ANIMAL GENETICS Immunogenetics, Molecular Genetics and Functional Genomics

SHORT COMMUNICATION

doi: 10.1111/age.12616

Molecular characterization of the llama *FGF5* gene and identification of putative loss of function mutations

M. S. Daverio*, L. Vidal-Rioja*, E. N. Frank† and F. Di Rocco*

*Laboratorio de Genética Molecular, Instituto Multidisciplinario de Biología Celular (IMBICE), CCT CONICET La Plata – CICPBA – UNLP, Calle 526 e/10 y 11, La Plata 1900, Buenos Aires, Argentina. †IRNASUS, CONICET-UCC Universidad Católica de Córdoba, Av. Armada Argentina 3555, X5014YIG Cordoba, Argentina.

Summary

Llama, the most numerous domestic camelid in Argentina, has good fiber-production ability. Although a few genes related to other productive traits have been characterized, the molecular genetic basis of fiber growth control in camelids is still poorly understood. Fibroblast growth factor 5 (FGF5) is a secreted signaling protein that controls hair growth in humans and other mammals. Mutations in the *FGF5* gene have been associated with long-hair phenotypes in several species. Here, we sequenced the llama *FGF5* gene, which consists of three exons encoding 813 bp. cDNA analysis from hair follicles revealed the expression of two *FGF5* alternative spliced transcripts, in one of which exon 2 is absent. DNA variation analysis showed four polymorphisms in the coding region: a synonymous SNP (c.210A>G), a single base deletion (c.348delA), a 12-bp insertion (c.351_352insCATATAACATAG) and a non-sense mutation (c.499C>T). The deletion was always found together with the insertion forming a haplotype and producing a putative truncated protein of 123 amino acids. The c.499C>T mutation also leads to a premature stop codon at position 168. In both cases, critical functional domains of FGF5, including one heparin binding site, are lost. All animals analyzed were homozygous for one of the deleterious mutations or compound heterozygous for both (i.e. c.348delA, c.351_352insCATATAACATAG/c.499T). Sequencing of guanaco samples showed that the *FGF5* gene encodes a full-length 270-amino acid protein. These results suggest that *FGF5* is likely functional in short-haired wild species and non-functional in the domestic fiber-producing species, the llama.

Keywords fiber, *fibroblast growth factor* 5 gene, camelids, polymorphisms, splicing variants

The hair growth cycle consists of three phases: active growth or anagen, regression or catagen, and quiescence or telogen (Chase 1954). The *FGF5* gene is the main regulator of hair growth. Two products of this gene, FGF5 and a shorter isoform FGF5-s, produced by alternative splicing of the mRNA, were first described in rat (Hattori *et al.* 1996) and isolated later from other species (Suzuki *et al.* 2000; Housley & Venta 2006; Higgins *et al.* 2014). Both isoforms are involved in the regulation of the hair cycle; FGF5 inhibits hair growth by inducing the passage from anagen to catagen, whereas FGF5-s acts as an antagonist of FGF5, suppressing its activity at anagen VI (Suzuki *et al.* 2000).

Address for Correspondence

F. Di Rocco, Laboratorio de Genética Molecular, Instituto Multidisciplinario de Biología Celular (IMBICE), CCT CONICET La Plata – CICPBA – UNLP, Calle 526 e/10 y 11, La Plata 1900, Buenos Aires, Argentina.
Email: fdirocco@imbice.gov.ar

Accepted for publication 11 September 2017

Deleterious mutations in *FGF5* predicted to alter protein function have been associated with long-hair phenotypes in species such as mice, cats, dogs and donkeys (Hébert *et al.* 1994; Drögemüller *et al.* 2007; Cadieu *et al.* 2009; Legrand *et al.* 2014). The inheritance mode of this trait has been shown to be recessive (Crary & Sawin 1953; Burns & Fraser 1966; Lloyd 1987), thus the presence of two nonfunctional *FGF5* alleles are required for the expression of the long-hair phenotype.

