Original Article

Cytogenet Genome Res 115:107–114 (2006)
DOI: 10.1159/000095229

Cytogenetic and Genome Research

Molecular characterization of the equine collagen, type IX, alpha 2 *(COL9A2)* gene on horse chromosome 2p16→p15

C. Böneker[a] H. Kuiper[a] C. Drögemüller[a] B.P. Chowdhary[b] O. Distl[a]

[a]Institute for Animal Breeding and Genetics, University of Veterinary Medicine Hannover, Hannover (Germany)
[b]Department of Veterinary Integrative Biosciences, College of Veterinary Medicine and Biomechanical Sciences, Texas A & M University, College Station, TX (USA)

Manuscript received 30 January 2006; accepted in revised form for publication by H. Hameister, 24 March 2006.

Abstract. The mammalian collagen, type IX, alpha 2 gene *(COL9A2)* encodes the alpha-2 chain of type IX collagen and is located on horse chromosome 2p16→p14 harbouring a quantitative trait locus for osteochondrosis. We isolated a bacterial artificial chromosome (BAC) clone containing the equine *COL9A2* gene and determined the complete genomic sequence of this gene. Cloning and characterization of equine *COL9A2* revealed that the equine gene consists of 32 exons spanning approximately 15 kb. The *COL9A2* transcript encodes a single protein of 688 amino acids. Thirty two single nucleotide polymorphisms (SNPs) equally distributed in the gene were detected in a mutation scan of eight unrelated Hanoverian warmblood stallions, including one SNP that affects the amino acid sequence of COL9A2. Comparative analyses between horse, human, mouse and rat indicate that the chromosomal location of equine *COL9A2* is in agreement with known chromosomal synteny relationships. The comparison of the gene structure and transcript revealed a high degree of conservation towards the other mammalian *COL9A2* genes. We chose three informative SNPs for association and linkage disequilibrium tests in three to five paternal half-sib families of Hanoverian warmblood horses consisting of 44 to 75 genotyped animals. The test statistics did not reach the significance threshold of 5% and so we could not show an association of *COL9A2* with equine osteochondrosis.

Multiple epiphyseal dysplasia (MED) is an osteochondrodysplasia characterized clinically by mild short stature and early-onset degenerative joint disease and radiographically by epiphyseal hypoplasia/dysplasia (Unger et al., 2001). MED is genetically heterogeneous, with autosomal dominant cases resulting from mutations in at least three genes: the cartilage oligomeric matrix protein gene *(COMP)* causes EDM1, the *COL9A2* and *COL9A3* genes of type IX procollagen cause EDM2 and EDM3 (Unger et al., 2001). EDM2 manifests in childhood and leads to pain and stiffness of several joints with repeatedly free articular bodies resulting in locking of the joint in some adult patients (Fiedler et al., 2002). Affected individuals are prone to the development of early degenerative joint disease (Fiedler et al., 2002). *COL9A2* (collagen, type IX, alpha 2) encodes the alpha-2 chain of type IX collagen. Type IX collagen is a cartilage-specific fibril-associated collagen. The human *COL9A2* gene consists of 32 exons spanning about 16.8 kb and was physically mapped to chromosome HSA1p33→p32.33 starting at 40,435,256 bp and ending at 40,452,032 bp (Perala et al.,

We gratefully acknowledge the German Research Council, DFG, Bonn, (DI 333/12-1) for their financial support of this study.

Request reprints from Ottmar Distl
Institute for Animal Breeding and Genetics
University of Veterinary Medicine Hannover
Bünteweg 17p, DE–30559 Hannover (Germany)
telephone: +49-511-953-8875; fax: +49-511-953-8582
e-mail: ottmar.distl@tiho-hannover.de

1993; Warman et al., 1994). A mutation in *COL9A2* in a Dutch Epiphyseal Dysplasia, Multiple 2, kindred showed linkage with DNA markers in the region of 1p32 (Muragaki et al., 1996). In 12 affected members from a family with a similar phenotype, they also identified a G-to-C mutation in the last nucleotide of exon 3 of *COL9A2* (Fiedler et al., 2002). In another study two families with distinctive oligoepiphyseal forms of multiple epiphyseal dysplasia were described in which heterozygosity for different mutations in *COL9A2* was found leading to skipping of exon 3 from *COL9A2* mRNA (Holden et al., 1999). A mutation in *COL9A2* in Finnish individuals with sciatica and radiologically documented intervertebral disc disease was also identified (Annunen et al., 1999).

