*Short Communication*

# Structural assessment of backcrossing using microsatellite markers

**N.O.M. Tshipuliso[1,2], L.J. Alexander[3], A. Kotze[2,4], K. Ehlers[2], V.L. Reisenauer Leesburg[3] and M.D. MacNeil[3#]**

[1] Agricultural Research Council, Private Bag X2, Irene 0062, South Africa
[2] Genetics Department, University of the Free State, Bloemfontein 9300; South Africa
[3] USDA Agricultural Research Service, Miles City, Montana 59301, USA
[4] National Zoological Gardens of South Africa, Pretoria 0001, South Africa

## Abstract

Backcrossing, coupled with marker or gene assisted selection, can be used to introgress a specific gene or chromosomal region from one population into another. The objective of this study was to assess the genomic structure of cattle produced by backcrossing for loci that are unlinked to a locus that was being introgressed. Genotypes of the two parental populations, their $F_1$ progeny, and two subsequent backcross generations of animals were determined at 34 microsatellite loci that were not linked to a locus being introgressed. There was little evidence to suggest any systematic genome-wide departure from pedigree derived expectation as a result of the breeding system. These data validate the desired intention of a backcrossing program that progressive generations migrate genotypically toward one of the parental type.

**Keywords:** Backcross, breeding system, introgression, genetic markers
[#] Corresponding author. Email: mike.macneil@ars.usda.gov

Backcrossing is a well-known and long established breeding method where a characteristic is introgressed from a donor parent into the genomic background of a recurrent parent. Because backcrossing can isolate a gene or chromosomal region in a different genetic background, it helps to discern genetic architecture of quantitative traits (Hospital, 2005). It is also one of the few reliable methods to validate the additive effect of a QTL or candidate gene after it is putatively detected. It is usually desired that all loci, except the one being introgressed migrate to the original parental genotype. Thus, the objective of this study was to assess the genomic structure of cattle produced by backcrossing for loci that are unlinked to a locus that was being introgressed.

Thirteen Line 1 Hereford (L1) bulls and 91 composite gene combination (CGC) heifers were joined to initiate a backcrossing experiment in 1999. The intent was to introgress a locus on BTA6 flanked by microsatellite markers *BM4621* and *BM415* from CGC into the L1 background. Line 1 Herefords have been maintained as an inbred line at the Fort Keogh Livestock and Range Research Laboratory (LARRL) in Miles City, Montana, USA since 1934 (MacNeil *et al.*, 1992). The CGC population was started in 1979 at LARRL in order to develop a line of cattle uniquely suited to their environment, specifically the Northern Great Plains (Newman *et al.*, 1993). Breed composition of CGC is ½ Red Angus, ¼ Charolais, and ¼ Tarentaise. Both L1 and CGC have been maintained as closed populations since their founding. Initial $F_1$ heifers were joined with L1 bulls to produce a first backcross generation ($B_1$) and $B_1$ heifers were joined with L1 bulls to produce a second backcross generation ($B_2$).

Blood samples were collected from the parental, $F_1$, $B_1$, and $B_2$ animals for DNA analysis. Standard DNA extraction protocols were used (e.g. Ausubel *et al*., 1994). A set of 34 microsatellite markers covering the 29 bovine autosomes, but unlinked with each other and the region of interest on BTA6 were chosen from http://www.marc.usda.gov/genome/cattle/cattle.html based on marker position, suitability for multiplex reactions and ease of scoring. The chosen markers were: *AGLA227*, *BL1029*, *BL1038*, *BM719*, *BM8126*, *BM827*, *BMC1013*, *BMC5227*, *BMS1247*, *BMS1282*, *BMS1315*, *BMS1316*, *BMS1967*, *BMS2177*, *BMS2533*, *BMS2614*, *BMS468*, *BMS510*, *BMS574*, *BMS713*, *BMS745*, *BMS836*, *BP28*, *CSSM036*, *CSSM038*, *IDVGA-2*, *IDVGA-45*, *ILSTS023*, *ILSTS028*, *ILSTS059*, *INRA063*, *RM044*, *RM321*, and *URB014*. Touchdown PCR was performed in MJ Research (Waltham, MA, USA) thermocyclers as described in

http://biosupport.licor.com/docs/whatsnew/4300_ApMan_07999.pdf. All genotypes were collected on a LiCor 4200 DNA Analysis System (Lincoln, NE, USA). Genotypes were then independently scored by two individuals, compared, and discrepancies then resolved. Unresolved discrepancies resulted in samples being run through a second PCR and rescored.

