optimal family size restrictions. The cases of complete or null marker information are also presented for comparison.

Table II. Rate of genome homozygosity (ΔHo, %) and inbreeding (ΔF, %) and homozygosity at generation 15 (Ho^{15}, %) for different tactics of selection based on molecular information.

Markers/ Chromosome	Alleles/ Marker	GCM + no restriction in family size			Frequency-dependent + OWFS			GCM + OWFS		
		ΔHo	ΔF	Ho	ΔHo	ΔF	Ho	ΔHo	ΔF	Ho
0	0	1.87	1.89	25.14	1.17	1.17	16.30	1.17	1.17	16.30
1	4	1.33	1.45	18.71	1.10	1.15	15.80	1.11	1.17	15.68
	10	1.15	1.34	16.53	1.07	1.20	15.39	1.00	1.16	14.68
2	4	1.21	1.38	17.37	1.07	1.17	15.47	1.01	1.15	15.13
	10	1.06	1.31	15.35	0.97	1.17	14.45	0.92	1.15	13.85
6	4	1.05	1.30	15.46	0.98	1.17	14.56	0.91	1.14	13.95
	10	0.94	1.28	14.01	0.86	1.16	13.31	0.82	1.16	12.75
10	4	1.02	1.31	14.89	0.93	1.15	14.00	0.88	1.15	13.56
	10	0.89	1.23	13.44	0.82	1.15	12.86	0.76	1.13	12.17
100	64	0.71	1.19	11.08	0.68	1.14	11.09	0.69	1.01	11.58

GCM, group coancestry calculated from markers; OWFS, optimal within family selection. Standard errors of Ho^{15} range from 0.06 to 0.13.

As the number of markers and alleles per marker increased, the genome homozygosity attained at generation 15 decreased although it was not adequately reflected in the inbreeding coefficient. This also confirmed our previous finding [12] that the value of a marker is related to the number of alleles: two markers with ten alleles are as valuable as six markers with four alleles.

The results also indicated that the use of the method of minimum average group coancestry (or expected heterozygosity) based only on molecular data without family restrictions was not a good criterion even with a huge amount of molecular information. The use of this method while applying the optimal restrictions on family sizes emerged from *table II* as a better criterion (10 % of advantage). Our results, not shown here, also confirmed that slight improvements in the conventional tactics could have an important impact on the maintenance of genetic variability. Thus, OWFS with three markers/chromosome and four alleles/marker was as efficient as WFS with ten markers/chromosome and four alleles/marker (14.80 of genome homozygosity at generation 15 in both cases). Finally, frequency-dependent selection with optimal family restriction, which was previously analysed in more detail [12], provided good results, and was more easy to implement.

Finally, *table III* shows a comparison of the values for genome homozygosity when using the method of minimizing average group coancestry for markers (GCM) together with restrictions on family sizes with the theoretically optimal method of minimizing average group coancestry based on marker information (GCPM). In order to diminish the high computing cost of the analysis of pedigree involved in the last method, the genome size has been reduced to just one chromosome of 100 cM. Due to this smaller genome size, selection was more efficient and the results of the method of the average group marker coancestry with optimal restrictions were now better than those shown in *table II.* Results shown in *table III* also indicated that the method of average group coancestry based on the markers was 20–30 % more efficient. This comparison was only strictly valid for the genome size considered, but it can be safely concluded that the last method could contribute substantially to the efficiency of a marker-assisted conservation programme.

Table III. Rate of genome homozygosity (ΔHo, %) and homozygosity at generation 15 (Ho^{15}, %) for two types of selection: average group coancestry of markers plus optimal family restrictions (GCM + OWFS) or average group coancestry conditional on markers (GCPM). (Results for a genome size of one chromosome of 100 cM.).

Markers	Alleles Marker	GCM + OWFS		GCPM	
		ΔHo	Ho^{15}	ΔHo	Ho^{15}
0	0	1.16	16.13		
1	4	1.09	15.61	0.80	11.41
	10	0.98	14.29	0.76	10.75
2	4	1.01	14.72	0.66	9.57
	10	0.82	12.58	0.57	8.38
6	4	0.85	12.78	0.40	6.28
	10	0.55	9.39	0.37	5.78
10	4	0.71	11.47	0.34	5.48
	10	0.44	8.09	0.33	5.20
100	64	0.20	4.50		

Standard errors of Ho^{15} range from 0.09 to 0.20.