Therefore *COL9A2* seems to be a suitable candidate gene for multiple epiphyseal dysplasia, sciatica and intervertebral disc disease in various mammalian species. There are similarities in clinical signs between EDM2 and osteochondrosis (OC). OC is a developmental orthopaedic disorder found in growing animals of many domestic species including horses (Olsson, 1978). In a recent study, the prevalence of OC in horses regarding the distal interphalangeal, proximal interphalangeal, fetlock or hock joints was 32% (Stock et al., 2005a). The previously reported heritability estimates of OC in Warmblood horses ranged between $h^2 = 0.10–0.34$ (KWPN, 1994; Willms et al., 1999; Pieramati et al., 2003; Stock et al., 2005b). A whole genome scan in Hanoverian warmblood horses revealed a quantitative trait locus (QTL) for equine osteochondrosis on horse chromosome (ECA) 2p16→p14 (Böneker et al., submitted). Due to its location and function, *COL9A2* appears as a well suited positional candidate gene for equine OC.

In this report we provide the cloning, genomic organization and the complete sequence of the equine *COL9A2* gene, respectively. Additionally, we present data on new single nucleotide polymorphisms (SNPs) in this gene and an association analysis with equine OC.

Materials and methods

Cloning and sequencing the equine COL9A2 gene

For the isolation of an equine BAC clone containing the *COL9A2* gene the equine BAC library CHORI-241 was initially screened with a ^{32}P-labeled insert of a human IMAGE cDNA clone (IMAGp998 A1610157) provided by the German Human Genome Resource Center/ Primary Database (http://www.rzpd.de/) from the orthologous human gene according to the CHORI protocols (http://www.chori.org/bacpac/). DNA from clone CH241-67C2 was isolated using the Qiagen Midi plasmid kit (Qiagen, Hilden, Germany) and insert size was determined by pulsed field gel electrophoresis. BAC DNA was restricted with different enzymes, separated on 0.8% agarose gels. Fragments were cloned into the polylinker of pGEM-4Z (Promega, Mannheim, Germany). Recombinant plasmid DNA was sequenced with the ThermoSequenase kit (Amersham Biosciences, Freiburg, Germany) and a LICOR automated sequencer (LI-COR, Lincoln, NE, USA). After sequencing a collection of plasmid subclones, remaining gaps were closed by a primer walking strategy until both strands were completely sequenced. Sequence data were analyzed with Sequencher 4.2 (GeneCodes, Ann Arbor, MI, USA).

Repetitive elements were detected with Repeatmasker 3.1.0 (http://www.repeatmasker.org/cgi-bin/WEBRepeatMasker). The genomic structure of equine *COL9A2* was determined by using the genomic DNA sequence as query in BLASTN (basic local alignment search tool nucleotide) analyses of the equine expressed sequence tag (EST) databases (http://www.ncbi.nlm.nih.gov/BLAST/). Equine 3′-ESTs (Accession nos. CX597674, CX605327 and CX600538) were used to determine the 3′ end of equine *COL9A2*. When no corresponding equine EST could be detected, the human *COL9A2* reference mRNA (Accession no. NM001852) was used to annotate the *COL9A2* exons on the genomic sequence. For the exact localization of the exon/intron boundaries the mRNA-to-genomic alignment program Spidey (http://www.ncbi.nlm.nih.gov/IEB/Research/Ostell/Spidey/index.html) was used. GC content was calculated with the EBI toolbox CpG Plot/CpGreport using a 300-bp window (http://www.ebi.ac.uk/Tools/sequence.html). Similarity between the coding sequence of horse, human (Accession no. NM001852), mouse (Accession no. NM007741) and rat (Accession no. XM342903) was calculated with the EBI toolbox Align. Protein alignment was done with EBI toolbox ClustalW. The dog genome (Boxer genome assembly 2.1, http://www.ncbi.nlm.nih.gov/mapview/map_search.cgi?taxid=9615) could not be aligned because of annotation errors in the predicted mRNA (Accession no. XM_539576) and the predicted protein (Accession no. XP_539576).

Mutation and association analysis

To identify variations within the equine *COL9A2* sequence, exons with flanking regions were PCR amplified and sequenced from eight unrelated Hanoverian Warmblood stallions. PCR primers for the amplification of *COL9A2* exons with flanking sequences are given in Table 1. PCR primers were developed with the Primer 3 program (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi). The PCR reactions were performed in a total of 30 μl containing 10 ng of genomic DNA as template, 10 pmol of each primer and 1 U *Taq* polymerase (MP Biomedicals, Eschwege, Germany). After a 5-min initial denaturation at 95°C, 35 cycles of 45 s at 94°C, 45 s at 60°C, and 1 min at 72°C were performed in PTC 100™ or PTC 200™ thermocyclers (MJ Research, Watertown, MA, USA). The obtained PCR products were directly sequenced with the DYEnamic ET Terminator kit (Amersham Biosciences) and a MegaBACE 1000 capillary sequencer (Amersham Biosciences), using the PCR primers as sequencing primers.