The llama (*Lama glama*) was domesticated in the Andes from its wild ancestor, the guanaco (*Lama guanicoe*), as a multipurpose animal. However, at present, its main product is fiber. In Argentina, the predominant llama phenotype is characterized by long fiber and usually by a wide coverage of the fleece that extends to the head and legs (Fig. S1) (Hick *et al.* 2009). Compared with the guanaco, which has short fibers of 1.4–3.8 cm (Von Thüngen *et al.* 2005), the mean staple length in the llamas, adjusted by 12-month growth periods, is about 15.5 cm (Frank *et al.* 2006).

Genes that control fiber growth and its characteristics are still unknown for camelids. Here, we have characterized the *FGF5* gene, described two alternative mRNA spliced products, and identified genetic variants in Argentine llamas. The three coding exons of llama *FGF5* gene were amplified by PCR with three pairs of primers designed on the alpaca (*Vicugna pacos*) genomic sequence (Ensembl: GeneScaffold _2359: 620885: 644690: 1) (Table S2) and then sequenced as described in Appendix S1. New DNA sequences obtained were deposited in GenBank under accession numbers KY357463, KY357464 and KY357465.

The full coding region of 813 bp was organized into three exons. RT-PCR analysis showed that two *FGF5* transcripts of different lengths were expressed in the llama hair follicle (Fig. 1). The shortest differed from the longest on account of the absence of the complete exon 2 (104 bp), corresponding to the FGF5-s isoform reported for other species.

The three *FGF5* exons and their adjacent intronic regions were screened for variation in 25 unrelated llamas. Five polymorphisms were discovered, three of which—c.348delA, c.351_352insCATATAACATAG and c.499C>T—affect the putative protein encoded (Fig. 2).

At these three positions, three individuals showed homozygous genotypes del-del/ins+ins+/C-C, 15 showed homozygous genotypes A-A/ins-ins-/T-T and the remaining were heterozygous A-del/ins+ins-/C-T (Table S3). Haplotype phases could be unambiguously inferred from the homozygous animals, and two haplotypes were identified, one carrying both c.348delA and c.351_352insCATATAACATAG mutations and the other bearing c.499C>T (Fig S2). Remarkably, both haplotypes code for truncated proteins, of 123 and 166 amino acids respectively (Fig. 3). A BLAST search showed 100% homology of this last variant with the alpaca protein (accession no. XP_006198283). To investigate whether llama polymorphisms were also found in the ancestral species, we examined three guanaco samples. Sequences obtained showed two synonymous SNPs in exon 1 and three intronic polymorphisms, all of them coding for the same full-length protein of 270 amino acids.

The combination of the llama *FGF5* c.348delA and c.351_352insCATATAACATAG mutations produces a frame-shift leading to the substitution of residues from positions 117 to 123 and to the introduction of a premature stop codon at amino acid 124. These mutations affect both FGF5 isoforms in the same way by generating a single protein of 123 amino acids.

To evaluate the putative functional impact of the substitutions, three-dimensional models of the wild and mutant protein were built (Fig. S3). Compared to the