All genotypes were also assessed relative to the purported pedigree using GenoProb (Thallman *et al*., 2001a; b). GenoProb is a computer program that analyzes genetic marker data in complex pedigrees with missing marker data using an iterative allelic peeling algorithm. Pedigree information for each individual animal and marker locus data were required (Thallman *et al*., 2001a; b). Pedigree errors were resolved based on reanalyzed genotypes and when this was not possible, animals with suspect parentage and their descendants were dropped from the study.

A model based on Bayesian clustering algorithm called **Structure,** version 2.2 (Pritchard & Wen, 2004) was used to quantify membership in populations using the genotype data. The L1 and CGC animals were assigned *a priori* to their respective populations. The probabilities of each animal then belonging to the K=2 foundation populations were calculated, with membership coefficients summing to 1 across the clusters (Pritchard *et al*., 2000). Results were interpreted as estimating the genomic similarity of $F_1$, $B_1$, and $B_2$ animals to the parental populations.

The estimated probability of genomic similarity of each animal, represented by a bar, to L1 is presented graphically in Figure 1. Sets of animals correspond to each experimental generation as indicated on the x-axis. The probability of genomic similarity to CGC is 1.0 minus the probability of genomic similarity to L1. The probabilistic assignment of parental animals to their respective populations was clearly consistent with the *a priori* expectations (i.e., probability of L1 animals being assigned to L1 ≈ 1.0 and probability of CGC animals being assigned to L1 ≈ 0.0).

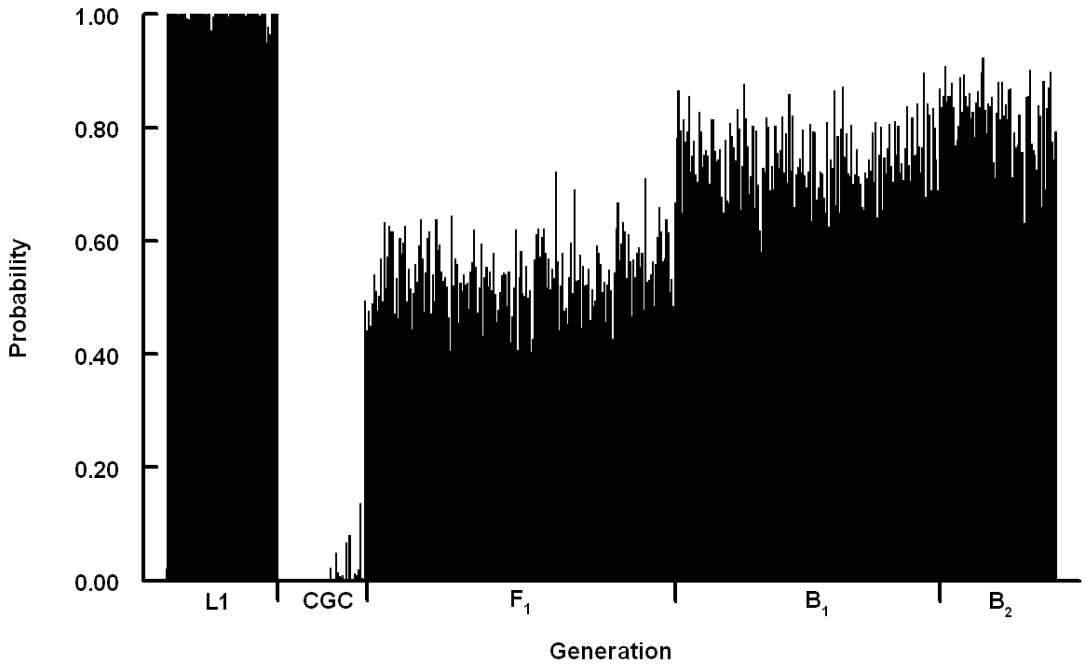


**Figure 1** Posterior probability of each animal being a member of the Line 1 Hereford population, categorized by generation (L1 = Line 1 Hereford, CGC, L1xCGC = $F_1$, L1x$F_1$ = $B_1$, L1x$B_1$ = $B_2$).

