The use of genomic coancestry matrices in the optimization of contributions for maintaining diversity at specific regions of the genome

Proceedings, 10th World Congress of Genetics Applied to Livestock Production

The use of genomic coancestry matrices in the optimization of contributions for maintaining diversity at specific regions of the genome

F. Gómez-Romano[1], B. Villanueva[1], J. Fernández[1], J. A. Woolliams[2] and R. Pong-Wong[2]
[1]Departamento de Mejora Genética Animal, INIA, Madrid, Spain,
[2]The Roslin Institute and the R(D)SVS, University of Edinburgh, UK

ABSTRACT: Optimal contribution methods have proved to be very efficient for controlling the rate at which coancestry and inbreeding increase and therefore for maintaining genetic diversity. These methods have usually used pedigree information for estimating expected genetic relationships between animals. However, the large amount of genomic information now available (dense chips containing thousand of SNP markers), provides us with a good opportunity of obtaining more accurate estimates of relationships. Genomic information also permits us now to target specific regions in the genome where there is an interest in maximising diversity. Using a semidefinite programming optimisation approach, we have investigated the effectiveness of using genomic coancestry matrices for controlling the loss of genetic variability in specific genomic regions while restricting the overall loss in the rest of the genome. The results show that genomic management was very successful for avoiding loss of diversity at specific genomic regions (even increased diversity). This management was also successful in restricting the loss of diversity in the remaining genome although the realised rate of coancestry resulted higher than the restriction imposed. There is thus a need of refining the theory of genetic contributions when realised genomic matrices are used.
Keywords: Genetic diversity; Inbreeding; Coancestry; Optimal contributions

Introduction

It is generally accepted that controlling the rate of coancestry provides a general framework for managing genetic variability. Optimal Contribution (OC) methods (Meuwissen (1997); Grundy et al. (1998)) permit to determine the optimal number of offspring that each breeding candidate should have to minimise coancestry. These methods were initially developed assuming the pedigree-based relationship matrix (**A**) which does not take into account variation due to Mendelian sampling. Thus, **A** represents expected relationships assuming neutrality. Although its use has proved to be efficient to manage diversity it has some limitations. For instance, individuals from the same (full-sib) families would inherit different set of alleles but they are assumed to be equally related. Additionally, since **A** does not consider variation between genomic regions, the optimisation of contributions would, in average, control the rate of coancestry to the chosen value, but some genomic regions may have substantially higher rates than those desired by the breeder. For example, regions harbouring QTL affecting a trait under selection will have higher rates of inbreeding than those expected from neutral regions (Roughsedge et al. (2008)). Thus, controlling loss of genetic diversity using pedigree information in a population under selection will, undoubtable, result in much higher loss of variation in regions under selection than the rest of genome. Relationship matrices calculated from markers at particular genomic regions can be used to set specific constraints for variability loss. In this way, diversity could be managed independently providing losses in diversity are not higher than a given predetermined rate for each genomic region.

The objective of this study was to assess the effectiveness of using dense SNP panels in controlling the loss of genetic variability in specific genomic regions while restricting the overall loss in the rest of the genome.

Materials and Methods

Optimisation of contributions. Let assume a set of N breeding candidates at a given time t. The pedigree-based genetic relationships between candidates are given in matrix **A**. Now, let **c** be the vector of genetic contributions of the candidates to the next (offspring) generation. When the main objective is to minimize the loss of genetic diversity, the OC problem can be formulated as to optimise **c** for minimising $\mathbf{c'Ac}/2$. In order to keep the solution for **c** within the valid range, the optimisation problem should also include the constraints that the sum of contributions of males and the sum of contributions of females are 0.5 each and that individual contributions are in the range [0,0.5]. The incorporation of realised genomic matrices (denoted here as **G**) calculated using SNP genotypes can be done with different levels of complexity. The simplest implementation is to calculate the average **G** matrix across the whole genome and use it to replace **A** in the optimisation problem described above. A more refined extension of the optimisation is the use of the relationship matrices specific for (some) genomic regions so their diversity can be managed independently. When minimising the loss of diversity at specific genomic regions, the OC problem can be formulated as to optimise **c** for minimising $\mathbf{c'G}_x\mathbf{c}/2$ subject to the restriction $\mathbf{c'G}_i\mathbf{c}/2 \leq f_i^*$, $i = 1, \ldots, k$, where x identifies a particular region in the genome, $\mathbf{G}_x$ represents the molecular relationship matrix calculated using only the markers from the target region where the loss of coancestry is to be minimised and $\mathbf{G}_i$ is the relationship matrix calculated with all the markers not located in that region. The optimisation method followed the semidefinite programming approach as described by Pong-