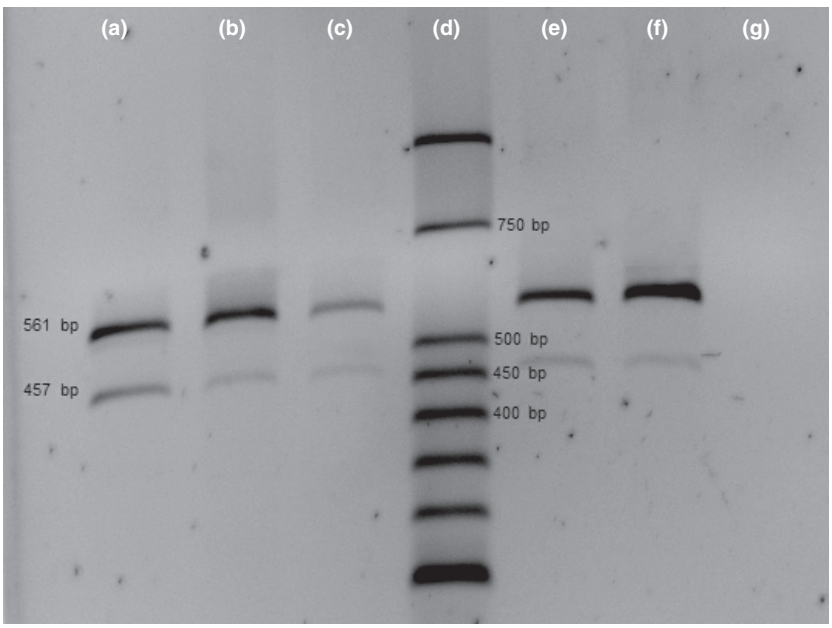

Figure 1 Agarose gel electrophoresis showing the two PCR products (561 and 467 bp) corresponding to the FGF5 isoforms found in llamas. Lanes a, b, c, e and f correspond to each of the cDNA samples amplified. Lane d indicates the 50 bp molecular weight marker and lane g, the negative control.

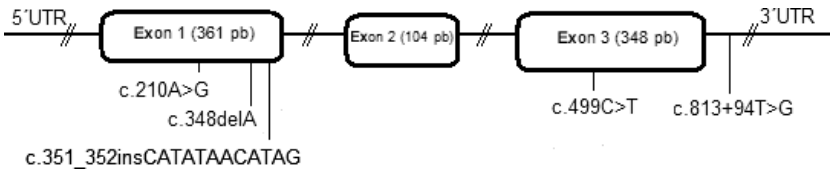

Figure 2 Structural organization and location of the polymorphisms identified in the llama *FGF5* gene. Position +1 was assigned to the translation start site.

Figure 3 Alignment of the predicted FGF5 amino acid sequence from the guanaco (*Lama guanicoe*) and the llama c.351_352insCATATAACATAG and c.499C>T variants. Amino acid substitutions at the C terminal end of the c.351_352insCATATAACATAG variant are shaded in grey.

270-amino-acid protein encoded by *FGF5* in guanaco and in other species (Drögemüller *et al.* 2007; Cadieu *et al.* 2009), this variant is 147 amino acids shorter and lacks one of the heparin binding sites and most of the conserved core of residues that make contact with the FGF1R receptor (Beenken & Mohammadi 2009).

The other substitution found, c.499C>T, also codes for a premature stop codon (p.Arg167*), resulting in the deletion of the 103 amino acids from the C terminus of the protein. Five of the 12 beta sheets involved in the FGF function are lost, including β10 and β12, which contain one of the heparin binding sites (Fig. S3). In this case as well, the protein loses critical domains necessary for its function. Because the mutation is located at exon 3, it affects only the long isoform, whereas FGF5-s can be translated normally.

FGF5 mutations similar to those described in llamas have been found to be associated with long-hair phenotypes in other species. In the long-haired Poitou donkey breed, two independent recessive mutations that lead to different truncated proteins of 82 and 159 amino acid length have been associated with this trait (Legrand *et al.* 2014). Similarly, two frame-shift *FGF5* polymorphisms that produce proteins of 255 and 261 amino acids respectively, c.556_571del16 and c.559_560dupGG, have been associated with the long-hair phenotype in dogs (Dierks *et al.* 2013).

In all species studied so far, the long-hair phenotype is inherited as a recessive trait. We found that all llamas studied were homozygous for one of the mutated alleles or heterozygous for a combination of both. Although the mode of inheritance of this trait in this species has not been established, our molecular data are also consistent with recessive inheritance. Therefore, even though functional assays need to be done to prove this hypothesis, it is very likely that the two variants of *FGF5* described here for llamas with the long-fiber phenotype could represent *FGF5* variants with loss of function.

On the other hand, the putatively functional allele was observed only in the short-haired wild species, guanaco. If the mutations identified are indeed associated with long hair and it is a recessive trait, we expect the functional allele to appear only in those llamas with short hair. This phenotype, known as 'kcara' is absent in most herds and has low overall frequency in llama populations of Argentina. Thus, the identification and study of populations in which this phenotype is segregated will be necessary to verify this proposition.

Acknowledgements

The authors thank M. Silbestro for her technical assistance and all llama breeders who allowed us to take samples. This work was supported by grant PIP-00370 from The National Scientific and Technical Research Council (CONICET) and funds from the Commission of Scientific Research of the Buenos Aires province.

