Animal Genetics, 1998, **29**, 173–177

Identification and analysis of the dog keratin 9 (*KRT9*) gene

P Lachaume, C Hitte, S Jouquand, C Priat, F Galibert

Summary

We have identified the gene coding for the canine ortholog of the human keratin 9 protein using the inverse-polymerase chain reaction (PCR) strategy. Sequence comparison and structure analysis of the gene show marked similarity with the human gene. This gene spans about 7 kb and spreads over eight exons. In the dog gene, the reading frame is extended by 20 codons, the first in-frame stop codon being in exon 8 in the dog rather than in exon 7 as in humans. Alignment of human and dog predicted amino acid sequences confirms the high analogy, reaching 75% identity and 95% similarity in the rod domain. Interestingly, the glycine-loop motif number in the C-terminal V2 variable subdomain of the protein increases from 19 in human to 43 in dog, generating a size difference of 12 kDa between the two proteins. Due to its restricted expression pattern in mammalian epidermis, dog keratin 9 gene was a good candidate gene for the genetic palmoplantar hyperkeratosis observed in the Dogue de Bordeaux. However, no polymorphism associated with the pathology was detected within an affected Dogue de Bordeaux pedigree ruling out this hypothesis.

Keywords: dog, keratin 9 gene, microsatellite, polymorphism

Introduction

Keratins belong to a multigenic family of structural proteins that constitute the intermediate filaments present in all epithelial cells (Moll *et al*. 1982; Fuchs 1988; Skalli *et al*. 1992).

Keratin 9 (K9) is a large type I keratin (64 kDa) known to be almost exclusively synthesized in the epidermis of palms and soles, as shown by K9 specific antibodies (Moll *et al*. 1987; Langbein *et al*. 1993). The protein is composed of three parts: a specific N-terminal (head) and a C-terminal (tail) domain separated by a central rod domain. The rod domain which is highly conserved among keratin family members, consists of 4 α-helices named coil 1A, 1B, 2A and 2B interrupted by linkers L1, L12 and L2.

The human keratin 9 gene (*KRT9*) has been assigned to the 17q21·1–2 chromosomal region (Reis *et al*. 1994). The *KRT9* cDNA has been cloned (Langbein *et al*. 1993) and a cosmid containing *KRT9* has been isolated and partially sequenced (Reis *et al*. 1994). The gene spans about 6 kb and spreads over eight exons.

Due to its expression pattern, *KRT9* was proposed as a candidate gene for two related skin diseases, the epidermolytic palmoplantar keratoderma (EPPK) and the non-epidermal palmoplantar keratoderma (NEPPK). EPPK is an autosomal dominantly inherited skin disease characterized by diffuse thickening of the epidermis on the entire surface of palms and soles (McKusick 1992). NEPPK are a group of very similar disease with no evidence of epidermolysis (Kelsell *et al*. 1995). Following this hypothesis, five point mutations altering the highly conserved residues in the coil 1A part of the human K9 protein were discovered in patients suffering from EPPK (Hennies *et al*. 1994; Reis *et al*. 1994; Torchard *et al*. 1994; Endo *et al*. 1997).

While analysing a large set of dog microsatellites the present authors isolated a CA-repeat marker analogous to that found in the human *KRT9* gene (Reis *et al*. 1994). From this clone, the present authors undertook the cloning and sequencing of the complete canine KRT9 using the inverse-polymerase chain reaction (PCR) strategy (Ochman *et al*. 1988).

We have also determined the sequence of the canine *KRT9* gene in a pure breed Dogue de Bordeaux family suffering from foot pad hyperkeratosis (Paradis 1992). However, in this family, the present authors found no evidence or related polymorphism to support the hypothesis that *KRT9* gene is responsible for the pathological trait.

Materials and methods

Inverse-PCR strategy

Total digestion of 1 μg of genomic DNA from a non-breed dog was performed during 5 h in a 20 μl reaction volume with 10 units of enzyme.

