THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

Short communication: Identification of variation in the ovine prolactin gene of Chios sheep with a cost-effective sequence-based typing assay

OPEN ACCESS

Short communication: Identification of variation in the ovine prolactin gene of Chios sheep with a cost-effective sequence-based typing assay

D. Miltiadou,*[1] M. Orford,* S. Symeou,* and G. Banos†
*Department of Agricultural Sciences, Biotechnology and Food Science, Cyprus University of Technology, PO Box 50329, Lemesos, Cyprus
†Scotland's Rural College and Roslin Institute, University of Edinburgh, Easter Bush, EH25 9RG, Scotland, UK

ABSTRACT

The present study identified single nucleotide polymorphisms (SNP) in the coding and untranslated regions of the ovine prolactin gene of Chios sheep. By developing a cost-effective direct sequence-based typing assay, around 600 bp of reliable sequencing data and clear identification of heterozygous positions was achieved. Five SNP were found, located in exons 2 (KC764410:g.567G > A, g.625C > T, g.683C > A) and 3 (KC764410:g.2015C > A, g.2101G > A), whereas the remaining exons were monomorphic. The identified SNP were synonymous, with the exception of the g.567G > A SNP, which results in an Arg to His amino acid change. As the sequencing cost of the sequence-based typing assay was 20 orders of magnitude lower compared with a standard Sanger method, the assay was also used as a genotyping tool. The identified polymorphism was genotyped for 247 ewes and was subsequently used in mixed model association analyses of milk yield, milk fat content, and litter size at birth. The association analysis revealed a significant dominance effect of 0.17 ± 0.07 of the g.2015C > A SNP on milk fat percentage, whereas a dominance effect of -21.33 ± 10.51 of the same SNP on total lactation milk yield was also estimated. The g.2015C > A SNP explained 2.47 and 3.68% of the total phenotypic variance of milk yield and milk fat percentage, respectively, whereas the corresponding values for the animal variance were 7.14 and 11.75%. A suggestive association of the nonsynonymous g.567G > A SNP with litter size at birth was also detected.

Key words: sheep, prolactin association, SNP identification, sequence-based typing (SBT)

Short Communication

Prolactin (*PRL*) is essential for milk production in mammals, as its suppression inhibits lactation. In dairy ruminants, a good body of evidence has shown that *PRL* is galactopoietic and increases feed intake to provide the nutrients necessary to support lactation (Lacasse and Ollier, 2015). Therefore, *PRL* could be regarded as a functional candidate gene associated with milk production and composition.

Whereas SNP identified within the bovine *PRL* gene have been associated with milk production traits (Dong et al., 2013; Lü et al., 2010; Raven et al., 2014), the polymorphism of the ovine *PRL* gene has only received limited attention. Among the polymorphisms identified, 2 variants (A and B), differing by a 23-bp deletion within intron 2 (Orford et al., 2010), have been associated with milk traits (Ramos et al., 2009; Staiger et al., 2010). These *PRL* intron 2 genotypes significantly affected milk yield and fat and protein content in the Serra da Estrela breed (Ramos et al., 2009). Staiger et al. (2010) also suggested that the *PRL* intron 2 polymorphism significantly affected milk yield in East Friesian sheep, and could therefore be used as a potential marker in breeding programs. In addition, an SNP within the 5′ flanking region of the ovine *PRL* gene has been associated with litter size in Small Tail Han sheep (Chu et al., 2009).

The objectives of the present study were to (1) develop a cost effective sequence-based typing (**SBT**) assay to identify and genotype variations in all 5 exons of the *PRL* gene, including the regulatory untranslated regions, and (2) conduct a preliminary association analysis of the identified polymorphism in (1) with sheep traits.

Genomic DNA was isolated from 247 randomly selected Chios sheep from a commercial farm in Cyprus, using the Genomic DNA Blood kit (Macherey-Nagel, Düren, Germany) according to the manufacturers' instructions. For SNP identification and genotyping of *PRL* exons, PCR primers were designed based on the sheep genome sequence Oar v3.0 (https://isgcdata.agresearch.co.nz/), using the program Primer3 (http://frodo.wi.mit.edu/primer3/), and PCR reactions were set up in 25-μL volumes containing 25 ng of genomic DNA, 0.5 U of Taq DNA polymerase (Qiagen Inc., Valencia, CA), 0.5 μM each primer (Table 1), and 1.5