We first tested eight unrelated stallions for heterozygosity of the SNPs developed. Then we chose the most informative SNPs for these stallions which were located in intron 4 (SNP_1), exon 12 (SNP_2) and exon 18 (SNP_3). Half-sib families with heterozygous stallions were then used for association analyses. So we could include three families (71 animals genotyped whereof 36 were progeny) for SNP_1, four families (44 animals genotyped whereof 23 were progeny) for SNP_2 and five families (75 animals genotyped whereof 35 were progeny) for SNP_3. The polymorphism information content (PIC) of these SNPs was 0.46, 0.36 and 0.45. Phenotypes were classified as affected by OC in fetlock or hock joints or not-affected. The prevalence for fetlock OC was about 50% and for hock OC about 30% in these families. Briefly, irregular bone fragments at the predilection sites of fetlock and hock joints were considered as signs of OC. Predilection site in fetlock joints was according to Kroll et al. (2001) the sagittal ridge of the 3rd metatarsal/metacarpal bone. Intermediate ridge of the distal tibia and lateral trochlea tali were the predilection sites for hock OC. Statistical analyses were performed using SAS/Genetics (Statistical Analysis System, version 9.1.3, Cary, NC, USA) to calculate PIC and to test for Hardy-Weinberg equilibrium. Phenotypes for OC and genotypes of SNPs of *COL9A2* were tested for genotypic and allelic association using χ^2 tests of the procedure CASECONTROL of SAS/Genetics. In addition, linkage disequilibrium tests and haplotype trait association tests were performed using the procedures FAMILY and HAPLOTYPE of SAS/Genetics. Significance of test statistics was assumed for error probabilities of $P < 0.05$.

Fluorescence in situ hybridization

The 220-kb equine BAC clone containing the equine *COL9A2* gene was labelled with digoxigenin by nick translation using a Dig-Nick-Translation-Mix (Roche Diagnostics, Mannheim, Germany). Fluorescence in situ hybridization (FISH) on GTG-banded equine chromo-

Table 1. PCR primers for the amplification of *COL9A2* exons

Exon	Primer F sequence (5′–3′)	Primer R sequence (5′–3′)	Product length (bp)
1	CCGGGCTTCTCAGGTGAC	GCTCATCCACCTCCTTTCAC	458
2	CCTTTCCCAGTGGAGCTG	CAGGGCGAGAGGTTCAGG	481
3–4	GTGAGGAGAGGAGCATGTGG	CAAACAGAACCCCTCTCTGC	510
5–7	TCATGGAGTGTGGTGAAAGG	AGTCCTCCGGCAAGTCTACC	588
8–9	CTCTGGGTGCCTGTTCCTAC	GGAGACGTGTGGAGATGGAG	618
10–11	CCGGTAAGTGTCCTCAGACC	TAGGTTCGTGGCTCCTTCTG	548
12–13	CTCACATTGAGGGCTTCGAG	CCAGAGCAGGGAAATGACTC	595
14–16	GTCGGAGAGGGAATCATGC	CGCCCCTGTACCAGTCAC	615
17	TTAACACACGTCTTGCTCAGG	AGTGTCCCAGATTGCTGGAG	565
18–19	ATGACCAGACCCACTTCCTG	GCGACTGTCCAAATGATGAG	679
20–21	ATCTGGACTGGGGATGTGAG	CTCACCCAGGCCCCATAC	679
22–23	TAGCCAGGTGTGGAAAGGAG	TAGGGCCCTGAGGAGAAAAG	558
24–25	CTCAAGGACGTCAGGTAATGC	AATGGAGAACAACGGGTCAG	587
26–27	CCGATCTCCCGCTTTCTG	AGAGCGGGGTTGAGAGGAC	528
28–29	CCTCTCAACCCCGCTCTG	TCCAGGCTCTCAGGCAAG	620
30	AAGGAGTTGGGGACCTGTG	TGGGGGAAAGTGTCTGTACC	521
31	GCAGTCTTTTCCAACGTTCC	TCACGGGCAATTATCCATTAC	522
32	AATGGATAATTGCCCGTGAC	GACACTTGCCCTGTCTGTTG	584

somes (Bowling et al., 1997) was performed using 750 ng of digoxigenin-labelled BAC DNA. 20 μg sheared total equine DNA and 10 μg salmon sperm were used as competitors in this experiment. After hybridization overnight, signal detection was performed using a Rhodamine Detection Kit (Qbiogene, Heidelberg, Germany). The chromosomes were counterstained with 4′,6-diamidino-2-phenylindole (DAPI) and embedded in Vectashield Mounting Medium (Vector Laboratories, Burlingame, CA, USA). Thirty metaphases that were previously photographed were re-examined after hybridization with a Zeiss Axioplan 2 microscope equipped for fluorescence.