Wong and Woolliams (2007) due to its flexibility for adding extra constraints.

Simulations and genetic model. The study considered a population of N animals (20 or 100) per generation. The sex of the individuals was randomly assigned at each generation but ensuring that half were males and half were females. The population was managed for 10 generations. The number of individuals per generation was kept constant. The genetic model assumed the genome divided into 20 chromosomes of one Morgan each. Each chromosome had 2,000 biallelic loci equally spaced. The genotypes of 1,000 of them (at alternate positions) were assumed to be known and used to calculate the genomic matrices, simulating SNP markers. The remaining 1,000 loci were used to assess the performance of the different managing strategies. The simulation of the base population was described in detail in Gomez-Romano et al. (2013). In brief, a historical population of size N is allowed to reproduce for 5,000 generations with random mating. The last generation of this process was considered to be the base population of this study. The historical population was initialised assuming that the 40,000 loci were fixed. Mutations were allowed and being retained or lost due to genetic drift creating a distinctive linkage disequilibrium and gene frequency pattern. It should be noted that there were no mutations when creating the generations where management took place (i.e., after creating the base population).

Estimation of genomic matrices. The calculation of the genomic relationship was based on the allelic relationship method proposed by Nejati-Javaremi et al. (1997). At a given SNP the allelic relationship between two individuals is $0.5\sum_{i=1}^{2}\sum_{j=1}^{2}\delta_{ij}$, where δ_{ij} is the allele sharing status being equal to 1 if allele i from the first individual is the same to allele j from the second individual and 0 otherwise. The estimated genomic relationship between two individuals is the average value across all SNP in the genome (for the whole genome matrix) or in the region of interest (for a regional genomic matrix).

Scenarios compared. Four different management scenarios were considered: i) minimising pedigree coancestry (f_p); ii) minimising overall molecular coancestry (f_{m_ov}); iii) minimising coancestry in chromosome 1 (f_{m_c1}); and iv) minimising coancestry in chromosome 1 but restricting the coancestry rate in the rest of the genome ($\Delta f_{m\_ov\text{-}c1}$) to be ≤ 1%. An extra scenario where contributions were randomly assigned was also considered. The restriction for scenario iv) was calculated as $f^{*}_{t+1} = 1 - (1 - \Delta f)(1 - f_t)$, where f_t is the average coancestry in the candidates' population using all the markers except those in the chromosome 1 and Δf is the preselected target rate of inbreeding (1%).

Criteria of comparison. The main criteria of comparison were the average coancestry (f) and its rate (Δf) which represents the amount of genetic diversity that has been lost at a given time. Three estimates of coancestry were calculated at each generation: (i) the pedigree-based coancestry calculated using **A**; (ii) the overall realised molecular coancestry calculated with the overall **G** matrix; and (iii) the molecular coancestry at chromosome 1 representing the diversity remaining in such chromosome. The average coancestry at a given generation is half therelationship matrix of all individuals from that generation.

In order to make the results comparable, the average molecular coancestry was adjusted so the value in the base population was the same as the value calculated using **A**. Since animals in the base population are assumed to be (pedigree) unrelated, their **A** is the identity matrix so their average coancestry is $1/2N$. Hence, the adjusted average coancestry for generation t was equal to $(f_t - f_0)/(1 - f_0) + 1/2N$, where f_t is the unadjusted average coancestry at generation t.