Table 1. PCR primers used to amplify prolactin (*PRL*) exons

Gene fragment[1]	Forward primer	Reverse primer	Size (bp)
5′UTR–exon 1-partial intron 1	GCCTTATAAAGCCAACATCTGG	TGATACCCCCATTGGAACAT	171
Partial intron 1-exon2-partial intron 2	ATGACAAACTCCTACAAGCTG	CCACATCTTATGAGCTAATGTCTTA	697
Partial intron2-exon3-partial intron3	GCCCAAACAACCCTAATGAA	CGTGAAGCCAGGTAACATCA	219
Partial intron3-exon4-partial intron4	TTTAATGAGATTGTTTCCTGTGG	TCATGAGAACAGCAAGGAAGA	250
Partial intron4-exon5–3′UTR	TCTTTCCTGTATCTTCCCCAAT	GAAACATTGACAAAATTGCCATC	468

[1]UTR = untranslated region.

mM MgCl$_2$ (2.5 mM MgCl$_2$ for *PRL* exon 2). After an initial 5-min denaturation step, the PCR reactions were subjected to 30 cycles at 94°C for 30 s, 56°C for 30 s, and 72°C for 30 s, followed by a final elongation step for 5 min at 72°C.

To reduce the sequencing cost, different cleaning procedures of the PCR products (agarose gel extraction and isopropanol precipitation) and different volumes of the fluorescent terminator mix were tested (8, 4, 2, 1, 0.5, and 0.25 µL) for direct sequencing of the amplified fragments. A comparative experiment was set up to ensure the accuracy and specificity of the sequencing results. Sequencing of all 5 exons of the *PRL* gene from 20 animals was performed with both a standard and the modified sequencing protocol for SBT, as described below. To ensure that the SBT assay works efficiently, a similar comparison experiment was performed for a second gene, the *ACAA2* gene, with primers and conditions described in Orford et al. (2012).

For standard sequencing, PCR products, generated with the primers shown in Table 1, were resolved on 2% agarose gels and the excised bands were gel purified using the Qiagen gel extraction kit (Qiagen Inc.). The purified templates were then subjected to cycle sequencing, using the Big Dye Terminator Reagent (Applied Biosystems, Foster City, CA) according to manufacturer's recommendations.

For the SBT assay, the same PCR products described above were generated in 96-well plates and precipitated by the addition of 5 µL of 3 M sodium acetate, pH 5.3, plus 25 µL of isopropanol. The reactions were then incubated at −70°C for 20 min and centrifuged at 4°C for 45 min at 2,720 × g. The supernatants were removed by inverting the plates on paper towels and briefly spinning them upside down for 2 s at 500 x g. The precipitated products were then washed by the addition of 100 µL of 70% ethanol without mixing, recentrifuged for 10 min at 2,720 × g at 4°C, and the supernatants were removed as before. Following a 5-min drying step at 50°C, 40 µL of deionized water was added to each well to redissolve the purified DNA.

Cycle sequencing was then performed in 10-µL reaction volumes using 2 µL of the purified templates, 0.5 µL of Big Dye Terminator Reagent v.3.1 (Applied Biosystems), 0.5 µL of the reverse primer (5 µM), 1.75 µL of 5× sequencing buffer, and 5.25 µL of distilled water. The sequencing reactions were subjected to an initial 2-min denaturation step at 96°C, 30 cycles at 96°C for 15 s, 50°C for 15 s, and 60°C for 4 min. After cycling, 10 µL of water was added, followed by 5 µL of 125 mM EDTA and 60 µL of 100% ethanol to precipitate the termination products. The plates were then incubated at room temperature for 15 min in the dark, centrifuged at 2,720 × g at 4°C for 45 min, and the supernatants were removed as before. Following a wash with 250 µL of 70% ethanol, the plates were dried for 5 min at 50°C and the termination products were dissolved in 10 µL of formamide before reading on an ABI 3130 genetic analyzer (Applied Biosystems).

The SBT protocol described above was used for identification of SNP in the coding and untranslated regions of the *PRL* gene from 20 additional ewes to those used for the validation experiment and for genotyping exons 2 and 3 of the *PRL* gene, which were found to be polymorphic. Individual records were collected for all animals, including lactation number, lambing date, and age of lambing. Phenotypic data included repeated records for total lactation milk yield, milk fat percentage, and litter size at birth. A total of 621 records were available for the 247 ewes of the study.

The effect of each genotype on the phenotypic traits was assessed with the following mixed linear model; each trait was analyzed separately:

$$Y_{jklmn} = \mu + YS_j + L_k + b_1 age + b_2 dur + G_l + A_m + e_{ijklmn},$$

where Y = total lactation milk yield, milk fat percentage, or litter size at birth record n of animal m, µ = overall population mean for the trait, YS = fixed effect of year (2009–2015) by season (1–2) of lambing interaction j, L = fixed effect of lactation k (1–4), b_1 = linear regression on age at lambing (age), b_2 = linear regression on lactation duration (dur; for milk yield only), G = fixed effect of genotype l at each locus (1–3, denoting

homozygous A, heterozygous, and homozygous B), A = random effect of animal m, and e = random residual effect.