Radiation hybrid (RH) mapping

A pair of primers was designed using the SP6 sequence of the BAC clone for the *COL9A2* gene and the Primer 3 program (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) (5′-TTTGCAGGCTTTTACAGAATG-3′ and 5′-TTACTACGTCTCCGTGAAGC-3′) generating a PCR product of 185 bp. Amplification of the marker was tested using standard conditions on the Texas A&M University equine 5,000-rad hybrid panel (Chowdhary et al., 2002). The PCR reactions were performed in a total of 20 μl containing 25 ng of RH cell line DNA as template, 15 pmol of each primer and 0.75 U Taq polymerase (MP Biomedicals). After a 4-min initial denaturation at 94°C, 35 cycles of 45 s at 94°C, 45 s at 59°C, and 45 s at 72°C were performed in PTC 200™ thermocyclers (MJ Research). PCR products were separated on a 1.5% agarose gel and visualized under UV light. Results were scored in terms of present, absent or ambiguous in the 92 hybrid cell lines. The typing data were incorporated into the latest radiation hybrid map made of 861 equine markers typed previously on the equine RH_{5000} panel (Chowdhary et al., 2003), using the two-point analysis (http://equine.cvm.tamu.edu/cgi-bin/ecarhmapper.cgi) of RHMAPPER-1.22 (Slonim et al., 1997).

Results and discussion

Analysis of the genomic organization of equine COL9A2

A human *COL9A2* cDNA clone was used to screen a genomic equine BAC library and two positive clones were isolated. Comparative BLAST analysis of the equine BAC clone end sequences with respect to the human genome suggested that the clone CH241-67C2 contained the entire *COL9A2* gene. The CH241-67C2 BAC end sequences were deposited in the EMBL nucleotide database (Accession nos. AM072944, AM072945). BLASTN analysis of the CH241-67C2 SP6 BAC end against build 35.1 of the human genome revealed two significant matches on HSA1p (BLAST E-value $2e^{-28}$). These hits were located in human *ZMPSTE24* over 103 bp (identity 86%) starting at 40,416,415 bp and over 223 bp (identity 83%) starting at 40,416,617 bp. BLASTN analysis of the CH241-67C2 T7 BAC end against build 35.1 of the human genome revealed a significant match on HSA1p (BLAST E-value $1e^{-157}$), located in human MGC27466 gene over 454 bp starting at 40,682,150 bp (identity 90%). A collection of plasmid subclones was sequenced to determine the complete DNA sequence of *COL9A2*. The remaining gaps were closed by a chromosome walking strategy and the complete sequence of the *COL9A2* gene was submitted to the EMBL nucleotide database (Accession no. AJ182498). Using available equine 3′-EST sequences and the human reference cDNA for cDNA-genomic sequence comparisons we detected that the equine *COL9A2* gene consists of 32 exons in agreement with human and murine *COL9A2* genes with exon/intron boundaries that conform perfectly to the GT/AG rule (Fig. 1). The exon sizes range from 24 to 767 bp, the introns between these exons span between 82 and 1,258 bp and the total size of equine *COL9A2* is approximately 15 kb. The repeat content in the 15-kb region of equine *COL9A2* is 16.6%. The fraction of the SINE (8.5%) and LINE (7.0%) elements are nearly balanced. Other repetitive elements constitute 1.2%. The entire equine *COL9A2* gene has an overall GC-content of 59.4% which is similar to the GC-content of human *COL9A2* with 57.2%. Equine *COL9A2* contains two CpG islands (Gardiner-Garden and Frommer, 1987). The first one is in the first intron (GC content of 76.1% over 385 bp) and the second one includes exons 25, 26, 27 and 28 (GC content of 74.4% over 802 bp). Sequence analysis of the genomic region upstream of the puta-