Conflict of interest

The authors declare no conflict of interests.

References

Beenken A. & Mohammadi M. (2009) The FGF family: biology, pathophysiology and therapy. *Nature Reviews Drug Discovery*, **8**, 235–53.

Burns M. & Fraser M.N. (1966) *Genetics of the Dog: The Basis of Successful Breeding*. Oliver and Boyd, Edinburgh, UK.

Cadieu E., Neff M., Quignon P. *et al.* (2009) Coat variation in the domestic dog is governed by variants in three genes. *Science* **326**, 150–3.

Chase H.B. (1954) Growth of the hair. *Physiological Reviews* **34**, 113–26.

Crary D.D. & Sawin P.B. (1953) Some factors influencing the growth potential of the skin in the domestic rabbit. *Journal of Experimental Zoology* **124**, 31–62.

Dierks C., Mömke S., Philipp U. & Distl O. (2013) Allelic heterogeneity of *FGF5* mutations causes the long-hair phenotype in dogs. *Animal Genetics* **44**, 425–31.

Drögemüller C., Rüfenacht S., Wichert B. & Leeb T. (2007) Mutations within the *FGF5* gene are associated with hair length in cats. *Animal Genetics* **38**, 218–21.

Frank E.N., Hick M.V.H., Gauna C.D., Lamas H.E., Renieri C. & Antonini M. (2006) Phenotypic and genetic description of fibre traits in South American domestic camelids (llamas and alpacas). *Small Ruminant Research* **61**, 113–29.

Hattori Y., Yamasaki M. & Itoh N. (1996) The rat *FGF*-5 mRNA variant generated by alternative splicing encodes a novel truncated form of FGF-5. *Biochimica et Biophysica Acta* **1306**, 31–3.

 doi: 10.1111/age.12616

Hébert J.M., Rosenquist T., Götz J. & Martin G.R. (1994) FGF5 as a regulator of the hair growth cycle: evidence from targeted and spontaneous mutations. *Cell* **78**, 1017–25.

Hick M., Lamas H., Echenique J., Prieto A., Castillo M. & Frank E. (2009) Estudio demográfico de los atributos morfológicos y productivos en poblaciones de llamas (*Lama glama*) de la provincia de Jujuy, Argentina. *Animal Genetic Resources Information* **45**, 71.

Higgins C.A., Petukhova L., Harel S., Ho Y.Y., Drill E., Shapiro L., Wajid M. & Christiano A.M. (2014) FGF5 is a crucial regulator of hair length in humans. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 10648–53.

Housley D.J. & Venta P.J. (2006) The long and the short of it: evidence that FGF5 is a major determinant of canine 'hair'-itability. *Animal Genetics* **37**, 309–15.

Legrand R., Tiret L. & Abitbol M. (2014) Two recessive mutations in *FGF5* are associated with the long-hair phenotype in donkeys. *Genetics Selection Evolution* **46**, 65.

Lloyd A.T. (1987) Cats from history and history from cats. *Endeavour* **11**, 112–5.

Suzuki S., Ota Y., Ozawa K. & Imamura T. (2000) Dual-mode regulation of hair growth cycle by two *Fgf*-5 gene products. *Journal of Investigative Dermatology* **114**, 456–63.

Von Thüngen J., Gálvez C.M., Sacchero D. & Duga L. (2005) Análisis de calidad de la fibra de guanaco (*Lama guanicoe* M.) en la Patagonia. *Revista Argentina de Producción Animal* **25**, 382–3.

Supporting information

Additional supporting information may be found online in the supporting information tab for this article:

Figure S1 Examples of the long-fiber llama phenotype characterized by wide coverage of the fleece that extends to the head and legs.

Figure S2 Partial alignment showing the polymorphisms found in the llama *FGF5* nucleotide sequence: c.348delA and c.351_352insCATATAACATAG.

Figure S3 3D model of the FGF5 protein.

Table S1 Origin and number of samples analyzed.

Table S2 Primers and PCR conditions.

Table S3 Llama *FGF5* genotypes with the three positions affecting the putative encoded protein.

Appendix S1 Materials and methods.