In all cases, predicted trait values for each genotype level and respective standard errors were derived; these values were reflective of the marginal genotypic effect on each trait adjusted for all other effects fitted in the model. The predicted trait values for each genotype were used to estimate additive and dominance SNP effects on the above traits, and the proportion of total phenotypic and animal variances for each trait accounted for by the SNP. The equations used were:

$$\text{Additive effect, a} =$$
$$(\text{homozygous A} - \text{homozygous B})/2,$$

$$\text{Dominance effect, d} = \text{heterozygous}$$
$$- [(\text{homozygous A} + \text{homozygous B})/2],$$

$$\text{Percentage of phenotypic variance } (V_P) \text{ due to SNP} =$$
$$100 \times \{2pq[a + d(q - p)]^2\}/V_P, \text{ and}$$

$$\text{Percentage of animal variance } (V_A) \text{ due to SNP} =$$
$$100 \times \{2pq[a + d(q - p)]^2\}/V_A,$$

where homozygous 1, heterozygous, and homozygous 2 are the predicted trait values of each genotype class, and p and q are the corresponding frequencies of alleles A and B at each locus. Variance components were estimated with the same model excluding the genotype effect. All analyses were conducted with ASReml 3 software (Gilmour et al., 2009).

The SBT protocol suggested in the present study was validated by comparison of the obtained sequencing results with the data from a standard Sanger sequencing assay of 2 different genes, in particular, for all 5 and 10 exons of the *PRL* and *ACAA2* genes, respectively. In a similar fashion to samples processed using standard Sanger sequencing and gel extraction for purification of PCR products, all traces generated using the SBT protocol showed an excellent signal-to-noise ratio, with quality values ranging from 48 to 62 (0.0015 to >0.0001% read error values) in the 2 genes analyzed. Furthermore, heterozygosity was easily detected and we found complete concordance between the sequencing data obtained by the standard sequencing and the SBT protocol (Figure 1). Among the different volumes of the fluorescent terminator mix tested, 0.5 μL produced sufficient quality reads for 500 to 600 bp and provided clear distinction of heterozygous positions. Regarding economic cost, it was estimated that using the validated SBT protocol reduced the expenditure of consumables for sequencing a 96-well plate by approximately 20 times compared with a standard Sanger sequencing protocol. Given the reduction of cost by several orders of magnitude, if appropriate primers are designed to generate specific PCR products, the cost-effective SBT assay suggested here could be used not only for SNP identification but also for genotyping of multiple SNP or other types of polymorphisms in any given target sequence without the need to optimize different protocols for each variation.

The SBT assay validated in the present study was successfully used for identification of polymorphisms within the entire coding and untranslated regions of the *PRL* gene and as a genotyping tool to simultaneously detect multiple SNP in the polymorphic exons. Comparison of the *PRL* coding sequence obtained here (submitted to GenBank with accession number KC764410.1), with the reference sequence of the sheep genome sequencing project (NM 001009306.1) confirmed the identity of the gene. In total, 5 SNP were detected on the ovine *PRL* gene, located in exons 2 (KC764410:g.567G > A, g.625C > T, g.683C > A) and 3 (KC764410:g.2015C > A, g.2101G > A), whereas the remaining 3 exons (1, 4, and 5), including the untranslated regions of the gene were found to be monomorphic in Chios sheep. With the exception of the g.567G > A SNP, detected in exon 2, which results in an Arg to His amino acid change, the remaining SNP were silent mutations.

All identified SNP in our study are common with those reported by the International Sheep Genomics Consortium in the public database European Variation Archive (http://www.ebi.ac.uk/eva/). An SNP that has been previously found in the 5′ flanking region of the prolactin gene of the Small Tail Han sheep, but not in Dorset, Hu, or Suffolk breeds (Chu et al., 2009), was also not detected in the same region of the *PRL* gene in Chios sheep. Allelic and genotypic frequencies for the *PRL* polymorphisms identified are presented in Table 2. Genotypic frequencies of all silent investigated mutations were found to be in Hardy-Weinberg equilibrium ($P > 0.05$), whereas genotypic frequencies of the nonsynonymous g.567G > A mutation deviated from Hardy-Weinberg equilibrium ($P < 0.05$).