P Lachaume
UMR 6745 CNRS Université Blaise Pascal, Campus de Cezeaux, 24, Avenue des Landais, 63177 Aubiere Cedex, France
C Hitte
S Jouquand
C Priat
F Galibert
Laboratoire de Recombinaisons Génétiques, UPR41 CNRS, Faculté de Médecine, Avenue du Professeur Léon Bernard 35043 Rennes Cedex, France

Correspondence: C Hitte.

Accepted 3 February 1998

Restriction enzymes were chosen according to the restriction map of the known part of the canine gene. The choice of an enzyme without a restriction site in the sequenced region allows bidirectional progress, whereas the choice of an enzyme with a restriction site near one of the ends of the known part results in unidirectional progress. Fifty nanograms of digested DNA was self-ligated overnight at 16 °C in a total volume of 200 µl, using T4 DNA ligase (Biolabs, Beverly, MA). Two successive rounds of PCR were performed with 18–21 bp primers selected from the known sequence of the CA clone using the software Oligo 5.0. The first PCR was set up with the Takara Ex Taq kit (Takara Bioproduct, Gagny, France) in a 30 µl volume, using 0·25 ng (1 µl) of the ligation mixture as template. Amplification was allowed to proceed for 30 cycles consisting each of 30 s at 94 °C, 1 min at the annealing temperature and 1–3 min at 72 °C. To increase PCR specificity and efficiency, a second nested PCR was performed in the same conditions, using internal primers and 1 µl of the first PCR as template.

Sequencing of the PCR products

Two to six microlitres of PCR products served as template for sequencing using a Taq dye terminator cycle sequencing kit and the ABI 373 DNA sequencer. The entire fragment was sequenced using both PCR primers and, if necessary, specific internal primers chosen from the sequence previously obtained.

DNA sequences were assembled with the SeqMan software of the DNASTAR package (DNASTAR, Madison, WI). Human and dog sequences were aligned with the Megalign software of the same package.

Table 1. Comparison of the different parts of human and dog KRT9 genes

Introns/exons	Length (bp) *Homo sapiens*	Length (bp) *Canis familiaris*	Identity %
5′UTR	66	69	67·0
Coding part of exon 1	641	594	80·9
Intron 1*	156†	1536	75·2‡
Exon 2	83	83	83·1
Intron 2	122	118	71·3
Exon 3	157	157	84·7
Intron 3	271	276	71·3
Exon 4	162	162	87·6
Intron 4§	813	723	72·4
Exon 5	126	126	82·6
Intron 5	122	124	77·2
Exon 6	224	224	84·8
Intron 6	410	363	75·0
Exon 7	518	995	ND
Intron 7*	329†	1090	70·1
Exon 8*	309†	173†	77·5**

*The human introns 1 and 7 (Reis *et al.* 1994) and the end of the non-coding part of the dog exon 8 have not been totally sequenced.
†Length of the known sequence is indicated.
‡Identity percentage is calculated from the known part of the two genes.
§Identity percentage of intron 4 is calculated in the flanking regions of the microsatellite.
**This exon being non-coding in human but partially coding in dog, the identity is calculated from the global known part.
ND, not determined: calculation of analogy is impossible because of the large modification of the glycine-loop number between human and dog K9 proteins.

Polymorphism analysis of the microsatellite

The polymorphism of the intron 4 CA-repeat was analysed in 15 unrelated non-breed dogs and in five pure-breed Dogue de Bordeaux.

The upper primer was labelled with 10 units of T4 polynucleotide kinase (Biolabs) using 1 µCi of [γ-P^{32}]ATP per ng. PCR was set up with the Perkin Elmer PCR kit using 10 ng of radiolabelled primers. Amplification was allowed to proceed for 30 cycles consisting each of 30 s at 94 °C, 30 s at the annealing temperature and 20 s at 72 °C. The annealing temperature of the reaction was decreased by 0·5 °C for every cycle from 59 °C to a touchdown of 49 °C (Don *et al.* 1991; Holmes *et al.* 1993).

PCR products were analysed by migration in a denaturing 6% polyacrylamide gel followed by autoradiography with HYPERfilm-β MAX (Amersham, Loulis, France).