3′-Splice site	Exon	5′-Splice site	Intron phase	Intron size
		-69		
	... (exon 1, >72 bp) ...	CAGATC**gt**aagtcccgcggcc	0	1258 bp
+73		+147		
tgtgtgtgttttc**ag**AGAGGT	... (exon 2, 75 bp) ...	ATCGAC**gt**gagtctaacctga	0	1223 bp
+148		+183		
acttgtatcttgc**ag**GGTGAC	... (exon 3, 36 bp) ...	CCTCCG**gt**gagtgccttttcc	0	82 bp
+184		+246		
atctctgttttca**ag**GGACCT	... (exon 4, 63 bp) ...	ACTGAT**gt**gagtacctgagtg	0	1043 bp
+247		+300		
ctctttcttcccc**ag**GGTCTA	... (exon 5, 54 bp) ...	GTCAAG**gt**aagtggcctgctg	0	123 bp
+301		+336		
ccattggtcttgc**ag**GGCCAG	... (exon 6, 36 bp) ...	CTGCCG**gt**aagtggactccat	0	93 bp
+337		+360		
ttccttctgttcc**ag**GGCCCT	... (exon 7, 24 bp) ...	CCCCCT**gt**aagtcctcaggga	0	317 bp
+361		+414		
cgtctctgttcct**ag**GGACCT	... (exon 8, 54 bp) ...	CCCAAG**gt**gagccccagccac	0	332 bp
+415		+468		
ctctcttgcttcc**ag**GGGGAT	... (exon 9, 54 bp) ...	AAACCG**gt**aagtgtcctcaga	0	106 bp
+469		+516		
cacgtctcctctc**ag**GGCCGC	... (exon 10, 48 bp) ...	TTCCTG**gt**gagaggccgggcc	0	217 bp
+517		+573		
ggattctctttcc**ag**TGTCCA	... (exon 11, 57 bp) ...	GTGAAG**gt**gagctccggaccc	0	138 bp
+574		+627		
ttttcttcgcttc**ag**GGGCAT	... (exon 12, 54 bp) ...	AAGCCG**gt**gagttgcaggcca	0	326 bp
+628		+681		
attcctgccccctt**ag**GGTCCC	... (exon 13, 54 bp) ...	CCGCCG**gt**aaggaaccctccc	0	398 bp
+682		+735		
cttctgtctccat**ag**GGTCCC	... (exon 14, 54 bp) ...	GAGACA**gt**aagtgatttggga	0	98 bp
+736		+789		
tctgcctcttccc**ag**GGTCCT	... (exon 15, 54 bp) ...	TCCCCG**gt**aagtctgctcctg	0	119 bp
+790		+843		
tgtccttcttgcc**ag**GGTGAG	... (exon 16, 54 bp) ...	GACGTG**gt**gagtcctcaggca	0	1136 bp
+844		+897		
tctgtctacttgc**ag**GGCAGC	... (exon 17, 54 bp) ...	GCAACC**gt**aagtggcccaggc	0	426 bp
+898		+951		
tttcctcctgagc**ag**GGTCCC	... (exon 18, 54 bp) ...	GTGAAG**gt**aggagtggggact	0	234 bp
+952		+1005		
ctctgtctctccc**ag**GGCAGT	... (exon 19, 54 bp) ...	CTCGCG**gt**gagtatgggctgg	0	978 bp
+1006		+1050		
agcttctttctcc**ag**GGGGTG	... (exon 20, 45 bp) ...	GACAAG**gt**gaggtgaccccaa	0	409 bp
+1051		+1104		
actgttttcctgc**ag**GGTGAG	... (exon 21, 54 bp) ...	AAGGAG**gt**aagtgtctcctcc	0	607 bp
+1105		+1158		
cactctgtcccccc**ag**GGAGAG	... (exon 22, 54 bp) ...	CAGAAG**gt**aagtgcctggcac	0	293 bp
+1159		+1212		
tccattccatttc**ag**GGCGAC	... (exon 23, 54 bp) ...	CGTCAG**gt**aatgcagggccgg	0	85 bp
+1213		+1284		
ttcttttctcctc**ag**GGCCCT	... (exon 24, 72 bp) ...	GACAAG**gt**accagaaggggct	0	241 bp
+1285		+1320		
tcccgctttctgc**ag**GGCTCC	... (exon 25, 36 bp) ...	GGAGCG**gt**gagtacagcgccg	0	145 bp
+1321		+1365		
gttctccatttgc**ag**GGCGAC	... (exon 26, 45 bp) ...	GAGAAG**gt**gagcgggctgcgc	0	86 bp
+1366		+1398		
ttccttgcgctgc**ag**GGCGAG	... (exon 27, 33 bp) ...	GGACAG**gt**gagcactgcccct	0	185 bp
+1399		+1545		
cctcgcccctggc**ag**CAAGGA	... (exon 28, 147 bp) ...	GTGGCG**gt**gagttggccctca	0	284 bp
+1546		+1600		
cgcctccgctttc**ag**GGCCGA	... (exon 29, 55 bp) ...	TGCAAG**gt**gaggggaggggc	1	308 bp
+1601		+1789		
ccttcctccctcc**ag**AGCAAC	... (exon 30, 189 bp) ...	CCAAGG**gt**gagtgctgttctg	1	791 bp
+1790		+1867		
cttctttctctgc**ag**GAAAAC	... (exon 31, 78 bp) ...	TCCCAG**gt**aagaccttggccc	1	445 bp
+1868		+2634		
ttctgcctttcac**ag**GACTTC	... (exon 32, 767 bp) ...	*TGTA**AATAAA**AATCCCCAACT*		