Preliminary association analysis, based on a small amount of data (621 records of 247 ewes), showed a significant dominance effect of 0.17 ± 0.07 of the *PRL* exon 3 g.2015C > A SNP on milk fat percentage ($P < 0.05$). Heterozygous CA animals exhibited significantly higher milk fat percentage compared with homozygous AA ewes ($P < 0.05$), whereas we noted a tendency for higher percentage compared with homozygote CC ewes ($P < 0.1$). Moreover, a dominance effect of -21.33 ± 10.51 of the same SNP on total lactation milk yield was
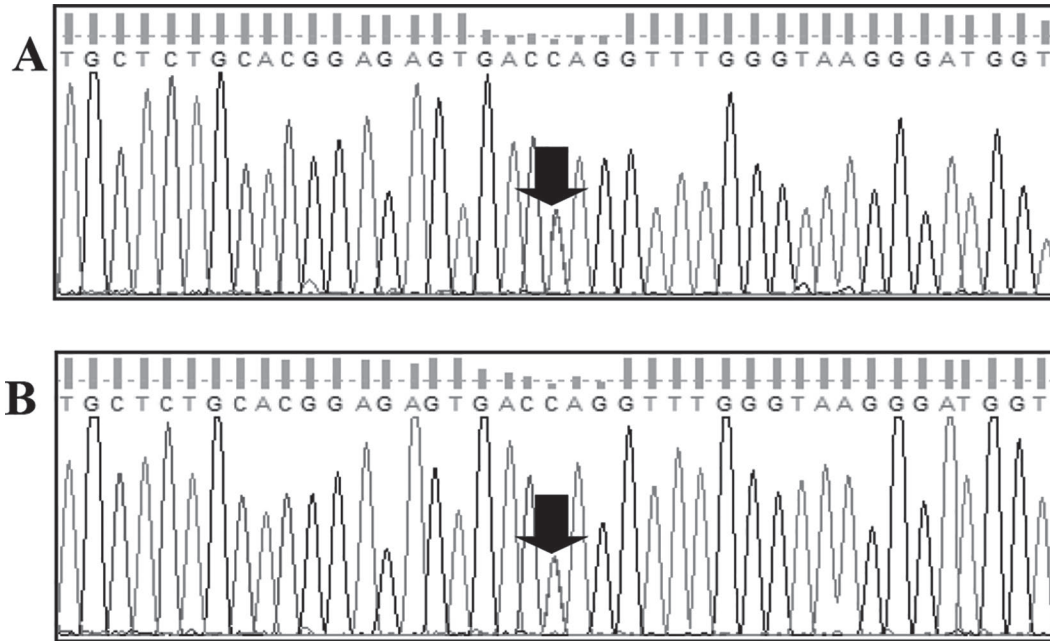
Figure 1. Comparison of sequencing assays and detection of heterozygosity. Sequencing reactions performed on the templates prepared by (A) standard sequencing- and (B) modified sequenced-based typing assay. In both traces, the position of a C/T SNP is indicated by a black arrow. Color version available online.

observed ($P < 0.05$). Based on the estimated allelic effects and the allele frequencies observed, the *PRL* exon 3 g.2015C > A SNP explained 2.47 and 3.68% of the total phenotypic variance for milk yield and fat percentage, respectively, whereas the corresponding values for the animal variance were 7.14 and 11.75%. The results suggest a significant overdominant effect on milk fat percentage and negative dominance on milk yield, which is consistent with the negative genetic correlation of those 2 traits (Fuertes et al., 1998). Although the g.2015C > A is a silent mutation that does not alter the functional conformation of the protein, strong evidence exists that even silent nucleotide substitutions in exon sequences may affect the speed and accuracy of translation (Drummond and Wilke, 2008). Moreover, synonymous mutations predominantly located near intron-exon junctions, as is the case of the g.2015C >

A SNP, may affect how the pre-mRNA is processed and arranged (Parmley et al., 2006). However, because our significant estimates do not pass the Bonferroni correction for multiple testing (Holm, 1979), further studies using larger populations and other breeds are needed to confirm these results.

