

**EVALUATION OF DIFFERENT GENOMIC COANCESTRY MATRICES TO  
MAINTAIN GENETIC VARIABILITY IN A TURBOT SELECTED POPULATION**  
Morales-González<sup>1,2\*</sup>, E., Saura<sup>1</sup>, M., Fernández<sup>1</sup>, A., Fernández<sup>1</sup>, J., Cabaleiro<sup>3</sup>, S.,  
Martínez<sup>4</sup>, P., Villanueva<sup>1</sup>, B.  
<sup>1</sup>INIA, Spain., <sup>2</sup>Universidad Politécnica de Valencia, Spain, <sup>3</sup>CETGA, Spain, <sup>4</sup>Universidade de  
Santiago de Compostela, Spain  
\*E-mail: elisa.em89@gmail.com

### INTRODUCTION

In the last decade there has been a considerable effort to increase the production of turbot (*Scophthalmus maximus* L.) through selective breeding programmes, particularly in Spain which is the leading European producer. Such programmes have a great potential, particularly due to the high fecundity of fish that allows to apply high selection intensities and consequently to obtain high genetic gains. However, this can also lead to high rates of inbreeding, with the consequent reduction of genetic variability and increased risk of inbreeding depression. Thus, the control of inbreeding and the loss of genetic variability in turbot selection programmes is fundamental to ensure their sustainability. However, to date there are no studies evaluating inbreeding in commercial populations of this species.

The most efficient method to control the loss of genetic variability and the increase of inbreeding is the optimal contributions method (OC). This method is based on optimising the contributions of candidates to minimise group coancestry of the selected breeders and thus minimise inbreeding in the next generation (Woolliams et al., 2015). Therefore, the central element of OC is the coancestry matrix.

Traditionally, the coancestry matrix has been computed from pedigree data and therefore gives only expectations of the proportion of genes that two particular individuals have in common. Genome-wide data can be used to estimate these proportions with a high degree of precision and coancestry coefficients can be re-calculated to reflect the true proportion of the genome in common. In fact, using genome-wide coefficients in OC has been proved to lead to higher diversity maintained than pedigree-based coefficients (de Cara et al., 2011; Gómez-Romano et al., 2013).

Several measures of genomic coancestry (and inbreeding) have been developed but their relative efficiency for maintaining genetic variability is unknown. The aims of this study were to i) obtain estimates of genome-wide coancestry coefficients in a turbot commercial population; and ii) evaluate the efficiency of different genome-wide coancestry matrices in retaining genetic variability when using OC.

### MATERIAL AND METHODS

Data available for this study came from an experiment developed within the framework of the European project FISHBOOST (<http://www.fishboost.eu/>). Genome-wide SNP data were available for 1,391 offspring belonging to 36 full-sib families (including 12 paternal and 11 maternal half-sibs) and for their parents (23 sires and 23 dams). Genotypes for 18,097 SNPs were obtained by genotyping-by-sequencing using a 2b-RAD-sequencing approach. Parents were sampled from a population maintained at CETGA (Aquaculture Cluster of Galicia, Spain) that is representative of the main turbot breeding programmes in Europe that have the same Atlantic origin.

Five different genome-wide coancestry matrices were compared: i)  $\theta_{IBS}$ , based on the proportion of alleles shared by two individuals (equivalent to a similarity index); ii)  $\theta_{L\&H}$ , based on the excess of SNP homozygosity following Li and Horvitz (1953); iii)  $\theta_{VR}$ , based on the genomic relationship matrix of VanRaden (VanRaden, 2008); iv)  $\theta_{YANG}$ , based on the genomic relationship matrix of Yang (2010) that differs from  $\theta_{VR}$  in the diagonal elements; and v)  $\theta_{SEG}$ , based on the proportion of identity by descent (IBD) segments shared by two individuals. Note that  $\theta_{IBS}$  reflects identity by state (IBS), and  $\theta_{L\&H}$ ,  $\theta_{VR}$  and  $\theta_{YANG}$  are standardized by the allele frequencies of a reference population (in this case, the parental population) in an attempt to

move from IBS to IBD. While  $\theta_{IBS}$  and  $\theta_{SEG}$  range between 0 and 1,  $\theta_{L\&H}$ ,  $\theta_{VR}$  and  $\theta_{YANG}$  can take negative values.