Fig. 1. Exon/intron boundaries of the equine *COL9A2* gene. Exon sequences are shown in uppercase letters, and intron sequences in lowercase letters. Untranslated regions are shown in italics. The conserved GT/AG exon/intron junctions are shown in boldface type. For exon 32 the polyadenylation signal is shown underlined instead of an exon/intron junction. Position +1 corresponds to the adenine of the translation initiation codon ATG.

```
horse           -MAAAAAPRSLLVLLQVLGLALAQIRGPPGEQGPPGPPGPPGVPGSDGIDGDKGPPGKAG 59
human           MAAATASPRSLLVLLQVVVLALAQIRGPPGERGPPGPPGPPGVPGSDGIDGDNGPPGKAG 60
mouse           -MTAVPAPRSLFVLLQVLWLALAQIRGPPGEPGPPGPPGPPGVPGSDGIDGDKGPPGKVG 59
rat             -MTALPAPRSLLLFLQVLWLALAQIRGPPGEPGLPGPPGPPGVPGSDGIDGDKGPPGKVG 59
                  :* .:****:::***: ************ * ******************:*****.*

horse           PPGPKGEPGQAGPDGPDGKPGTDGLTGAKGEPGPMGIPGVKGQPGLPGPPGLPGPGFAGP 119
human           PPGPKGEPGKAGPDGPDGKPGIDGLTGAKGEPGPMGIPGVKGQPGLPGPPGLPGPGFAGP 120
mouse           PPGSKGEPGKPGPDGPDGKPGIDGLMGAKGEPGPVGTPGVKGQPGLPGPPGLPGPGFAGP 119
rat             PPGSKGEPGKPGPDGPDGKPGIDGLMGAKGEPGPMGTPGIKGQPGLPGPPGLPGPGFAGP 119
                ***.*****:.********** *** ********:* **:*******************

horse           PGPPGPVGLPGEIGITGPKGDPGPDGPSGPPGPPGKPGRPGTIAGLEGSADFLCPTNCPA 179
human           PGPPGPVGLPGEIGIRGPKGDPGPDGPSGPPGPPGKPGRPGTIQGLEGSADFLCPTNCPP 180
mouse           PGPPGPVGLPGEIGTPGPKGDPGPEGPSGPPGPPGKPGRPGTIQGLEGSADFLCPTNCPA 179
rat             PGPPGPVGLPGEIGTPGPKGDPGPEGPSGPPGPPGKPGRPGTIQGLEGSADFLCPTNCPA 179
                ************** ********:****************** ***************.

horse           GAKGPPGLQGVKGHPGKRGALGDSGRQGKPGPKGDVGASGEQGIPGPPGPQGIRGYPGME 239
human           GMKGPPGLQGVKGHAGKRGILGDPGHQGKPGPKGDVGASGEQGIPGPPGPQGIRGYPGMA 240
mouse           GVKGPQGLQGVKGHPGKRGILGDPGRQGKPGPKGDVGASGEQGIPGPPGPQGIRGYPGMA 239
rat             GVKGPQGLQGVKGHPGKRGILGDPGRQGKPGPKGDVGASGEQGIPGPPGPQGIRGYPGMA 239
                * *** ********.**** ***.*:********************************

horse           GPKGETGPRGYKGMLGSIGAAGSPGEEGPRGPPGRAGEKGDVGSQGVRGPQGITGPKGAT 299
human           GPKGETGPHGYKGMVGAIGATGPPGEEGPRGPPGRAGEKGDEGSPGIRGPQGITGPKGAT 300
mouse           GPKGEMGPRGYKGMVGSIGAAGPPGEEGPRGPPGEAGEKGDVGSQGARGPQGITGPKGIT 299
rat             GPKGEMGPRGYKGMVGSIGAAGPPGEEGPRGPPGRAGEKGDVGSQGARGPQGITGPKGTT 299
                ***** **:*****:*:***:*.***********.****** ** * *********** *

horse           GPPGIDGKDGTPGTPGVKGSAGQAGRPGNPGHQGLAGVPGQPGTKGGPGDKGEPGQQGLP 359