Results and Discussion

Table 1 shows the average rates of pedigree and molecular coancestries for the four scenarios simulated in populations with N = 20 and 100. The scheme minimising the rate of molecular coancestry (Min Δf_{m_ov}) was clearly more efficient in maintaining diversity (i.e., it gave substantially lower Δf_{m_ov}) than that minimising the rate of pedigree coancestry (Min Δf_p). For instance, with N = 100, Δf_{m_ov} ranged from 0.24 to 0.30% when minimising Δf_p and from −0.33% to 0.06% when minimising Δf_{m_ov}. Hence, substantial improvements in retaining true realised genetic diversity can be achieved by using the genomic matrix rather than the pedigree-based **A**. The estimated Δf_p was very similar to Δf_{m_ov} when the contributions were assigned wat random or when they were optimised using pedigree information, suggesting that the use of pedigree coancestry as an indicator of the true realised molecular coancestry is only justified under this condition. The other schemes resulted in the pedigree coancestry not being a reliable estimator of the true molecular coancestry.

The optimisation method was very successful in reducing the rate of increase of their target coancestries (see also f_{m_c1} in Table 2) but they were less efficient in reducing other coancestries. The relatively high $\Delta f_{m\_ov\text{-}c1}$ obtained under scenario iii) highlight the need of making the optimisation more complex by including a restriction on the rate of coancestry in the rest of the genome. Table 2 compares results of scenarios where coancestry was minimised in chromosome 1 with and without a restriction on $\Delta f_{m\_ov\text{-}c1}$ ($\Delta f_{m\_ov\text{-}c1} \leq 1\%$). The scheme aiming to avoid loss of diversity at chromosome 1 was very successful in that it even increased genetic diversity in this chromosome (i.e. negative Δf_{m_c1}). After 10 generations, the diversity in chromosome 1 increased by 3% with N = 20 and about 8% with N = 100. However, the success in retaining diversity in this chromosome comes at a price in the remaining genome, where the rate of increase in coancestry was substantially larger than the other scenarios (even larger than the scenario of random selection). These results clearly suggest that when the main objective is to retain diversity at a specific region of the

genome, a restriction should also be added to avoid excessive loss of diversity of other areas of the genome. The optimisation performed imposing such restriction (i.e., scenario iv) has some success in reducing the loss of diversity in these other areas of the genome. Note that $\Delta f_{m\_ov\text{-}c1}$ was lower with than without the restriction. However, the realised Δf was higher than 1% (across the 10 generations it averaged 2.43% and 1.12% for $N = 20$ and $N = 100$, respectively).

Table 1. Rates of pedigree (Δf_p) and overall molecular (Δf_{m_ov}) coancestry across generations (t) when contributions are selected at random (Rand) or optimized for minimising f_p (Min Δf_p), f_{m_ov} (Min Δf_{m_ov}) or f_{m_c1} (Min Δf_{m_c1}).

	Δf_p (%)				Δf_{m_ov} (%)			
t	Rand	Min Δf_p	Min Δf_{m_ov}	Min Δf_{m_c1}	Rand	Min Δf_p	Min Δf_{m_ov}	Min Δf_{m_c1}
				$N = 20$				
1	2.44	1.28	2.52	5.10	2.69	1.33	0.25	3.98
2	2.64	1.30	1.77	3.66	2.99	1.34	0.93	2.95
3	2.17	1.31	1.76	3.23	2.28	1.18	0.94	3.46
4	2.81	1.30	1.73	3.03	2.95	1.30	1.14	2.67
5	2.76	1.30	1.78	3.40	3.40	1.25	0.85	3.11
10	2.26	1.30	1.78	3.14	2.21	1.42	0.97	2.59
				$N = 100$				
1	0.50	0.30	0.70	2.21	0.51	0.24	-0.33	1.86
2	0.48	0.20	0.52	1.36	0.51	0.30	-0.06	1.07
3	0.47	0.25	0.47	1.22	0.45	0.27	-0.03	1.05
4	0.50	0.25	0.47	1.44	0.52	0.24	0.00	1.40
5	0.53	0.25	0.48	1.35	0.52	0.27	0.03	1.17
10	0.50	0.26	0.45	0.97	0.53	0.25	0.06	0.87

Table 2. Rate of coancestry at chromosome 1 (Δf_{m_c1}) and at the whole genome except chromosome 1 ($\Delta f_{m\_ov\text{-}c1}$) across generations ($t$) when contributions are optimized for minimising coancestry at chromosome 1 restricting or not $\Delta f_{m\_ov\text{-}c1}$ with N = 20.