Although the conclusions obtained through simulation probably have some generalities, it should be recognized that some theoretical developments on marker-assisted conservation are needed. In recent years, substantial work has been carried out on the joint prediction of inbreeding and genetic gain when selecting for a quantitative trait (see [15], for the latest development of the theory). However, predictions on the rate of advance of the true homozygosity by descent when the selected trait is the heterozygosity itself, measured either by molecular or pedigree information, is lacking.

The use of an optimal method enhances the prospectives of the application of molecular markers in conservation programmes, although the future will depend critically on DNA extraction and genotyping costs. Microsatellite DNA markers have been considered until now as the most useful markers, especially when

multiplex genotyping is used, but in the near future other DNA polymorphisms such as SNP could be the most adequate for routine scoring [6]. It is also interesting to emphasize that the adequate use of molecular tools requires increasingly sophisticated methods of Monte Carlo analysis of pedigree and more powerful methods of combinatorial optimization.

ACKNOWLEDGEMENT

This work was supported by the INIA grant SC98-083.

REFERENCES

[1] Avise J.C., Molecular Markers, Natural History and Evolution, Chapman & Hall, New York, 1994.

[2] Ballou J.D., Lacy R.C., Identifying genetically important individuals for management of genetic variation in pedigreed populations, in: Ballou J.D., Gilpin M.A., Foose T.J.R. (Eds.), Population Management for Survival and Recovery. Analytical Methods and Strategies in Small Population Conservation, Columbia University Press, 1995, pp. 76–111.

[3] Brisbane J.R., Gibson J.P., Balancing selection response and rate of inbreeding by including genetic relationship in selection decisions, Theor. Appl. Genet. 91 (1995) 421–431.

[4] Caballero A., Toro M.A., Interrelations between effective population size and other pedigree tools for the management of conserved populations, Genet. Res. (submitted).

[5] Fernández J., Toro M.A., The use of mathematical programming to control inbreeding in selection schemes, J. Anim. Breed. Genet. (1999) (in press).

[6] Kruglyak L., The use of genetic map of biallelic markers in linkage studies, Nat. Genet. 17 (1997) 21–24.

[7] Meuwissen T.H.E., Maximizing the response of selection with a predefined rate of inbreeding, J. Anim. Sci. 75 (1997) 934–940.

[8] Meuwissen T.H.E., Woolliams J.A., Maximizing genetic response in breeding schemes of dairy cattle with constraints on variance of response, J. Dairy Sci. 77 (1994) 1905–1916.

[9] Nei M., Analysis of gene diversity in subdivided populations, Proc. Nat. Acad. Sci. USA 70 (1973) 3321–3323.

[10] Press W.H., Flannery B.P., Teukolsky S.A., Vetterling W.T., Numerical Recipes, Cambridge University Press, Cambridge, 1989.

[11] Rodrigañez J., Silió L., Toro M.A., Rodriguez C., Fifty years of conservation of a black hairless strain of Iberian pigs, in: Matassino D., Boyazoglou J., Cappuccio A. (Eds.), International Symposium on Mediterranean Animal Germplasm and Future Challenges, EAAP Publication no. 85, Wageningen Pers, 1997, pp. 183–186.

[12] Toro M.A., Silió L., Rodrigañez J., Rodriguez C., The use of molecular markers in conservation programmes of live animals, Genet. Sel. Evol. 30 (1998) 585–600.

[13] Varona L., Pérez-Enciso, M., Detección de QTLs mediante la partición de la varianza genética en función del parentesco atribuible a segmentos del genoma, ITEA Producción Animal 94A (3) (1998) 265–270.

[14] Wang J., More efficient breeding systems for controlling inbreeding and effective size in animal populations, Heredity 79 (1997) 591–599.

[15] Woolliams J., A recipe for the design of breeding schemes, Proc. 6th World Cong. Genet. Appl. Livest. Prod. 25 (1998) 427–430.

[16] Wray N., Goddard M.E., Increasing long-term response to selection, Genet. Sel. Evol. 26 (1994) 431–451.