The OC method was used for maximising genetic variability. The problem to be solved in OC is concerned with the allocation of contributions of the candidates to produce the next generation so as to minimise the global coancestry, and can be formulated as to minimise  $\mathbf{c}^T\boldsymbol{\theta}\mathbf{c}$ , where  $\mathbf{c}$  is the vector of solutions (i.e. contributions or proportions of total offspring left by each candidate) and  $\boldsymbol{\theta}$  is the coancestry matrix. Restrictions were imposed to ensure that half of the contributions come from males and half come from females and that  $c_i \geq 0$ , for  $i = 1, \dots, n$  candidates. The different coancestry matrices described above were used in the optimisation that was solved using a simulated annealing algorithm (Kirkpatrick et al., 1983). The amount of genetic variability retained after the optimisation was measured as the percentage of fixed loci and the expected homozygosity in the selected candidates. Here, we considered as potential candidates the offspring for which sex was available (i.e. 1,152 out of the 1,391 genotyped offspring).

## RESULTS AND DISCUSSION

The different coancestry coefficients were compared for four groups of different degrees of relationships including parent-offspring, full-sibs, half-sibs and less related individuals (Table 1). Estimates of  $f_{IBS}$  were much higher than estimates of the other coefficients as expected given that  $f_{IBS}$  reflects IBS while the rest try to reflect IBD. Also, the magnitude of  $f_{SEG}$  was higher than that of  $f_{L\&H}$ ,  $f_{VR}$  and  $f_{YANG}$ . Although the magnitude of the different coancestry coefficients differed greatly, correlations between them were high ( $\geq 0.8$ ). All coefficients had the ability to discriminate different degree of relationships, and  $f_{VR}$  and  $f_{YANG}$  were those that more approximated to the theoretically expected values derived from pedigree relationships (e.g., 0.25 for parents-offspring and full-sibs and 0.125 for half-sibs).

**Table 1.** Average (and standard deviation) of the different coancestry coefficients for different degree of relationships and number of values averaged ( $N$ ).

	$N$	$f_{IBS}$	$f_{L\&H}$	$f_{VR} = f_{YANG}$	$f_{SEG}$
All	1,035,770	0.777 (0.018)	0.004 (0.081)	0.002 (0.064)	0.134 (0.064)
Parents-offspring	2,783	0.828 (0.012)	0.231 (0.055)	0.238 (0.136)	0.325 (0.042)
Full-sibs	26,296	0.829 (0.013)	0.235 (0.059)	0.241 (0.106)	0.328 (0.047)
Half-sibs	46,762	0.799 (0.014)	0.104 (0.063)	0.115 (0.080)	0.219 (0.048)
Less-related	959,929	0.774 (0.015)	-0.008 (0.067)	-0.011 (0.033)	0.124 (0.051)

Coefficients  $f_{L\&H}$ ,  $f_{VR}$  and  $f_{YANG}$  were developed for correcting for the fact that alleles shared among individuals were already present in the base population, which is assumed to be composed by unrelated and non-inbred individuals. Thus, the allele frequencies used for computing these coefficients should be from that base population. However, these frequencies are usually unknown and the current population from which the sample of individuals is taken is frequently used as the reference. Here, for these coefficients the reference population was the parental generation and their possible negative values can be better interpreted in terms of Wright's (1921) original correlation concept of relatedness than in terms of Malecot's (1948) probability of IBD (Wang, 2014). A negative value also has a biological meaning as it signifies that the probability of two homologous alleles from two individuals being IBD is smaller than that of two homologous alleles drawn at random from the population. In contrast,  $f_{IBS}$  reflects relationships that are caused by common ancestors going back to a very distant theoretical base population in which all alleles were unique. Coefficient  $f_{SEG}$  may be used to distinguish recent from distant relatedness, given that, as a simplification, the length of a segment (in cM) equals  $100/2g$ , where  $g$  is the number of ancestral generations to the common ancestor from which the segment was derived (Thompson, 2013; Purfield et al., 2017). Here,

the minimum length of a segment chosen was 0.4 Mb and thus it is expected that segment of this length come from common ancestors born 50 generations ago when assuming a recombination rate of 2.5 cM/Mb (Bouza et al., 2007).