human           GPPGINGKDGTPGTPGMKGSAGQAGQPGSPGHQGLAGVPGQPGTKGGPGDQGEPGPQGLP 360
mouse           GPPGIDGKDGTPGIPGMKGSAGQVGRPGSPGHQGLAGVPGQPGTKGGPGDKGEPGQQGLP 359
rat             GPPGIDGKDGTPGIPGMKGSAGQVGRPGSPGHQGLAGVPGQPGTKGGPGDKGEPGQQGFP 359
                *****:******* **:******.*:**.*********************:**** **:*

horse           GFSGPPGKEGEPGPRGEIGPRGIVGQKGDQGERGPVGQPGPQGRQGPKGEQGPPGIPGPQ 419
human           GFSGPPGKEGEPGPRGEIGPQGIMGQKGDQGERGPVGQPGPQGRQGPKGEQGPPGIPGPQ 420
mouse           GVSGPPGKEGEPGPRGEIGPQGIMGQKGDQGERGPVGQPGPQGRQGPKGEQGPPGIPGPQ 419
rat             GISGPPGKEGEPGPRGETGPQGIMGEKGDQGERGPVGQPGPQGRQGPKGEQGSPGIPGPQ 419
                *.*************** **:**:*:************************.*******

horse           GLPGIKGDKGSPGKTGPRGGAGDPGVAGLPGEKGEKGESGEPGPKGQQGVRGEPGYPGPS 479
human           GLPGVKGDKGSPGKTGPRGKVGDPGVAGLPGEKGEKGESGEPGPKGQQGVRGEPGYPGPS 480
mouse           GLPGIKGDKGSPGKTGPRGGVGDPGVAGLPGEKGEKGQSGEPGLKGQQGVRGETGYPGPS 479
rat             GLPGIKGDKGSPGKTGPRGGVGDPGVAGLPGEKGEKGLSGEPGLKGQQGVRGEPGYPGPS 479
                ****:*************** .**************** ***** *********.******

horse           GDAGAPGVQGYPGPPGPRGLAGDRGVPGQPGRQGVAGRDASDQHIVDVVLKMLQEQLAEV 539
human           GDAGAPGVQGYPGPPGPRGLAGNRGVPGQPGRQGVEGRDATDQHIVDVALKMLQEQLAEV 540
mouse           GDIGAPGVQGYPGLPGPRGLVGDRGVPGQPGRQGVVGRAASDQHIVDVVLKMIQEQLAEV 539
rat             GDAGAPGVQGYPGLPGPRGLVGDRGVPGQPGRQGVVGRAASDQHIVDVVLKMIQEQLAEV 539
                ** ********** ******.*:************ ** *:*******.***:*******

horse           AVSAKREALGATGMVGPPGPPGPPGYPGKQGPHGHPGPRGIPGIVGAVGQIGNTGPKGKR 599
human           AVSAKREALGAVGMMGPPGPPGPPGYPGKQGPHGHPGPRGVPGIVGAVGQIGNTGPKGKR 600
mouse           AVSAKREALGAAGMVGLPGPPGPPGYPGKQGPNGHPGPRGIPGIVGAVGQIGNTGPKGKR 599
rat             AVSAKREALGATGMVGLPGPPGPPGYPGKQGPNGHPGPRGIPGIVGAVGQIGNTGPKGKR 599
                ***********.**:* ***************:*******:******************

horse           GEKGDQGDTGRGHPGMPGPPGIPGLPGRPGQAINGKDGDRGSPGAPGEAGRPGLPGPVGL 659
human           GEKGDPGEVGRGHPGMPGPPGIPGLPGRPGQAINGKDGDRGSPGAPGEAGRPGLPGPVGL 660
mouse           GEKGDRGEMGHGHPGMPGPPGIPGLPGRPGQAINGKDGDRGSPGAPGEAGRPGRPGPVGL 659
rat             GEKGDQGEMGRGHPGMPGPPGIPGLPGRPGQAINGKDGDRGSPGAPGEAGRPGRPGPVGL 659
                ***** *: *:****************************************** ******

horse           PGFCEPAACLGASAYASARLTEPGSIKGP 688
human           PGFCEPAACLGASAYASARLTEPGSIKGP 689
mouse           PGFCEPAACLGASAYTSARLTEPGSIKGP 688
rat             PGFCEPAACLGASAYTSARLTEPGSIKGP 688
                ***************:*************
```