	No restriction		With restriction	
t	Δf_{m_c1}	$\Delta f_{m\_ov\text{-}c1}$	Δf_{m_c1}	$\Delta f_{m\_ov\text{-}c1}$
1	−3.83	3.98	−3.14	2.54
2	−0.70	2.95	−1.18	2.53
3	−1.00	3.46	−0.50	2.41
4	0.07	2.67	−0.17	1.86
5	−0.69	3.11	−0.56	2.55
10	0.10	3.98	−0.20	2.38

Surprisingly, the solutions from the optimisation were valid in the sense that those found as optimum did, indeed, fulfil the restriction that the expected $\Delta f_{m\_ov\text{-}c1}$ should not be greater than 1%. Figure 1 shows the expected $\Delta f_{m\_ov\text{-}c1}$ given the assigned contributions to candidates and the realised value observed in the offspring generation. Across the 10 generations of management, the optimisation procedure yielded solutions with expected $\Delta f_{m\_ov\text{-}c1}$ lower than the imposed 1% restriction. However, the realised $\Delta f_{m\_ov\text{-}c1}$ observed in the offspring was consistently higher than the value aimed during the optimisation.

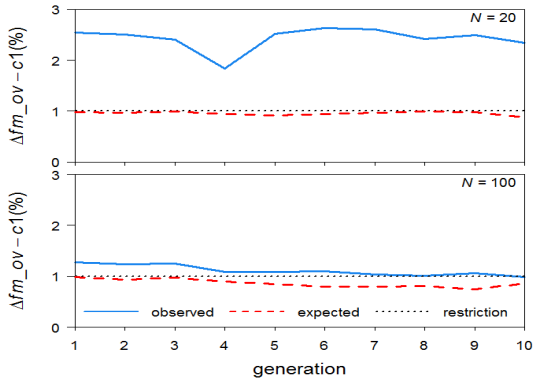

Figure 1: Expected and observed $\Delta f_{m\_ov\text{-}c1}$ when the optimisation was performed for minimising coancestry in chromosome 1 while restricting the rate of coancestry at the rest of the genome to 1%.

Conclusion

This study shows that the use of molecular coancestry in the optimisation is substantially more efficient in retaining genetic diversity than the use of pedigree coancestry. Moreover, the use of molecular coancestry permits us to target specific genomic regions for minimising the loss of diversity and extend the optimisation procedure to include restrictions in regions different to those where the loss of diversity is minimized. Under this scenario, however, the realised Δf did not fullfil the restriction that was imposed. This clearly highlight the need of refining the theory of genetic contributions using realised genomic matrices in order to assure that constraints on Δf are properly included in the model.

Acknowledgements

This work was funded by the MINECO and INIA, Spain (CGL2012-39861-C02-02). The Roslin Institute receives financial support from the BBSRC. We also thank funding from the QUANTOMICS EU Project (222664-2).

Literature Cited

Gómez-Romano, F., Villanueva, B., de Cara, M. A. R. et al. (2013). *Genet. Sel. Evol.* 45:38-45.

Grundy, B., Villanueva, B., and Woolliams, J. A. (1998). *Genet. Res.* 72: 159-168.

Meuwissen, T. H. E. (1997). *J. Anim. Sci.* 75: 934-940.

Nejati-Javaremi, A., Smith, C. and Gibson, J. P. (1997). *J. Anim. Sci.* 75: 1738-1745.

Pong-Wong, R. and Woolliams, J. A. (2007). *Genet. Sel. Evol.* 39**:** 3-25.

Roughsedge, T., Pong-Wong, R., Woolliams, J. A. et al. (2008). *Genet. Res.* 90: 199-208.