Dogue de Bordeaux family

Five members, a mother and four siblings of a Dogue de Bordeaux family were analysed. Three of the offspring showed footpad with papillated epidermal hyperplasia with diffuse orthokeratotic hyperkeratosis (Paradis 1992). The mother and the fourth sibling presenting none of these symptoms served as controls.

Lymphocyte DNA from each dog was extracted from peripheric blood. Each KRT9 exon was amplified in the five Dogue de Bordeaux and in a non-breed dog using specific primers (Table 1) and the touchdown method as described above.

Results and discussion

A genomic library made in pBluescript from a non-breed dog was screened with a (CA)10 probe to identify tandem repeat sequences. The positive clones were sequenced and evaluated for nucleic acid similarity against the GenBank database using BLASTN on-line application

Table 2. Polymorphism analysis of the KRT9 associated CA-repeat

	Alleles names	PCR product size (bp)	No. of alleles	Heterozygosity level
	1	130	4	
	2	132	1	
Unrelated	3	134	16	0·67
Non-breed dogs	4	136	2	
	5	140	8	
	6	148	1	
Dogue de Bordeaux	5	140	10	0·00

Heterozygosity level is calculated with the formula: $(1 - \Sigma)_{i=1}^{n}$(frequency of allele i)^2.
Primers used in CA-polymorphism analysis: KRT9CAu, CCCCTGGCCTATGAGAGC; KRT9CAl, TGTCCCCAAACCACCTT.

(Altschul *et al.* 1990). The 962-bp insert of the 01E05 clone revealed 80% analogy with the *KRT9* gene (Reis *et al.* 1994). This clone contains part of intron 4, exon 5, intron 5 and part of exon 6 of the dog ortholog of *KRT9* (Fig. 1).

We used a chromosome walking strategy based on inverse-PCR to sequence the whole dog *KRT9* gene. Two PCR fragments of 1·1 and 1·6 kb were generated with *Bam*HI and *Hind*III digests, respectively (Fig. 1C). Both were completely sequenced, giving 1541 bp of additional sequence to the *KRT9* sequence.

A second run of unidirectional inverse PCR was performed to identify the 5′ and 3′ parts of the gene. In order to progress toward the 5′ end, the present authors generated a *Pvu*II digest from which they obtained a 2·8 kb PCR product (Fig. 1D). Sequencing of the fragment gave 2467 bp of novel sequence, while the remaining 250 bp exactly matched the already known sequence. These 2467 bp comprise part of the promoter region, the first exon and the first intron. In the meantime, the present authors generated from a *Rsa*I digest a 1·2 kb PCR