The individual probabilities of genomic similarity to L1 are then summarized by generation and compared to their expectation based on the breeding system in Table 1. It is plausible that backcross generations could differ from pedigree based expectation due to heterozygote advantage or greater genetic fitness of one population than another. Here, there is little evidence to suggest any systematic genome-wide

**Table 1** Numbers of animals in successive generations of backcrossing ($B_1$ and $B_2$) an $F_1$ population formed from Line 1 Hereford (L1) and composite gene combination (CGC) parental stocks to L1, and means (± s.e.) and their expectations of the genomic probability of each animal being a member of the L1 population

| | Generation | | | | |
|---|---|---|---|---|---|
| Item | L1 | CGC | $F_1$ | $B_1$ | $B_2$ |
| Number of animals | 64 | 50 | 179 | 153 | 69 |
| Expectation | 1.000 | 0.000 | 0.500 | 0.750 | 0.875 |
| Mean | 0.999 ± 0.006 | 0.002 ± 0.011 | 0.540 ± 0.018 | 0.746 ± 0.020 | 0.819 ± 0.029 |

departure from the pedigree derived expectations. However, microsatellites are generally believed to have neutral alleles and thus not acted upon directly by selection. Note, there was greater variation in probabilistic assignment to population in the $F_1$, $B_1$, and $B_2$ generations than among parental animals.

These data validate the desired intention of a backcrossing program that progressive generations migrate genotypically toward one of the parental types. However, variation among individuals within generations in the proportional assignment to founder populations is noteworthy.

**Acknowledgements**

The authors thank L. French, S. Alexander, K. Neary and H. Stroh for excellent technical assistance. USDA-ARS is an equal opportunity/affirmative action employer. All agency services are available without discrimination. This research was conducted under cooperative agreements between USDA-ARS, Montana Agric. Exp. Stn., ARC, and was part of a MSc study at the University of the Free State. Mention of a proprietary product does not constitute a guarantee or warranty of the product by USDA, Montana Agric. Exp. Stn., ARC, or the authors and does not imply its approval to the exclusion of other products that also may be suitable.

## References

Ausubel, F., Brent, M.R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. & Struhl. K., 1994. In: Current Protocols in Molecular Biology 1. John Wiley & Sons, Inc. New York.

Hospital, F., 2005. Selection in backcross program. Phil. Trans. R. Soc. B. 360, 1503-1511.

MacNeil, M.D., Urick, J.J., Newman, S. & Knapp, B.W., 1992. Selection for postweaning growth in inbred Hereford cattle - The Fort-Keogh, Montana Line-1 example. J. Anim. Sci. 70, 723-733.

Newman, S., MacNeil, M.D., Reynolds, W.L., Knapp, B.W. & Urick, J.J., 1993. Fixed effects in the formation of a composite line of beef cattle.1. Experimental design and reproductive performance. J. Anim. Sci. 71, 2026-2032.

Pritchard, J.K. & Wen, W., 2004. Documentation for STRUCTURE software: Version 2.2. http://pritch.bsd.uchcago.edu/software/structure22/readme.pdf.

Pritchard, J.K., Stephens, M. & Donnelly, P., 2000. Inference of population structure using multilocus genotype data. Genetics 155, 945-959.

Thallman, R.M., Bennett, G.L., Keele, J.W. & Kappes, S.M., 2001a. Efficient computation of genotypes probabilities for loci with alleles. Allelic peeling. J. Anim. Sci. 79, 26-33.

Thallman, R.M., Bennett, G.L., Keele, J.W. & Kappes, S.M., 2001b. Efficient computation of genotype probalities for loci with many alleles. II. Iterative method for large, complex pedigrees. J. Anim. Sci. 79, 34-44.