A suggestive association of the nonsynonymous G to A substitution with litter size at birth ($P = 0.058$), with homozygous GG and heterozygous GA animals differing significantly from AA homozygotes ($P < 0.05$), was also detected. The ovine *PRL* gene has been previously associated with litter size at birth (Chu et al., 2009). Interestingly, genotypic frequencies for this SNP deviated from Hardy-Weinberg equilibrium. As the other 2 closely linked silent SNP in the same exon were found to be in Hardy-Weinberg equilibrium, it is unlikely that the deviation observed could be a result of genetic drift

Table 2. Genotypic and allelic frequencies of identified SNP of the prolactin gene

SNP	SNP position[1]	Genotypic frequencies	Allelic frequencies
PRL exon 2			
g.567C > T	7 G > A	GG:0.62, GA:0.36, AA:0.02	G:0.80, A:0.20
g.625C > T	65 T > C	TT:0.53, TC:0.39, CC:0.08	T:0.73, C:0.27
g.683C > A	122 C > A	CC:0.85, CA:0.14, AA:0.01	C:0.92, A:0.08
PRL exon 3			
g.2015C > A	7 C > A	CC:0.52, CA:0.37, AA:0.11	C:0.70, A:0.30
g.2101G > A	93 A > G	AA:0.58, GA:0.36, GG:0.06	A:0.76, G:0.24

[1]Number of nucleotide where the SNP is located from the beginning of the corresponding exon.

or migration. Directional selection could be a possible reason for deviation, because litter size at birth is an easily monitored trait by farmers who usually select against singletons.

Our results offer a cost effective SBT assay for identification of polymorphism and the first ever indication of possible genetic associations of the newly studied *PRL* regions with important dairy sheep traits.

REFERENCES

Chu, M. X., X. C. Wang, M. Jin, R. Di, H. Q. Chen, G. Q. Zhu, L. Fang, Y. H. Ma, and K. Li. 2009. DNA polymorphism of 5′ flanking region of prolactin gene and its association with litter size in sheep. J. Anim. Breed. Genet. 126:63–68.

Dong, C. H., X. M. Song, L. Zhang, J. F. Jiang, J. P. Zhou, and Y. Q. Jiang. 2013. New insights into the prolactin-RsaI (PRL-RsaI) locus in Chinese Holstein cows and its effect on milk performance traits. Genet. Mol. Res. 12:5766–5773.

Drummond, D. A., and C. O. Wilke. 2008. Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. Cell 134:341–352.

Fuertes, J. A., C. Gonzalo, J. Carriedo, and F. San Primitivo. 1998. Parameters of test day milk yield and milk components for dairy ewes. J. Dairy Sci. 81:1300–1307.

Gilmour, A. R., B. Gogel, B. Cullis, R. Thompson, and D. Butler. (2009) ASReml user guide release 3.0. VSN International Ltd., Hemel Hempstead, UK.

Holm, S. 1979. A simple sequentially rejective multiple test procedure. Scand. J. Stat. 6:65–70.

Lacasse, P., and S. Ollier. 2015. The dopamine antagonist domperidone increases prolactin concentration and enhances milk production in dairy cows. J. Dairy Sci. 98:7856–7864.

Lü, A., X. Hu, H. Chen, J. Jiang, C. Zhang, H. Xu, and X. Gao. 2010. Single nucleotide polymorphisms in bovine PRL gene and their associations with milk production traits in Chinese Holsteins. Mol. Biol. Rep. 37:547–551.

Orford, M., G. Hadjipavlou, O. Tzamaloukas, D. Chatziplis, A. Koumas, A. Mavrogenis, C. Papachristoforou, and D. Miltiadou. 2012. A single nucleotide polymorphism in the acetyl-coenzyme A acyltransferase 2 (ACAA2) gene is associated with milk yield in Chios sheep. J. Dairy Sci. 95:3419–3427.

Orford, M., O. Tzamaloukas, C. Papachristoforou, and D. Miltiadou. 2010. Technical note: A simplified PCR-based assay for the characterization of two prolactin variants that affect milk traits in sheep breeds. J. Dairy Sci. 93:5996–5999.

Parmley, J. L., J. V. Chamary, and L. D. Hurst. 2006. Evidence for purifying selection against synonymous mutations in mammalian exonic splicing enhancers. Mol. Biol. Evol. 23:301–309.

Ramos, A. M., C. A. P. Matos, P. A. Russo-Almeida, C. M. V. Bettencourt, J. Matos, A. Martins, C. Pinheiro, and T. Rangel-Figueiredo. 2009. Candidate genes for milk production traits in Portuguese dairy sheep. Small Rumin. Res. 82:117–121.

Raven, L. A., B. G. Cocks, M. E. Goddard, J. E. Pryce, and B. J. Hayes. 2014. Genetic variants in mammary development, prolactin signalling and involution pathways explain considerable variation in bovine milk production and milk composition. Genet. Sel. Evol. 46:29.

Staiger, E. A., M. L. Thonney, J. W. Buchanan, E. R. Rogers, P. A. Oltenacu, and R. G. Mateescu. 2010. Effect of prolactin, β-lactoglobulin, and κ-casein genotype on milk yield in East Friesian sheep. J. Dairy Sci. 93:1736–1742.