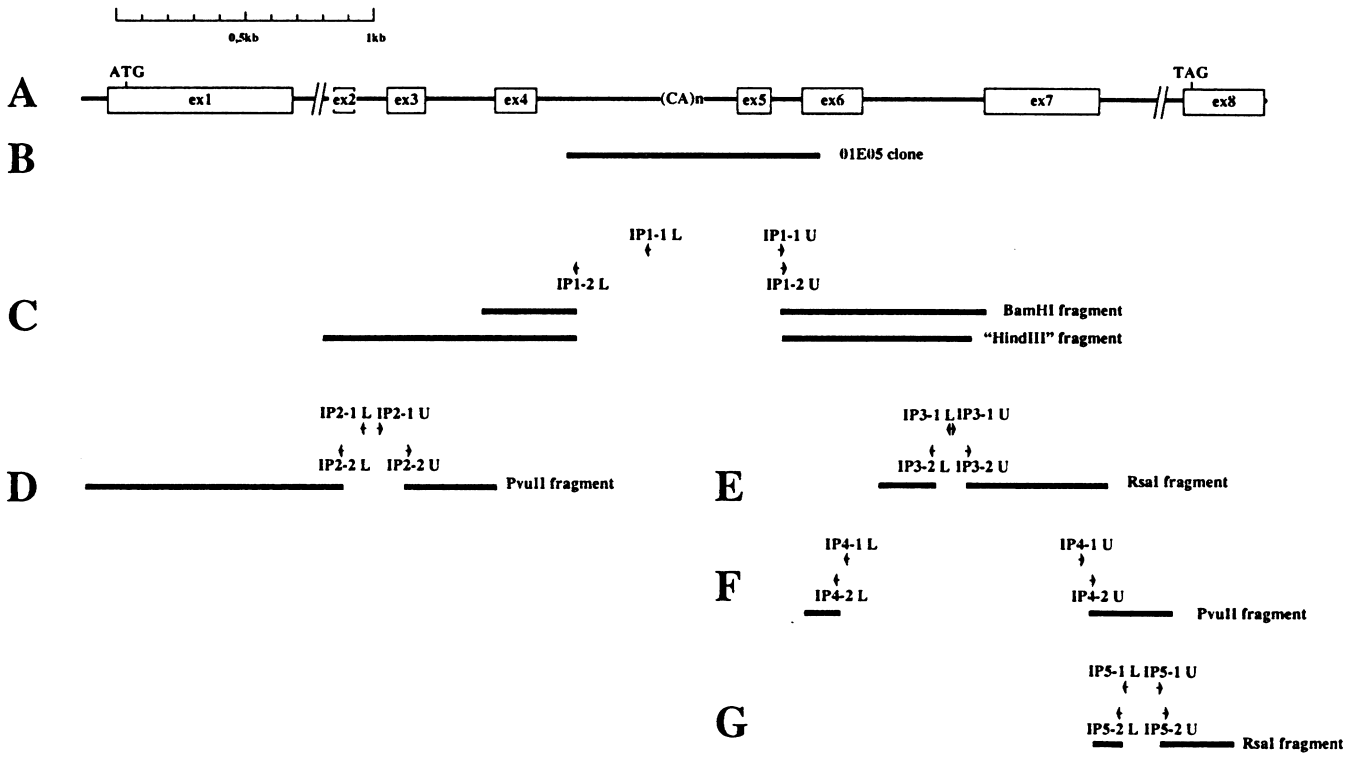

Fig. 1. Dog KRT9 gene and cloning strategy. (A) Organization of KRT9 gene. Open boxes denote exonic parts, black horizontal dashes are introns. Start and stop codons are mapped. The polymorphic microsatellite in intron 4 is indicated as (CA)n. (B) Position of the initial clone 01E05, containing the polymorphic microsatellite. (C–G) Successive steps of inverse-PCR used to sequence KRT9. Horizontal dashes indicate the location of the different PCR fragment obtained at each step. The arrows above represent the two pairs of PCR primers used to generate these fragments. The names of these primers are indicated.

product (Fig. 1E) to progress toward the 3′ end. We determined by sequencing 967 additional nucleotides toward the 3′ end comprising exon 7 and the beginning of intron 7.

Two additional inverse-PCR runs, generating a 1·2kb *Pvu*II fragment (Fig. 1F) followed by a 900 bp *Rsa*I fragment (Fig. 1G) were required to obtain the whole sequence of intron 7 and the coding part of exon 8 (Fig. 1A).

The complete genomic sequence of the dog *KRT9* spans 7135 bp (Fig. 1A). DNA sequence analysis reveals high analogy with the human *KRT9*, reaching 87·6% identity in the coding regions (Table 1). The position of splice boundaries is identical in human and dog genes and the sequences of the splice sites are well conserved. All known exons were recognized by the GRAIL software (Oak Ridge National Laboratories, Oak Ridge, TN) (Uberbacher & Mural 1991) with 'excellent scores'. All exons have identical sizes, except for exons 1 and 7 that encode the variable parts of the protein ends. The present authors can then conclude that the structure of the human and dog genes is identical.

Protein comparison

The alignment of human and dog predicted amino acid sequences confirms their strong analogy (72% identity and 88% similarity). This analogy reaches 75% identity and 95% similarity in the rod domain. The C-terminal tail (V2 subdomains) is more divergent with 335 aa in dog vs. 156 aa in human protein generating a 12 kDa size difference (K9 dog protein calculated size is 76 kDa vs. 64 kDa for human K9). Analysis of the C-terminal part of the dog K9 shows that it is essentially composed of 43 tandem repeats of a very conserved motif of 6 aa or 8 aa. This motif is conserved in the human protein, although with a more variable length and only 19 repeats. On the contrary, the head V1 subdomain is shorter by 15 aa in the dog protein, with no dramatic consequences for the number and integrity of the glycine loops.