The genetic variability retained in the selected candidates when using the different coancestry matrices in OC was similar in terms of the expected heterozygosity ( $1-EH$ ) and also the percentage of alleles fixed (<1%) (Table 2). However, the different matrices achieved similar variability by selecting very different number of individuals. The less fish selected the higher was the number of SNPs fixed. The reduced number of fish selected could have subsequent negative consequences. In particular, with  $\theta_{L\&H}$  and  $\theta_{SEG}$  only 9% and 13% of the initial number of candidates were selected, respectively, in comparison with 47-85% of candidates selected with the other matrices. These differences can be explained because  $\theta_{L\&H}$  and  $\theta_{SEG}$  presented higher variances for coancestry coefficients estimated for the group of 'less related individuals' (Table 1). Conversely,  $\theta_{IBS}$ , which selected the highest number of candidates, was also the matrix with lower ability to differentiate relationships (Table 1).

**Table 2.** Number of fish selected to contribute ( $N_{sel}$ ), mean contribution and corresponding standard deviation (SD), minimum and maximum contributions, and number (percentage) of SNPs fixed and expected homozygosity (EH) of the candidates selected after implementing OC using different coancestry matrices.

	$N_{sel}$	Mean	SD	Min	Max	SNPs fixed	EH
$\theta_{IBS}$	976	0.001	0.002	0.000	0.017	0 (0.00%)	0.775
$\theta_{L\&H}$	94	0.011	0.008	0.000	0.042	176 (0.97%)	0.764
$\theta_{VR}$	636	0.002	0.001	0.000	0.007	1 (0.00%)	0.774
$\theta_{YANG}$	544	0.002	0.003	0.000	0.027	45 (0.25%)	0.787
$\theta_{SEG}$	152	0.007	0.006	0.000	0.030	56 (0.31%)	0.766

## REFERENCES

- Bouza, C., Hermida, M., Pardo, B.G. *et al.* 2007. *Genetics* 177: 2457–2467
- de Cara, M.A.R., Fernández, J., Toro, M.A., Villanueva, B. 2011. *J. Anim. Breed. Genet.* 128: 456–464.
- Gómez-Romano, F., Villanueva, B., de Cara, M.A.R., Fernández, J. 2013. *Genet. Sel. Evol.* 45: 1–8
- Kirkpatrick, S., Gelatt, C.D., Vecchi, M.P. 1983. *Science* 220: 671–680
- Li, C., Horvitz, D. 1953. *Am. J. Hum. Genet.* 5: 107–117
- Malécot, G. 1948. *The mathematics of heredity.* Masson & Cie, Paris
- Purfield, D.C., McParland, S., Wall, E. & Berry, D.P. 2017. *Plos One.* 12: 1–23
- Thompson, E. 2013. *Genetics.* 194: 301–326.
- VanRaden, P.M. 2008. *J. Dairy Sci.* 91: 4414–4423
- Wang, J. 2014. *J. Evol. Biol.* 27: 518–530
- Woolliams, J., Berg, P., Dagnachew, B.S., Meuwissen, T.H.E. 2015. *J. Anim. Breed. Genet.* 132: 89–99
- Wright, S. 1921. *Genetics* 6: 124
- Yang, J.A., Benyamin, B., McEvoy, B.P., *et al.* 2010. *Nat. Genet.* 42: 565–569.

**Acknowledgements:** The research leading to these results has received funding from the European Union's Seventh Framework Programme (KBBE.2013.1.2-10) under grant agreement n° 613611.

## EVALUATION OF DIFFERENT GENOMIC COANCESTRY MATRICES TO MAINTAIN GENETIC VARIABILITY IN A TURBOT SELECTED POPULATION

**ABSTRACT:** Due to their high fertility, aquaculture species are particularly susceptible to the accumulation of inbreeding since few parents can be used to create a whole generation. However, studies evaluating inbreeding in farm fish populations are very scarce. The most efficient method to control inbreeding is the optimal contributions method (OC) for which the central element is the coancestry matrix. In this study, we make use of ~18,000 SNP genotypes obtained by a 2b-RAD-sequencing approach to estimate genome-wide coancestry coefficients in a turbot farm population, and to evaluate the efficiency of different genome-wide coancestry matrices in retaining genetic variability when using OC. Coancestry matrices evaluated included those based on the i) proportion of shared alleles; ii) deviation of the expected homozygosity; iii) the genomic relationship matrix of VanRaden; iv) the genomic relationship matrix of Yang; and v) proportion of shared segments. Our results showed that the different genome-wide coancestry matrices presented a similar efficiency in retaining genetic variability when used in OC. However, important differences in the number of individuals selected in the optimisation were observed.

**Keywords:** coancestry, inbreeding, optimal contributions, turbot.