Short variations of glycine-loop number in K1, K10 and loricrin do not affect the structural organization or function of these motifs (Korge *et al*. 1992a,b). However, large interspecies size variations constitute a peculiarity of the highly specialized keratins (Steinert & Roop 1988; Herzog *et al*. 1994). In the case of K9, the present authors are led to suspect a correlation between the very large variation in glycine-loop number in V2 subdomain and the difference in the nature of the palm and sole epidermis in human and dog.

Moreover, the stop codon of the human gene is in exon 7, exon 8 being a non-coding exon (Reis *et al*. 1994). The TAG codon is replaced by CAG in the dog gene, extending the reading frame by 20 codons, with the first stop codon being in exon 8. This surprising discrepancy prompted the present authors to verify the human exon 7 sequence by PCR amplification and sequencing. This confirmed the presence of the stop codon in the penultimate exon in the human gene (data not shown).

Promoter region analysis

A unique promoter sequence is predicted at the beginning of the sequence. The TATA box is localized at position –22 and the transcription initiation site is predicted at position 1, in complete accordance with the human cDNA sequence.

Search for (CA) polymorphism in the Dogue de Bordeaux

In human, *KRT9* mutations in the rod domain induce a disorder known as EPPK. Thus, the dog *KRT9* gene was a good candidate for the genetic palmoplantar hyperkeratosis observed in Dogue de Bordeaux. (Paradis 1992; Gaguère, personal communication). The present authors used the CA-repeat of intron 4 as a possible marker for the disorder, and first looked for a polymorphism of this microsatellite in 15 unrelated non-breed dogs (see Materials and methods). As in human, the dog CA-repeat is highly polymorphic (six alleles) and the calculated heterozygosity in this cohort is 0·67 (Table 2). However, no polymorphism was detected in the Dogue de Bordeaux pedigree used in this study (Table 2) rendering this microsatellite useless for investigating a possible involvement of *KRT9* in the disorder.

Polymorphism analysis by sequencing

To analyse the possible involvement of *KRT9* in the foot pad hyperkeratosis syndrome observed in the Dogue de Bordeaux, the present authors sequenced the proximal promoter region (254 bp before initiation transcription site) and all the coding parts of *KRT9* in five members of the Dogue de Bordeaux family, three of which presented the pathology and two were healthy, and in two additional healthy mongrel dogs.

The sequence data of each exon were analysed and compared in the different dogs. No difference was observed between the five Dogue de Bordeaux. Two discrepancies only were

detected by comparing *KRT9* exonic sequences in the Dogue de Bordeaux and the two mongrels dogs. A C → T substitution in exon 3 without any change of the corresponding aa, and an A → G substitution in exon 4 that converts an asparagine residue to a glycine (aa 309). However, as this mutation is present in all the five Dogue de Bordeaux, it cannot be implicated in the foot pad hyperkeratosis observed in three of them. So no mutation in the K9 protein was detected in the Dogue de Bordeaux hyperkeratosis. Consequently it may be hypothesized either that the mutation is located in a non-coding part of the gene (in introns or upstream part of the proximal promoter region analysed), or that *KRT9* is not responsible for the Dogue de Bordeaux hyperkeratosis.

Acknowledgements

The authors thank Dr D. Gaguère for providing blood samples from the Dogue de Bordeaux pedigree, J. C. Chuat for reading the manuscript, F. Vignaux and C. André for helpful comments. This work was supported by a GREG grant for P. Lachaume and C. Hitte and French governmental fellowship for S. Jouquand and C. Priat. We thank the Centre National de la Recherche Scientifique (CNRS) and the Conseil Régional de Bretagne for supporting this work.

References

Altschul S.F., Gish W., Miller W., Myers E.W. & Lipman D.J. (1990) Basic local alignment search tool. *Journal of Molecular Biology* **215,** 403–10.

Don R.H., Wainwright B.J., Baker K. & Mattick J.S. (1991) 'Touchdown' PCR to circonvent spurious priming during gene amplification. *Nucleic Acids Research* **19,** 4008.

Endo H., Hatamochi A. & Shinkai H. (1997) A novel mutation of a leucine residue in coil 1A of keratin 9 in epidermolytic palmoplantar keratoderma. *Journal of Investigations Dermatology* **109** (1), 113–5.

Fuchs E. (1988) Keratins as biochemical markers of epithelial differentiation. *Trends in Genetics* **4,** 277–81.

Hennies H.-C., Zehender D., Kunze J., Küster W. & Reis A. (1994) Keratin 9 gene mutational heterogeneity in patients with epidermolytic palmoplantar keratoderma. *Human Genetics* **93,** 649–54.

Herzog F., Winter H. & Schweizer J. (1994) The large type II 70-kDa keratin of mouse epidermis is the ortholog of human keratin K2e. *Journal of Investigations Dermatology* **102,** 165–70.

Holmes N.G., Mellersh C.S., Humphreys S.J. *et al.* (1993) Isolation and characterization of microsatellites from the canine genome. *Animal Genetics* **24,** 289–92.

Kelsell D.P., Stevens H., Ratnavel R. *et al.* (1995) Genetic linkage studies in non-epidermolytic palmoplantar keratoderma: evidence for heterogeneity. *Human Molecular Genetics* **4,** 1021–5.

Korge B.P., Compton J.G., Steinert P.M. & Mischke D. (1992a) The two size alleles of human keratin 1 are due to a deletion in the glycine-rich carboxyl-terminal V2 subdomain. *Journal of Investigations Dermatology* **99,** 697–702.

Korge B.P., Gan S.-Q., McBride O.W., Mischke D. & Steinert P.M. (1992b) Extensive size polymorphism of the human keratin 10 chain resides in the C-terminal V2 subdoamin due to variable numbers and sizes of glycine loops. *Genetics* **89,** 910–4.

Langbein L., Ileid H.W., Moll I. & Franke W.W. (1993) Molecular characterization of the body site-specific human epidermal cytokeratin 9: cDNA cloning, amino acid sequence, and tissue specificity of gene expression. *Differentiation* **55,** 57–71.

McKusick V.A. (1992) *Mendelian Inheritance in Man*, 10th edn. Johns Hopkins University Press, Baltimore, MD.

Moll R., Franke W.W., Schiller D.L., Geiger B. & Krepler R. (1982) The catalogue of human cytokeratins: patterns of expression in normal epithelia, tumours and cultured cells. *Cell* **31,** 11–24.

Moll I., Heid H., Franke W.W. & Moll R. (1987) Distribution of a special subset of keratinocytes characterized by the expression of cytokeratin 9 in adult and foetal human epidermis of various body sites. *Differentiation* **33,** 254–65.

Ochman H., Gerber A. & Hartl D.L. (1988) Genetic application of an inverse polymerase chain reaction. *Genetics* **120,** 621–3.

Paradis M. (1992) Footpad hyperkeratosis in a family of Dogue de Bordeaux. *Veterinary Dermatology* **3,** 75–8.

Reis A., Hennies H.-C., Langbein L. *et al.* (1994) Keratin 9 gene mutations in epidermolytic palmoplantar keratoderma (EPPK). *Nature Genetics* **6,** 174–9.

Skalli O., Chou Y.-H. & Goldman R.D. (1992) Intermediate filament: not so tough after all. *Trends in Cell Biology* **2**, 308–12.

Steinert R.M. & Roop D.R. (1988) Molecular and cellular biology of intermediate filaments. *Annual Review of Biochemistry* **57,** 593–639.

Torchard D., Blanchet-Bardon C., Serova O. *et al.* (1994) Epidermolytic palmoplantar keratoderma cosegregates with a keratin 9 mutation in a pedigree with breast and ovarian cancer. *Nature Genetics* **6,** 106–10.

Uberbacher E.C. & Mural R.J. (1991) Locating protein coding regions in human DNA sequences by a multiple sensor-neural network approach. *Proceedings of the National Academy of Sciences, USA* **88,** 11261–5.