Fig. 2. Alignment of the equine COL9A2 protein (688 amino acids) with known orthologous COL9A2 protein sequences. The sequences were derived from GenBank entries with the accession numbers NP_001843 (human COL9A2), NP_031767 (mouse COL9A2) and XP_342904 (rat COL9A2). Identical residues are indicated by asterisks beneath the alignment, while colons and dots represent very similar and similar amino acids, respectively.

Table 2. Single nucleotide polymorphisms within the equine *COL9A2* gene

Location of polymorphic site	Position[a]	Equine cDNA position	Nucleotide polymorphism	Amino acid substitution	Allele frequencies	Genotype frequencies[b]
intron 2	212		C→T		0.94/0.06	87.5/12.5/0.0
intron 4	94		G→A		0.81/0.19	62.5/37.5/0.0
intron 4	117		T→C		0.94/0.06	87.5/12.5/0.0
intron 5	17		G→A		0.75/0.25	**50.0/50.0/0.0**
intron 8	147		G→A		0.75/0.25	**50.0/50.0/0.0**
intron 10	59		G→A		0.88/12.5	87.5/0.0/12.5
intron 11	29		C→T		0.94/0.06	87.5/12.5/0.0
exon 12	18	591	C→T	silent	0.75/0.25	**50.0/50.0/0.0**
intron 12	244		A→G		0.94/0.06	87.5/12.5/0.0
exon 13	18	645	G→A	silent	0.94/0.06	87.5/12.5/0.0
intron 14	10		T→C		0.94/0.06	87.5/12.5/0.0
intron 15	93		C→T		0.94/0.06	87.5/12.5/0.0
intron 16	85		G→A		**0.56/0.44**	**25.0/62.5/12.5**
intron 16	108		C→T		0.75/0.25	**50.0/50.0/0.0**
exon 17	42	885	G→A	silent	0.94/0.06	87.5/12.5/0.0
exon 17	54	897	C→G	silent	0.94/0.06	87.5/12.5/0.0
intron 17	141		C→T		0.94/0.06	87.5/12.5/0.0
exon 18	31	928	A→G	^{310}Thr→^{310}Ala	**0.56/0.44**	**25.0/62.5/12.5**
intron 20	260		A→G		0.94/0.06	87.5/12.5/0.0
intron 24	80		C→T		**0.50/0.50**	**25.0/50.0/25.0**
intron 24	176		C→T		0.75/0.25	**50.0/50.0/0.0**
intron 24	211		T→C		0.75/0.25	**50.0/50.0/0.0**
exon 25	12	1296	G→A	silent	0.94/0.06	87.5/12.5/0.0
intron 26	7		G→T		0.69/0.31	**50.0/37.5/12.5**
intron 29	14		G→A		0.94/0.06	87.5/12.5/0.0
exon 30	170	1770	C→T	silent	0.94/0.06	87.5/12.5/0.0
intron 30	53		G→C		0.94/0.06	87.5/12.5/0.0
intron 30	731		C→T		0.94/0.06	87.5/12.5/0.0
intron 31	229		G→A		0.94/0.06	87.5/12.5/0.0
exon 32	110	1977	C→T	silent	0.69/0.31	**37.5/62.5/0.0**
exon 32/^{UTR}	235	2102	G→A	silent	0.94/0.06	87.5/12.5/0.0
exon 32/^{UTR}	258	2125	C→G	silent	0.94/0.06	87.5/12.5/0.0

[a] Numbering refers to the position of the polymorphic nucleotide within the given exon or intron respectively.
[b] Genotypes are given as number of animals (homozygous for allele 1/heterozygous/ homozygous for allele 2).

tive transcription start site indicated the absence of TATA and CCAAT boxes. The mRNA of equine *COL9A2* (Accession no. AJ182497) contains an open reading frame of 2,064 nt coding for a protein of 688 amino acids. The translation start codon was assigned based on the homology to the human ortholog. The polyadenylation signal AATAAA is located approximately 560 bp downstream of the stop codon resulting in a transcript of approximately 2.6 kb (Fig. 1). In the coding sequence the equine *COL9A2* gene displays 89.5% similarity to the human *COL9A2* gene, 86.0% similarity to the murine *Col9a2* gene and 86.4% similarity to the rat *Col9a2* gene, respectively. In the untranslated regions the sequence similarity between human, mouse, rat and horse is rather low. The equine COL9A2 protein displays 95.9%, 93.9% and 93.9% similarity to the human, murine and rat COL9A2 proteins (Fig. 2).

Polymorphisms within the equine COL9A2 gene

The search for sequence variations within the *COL9A2* gene revealed a total of 32 SNPs shown in Table 2. One SNP located in exon 18 affects the amino acid sequence at position 310 of equine COL9A2 protein. This SNP causes an amino acid change from threonine to alanine. In the human, murine, rat and dog COL9A2 proteins, threonine is found at the corresponding amino-acid position. Quality evaluation of the observed SNPs revealed four SNPs with nearly equal allele frequencies and all three possible genotypes and seven SNPs with nearly equal genotype frequencies for genotypes homozygous for allele 1 and heterozygous genotypes. The eleven SNPs suitable for association studies in horses are shown in bold face in Table 2.

Chromosomal assignment and RH mapping results

For the chromosomal localization of BAC clone CH241-67C2, the BAC DNA was used as probe in a FISH experiment on equine metaphase chromosomes. The assignment localized the *COL9A2* to ECA2p16→p15 (Fig. 3). In order to confirm the FISH mapping of the BAC clone CH241-67C2 an SP6 BAC end marker was typed on the equine RH_{5000} panel. RH analysis of the SP6 BAC end marker showed a retention frequency of 14.1%. *COL9A2* was localized on ECA2, close to marker COR037 with a lod score of

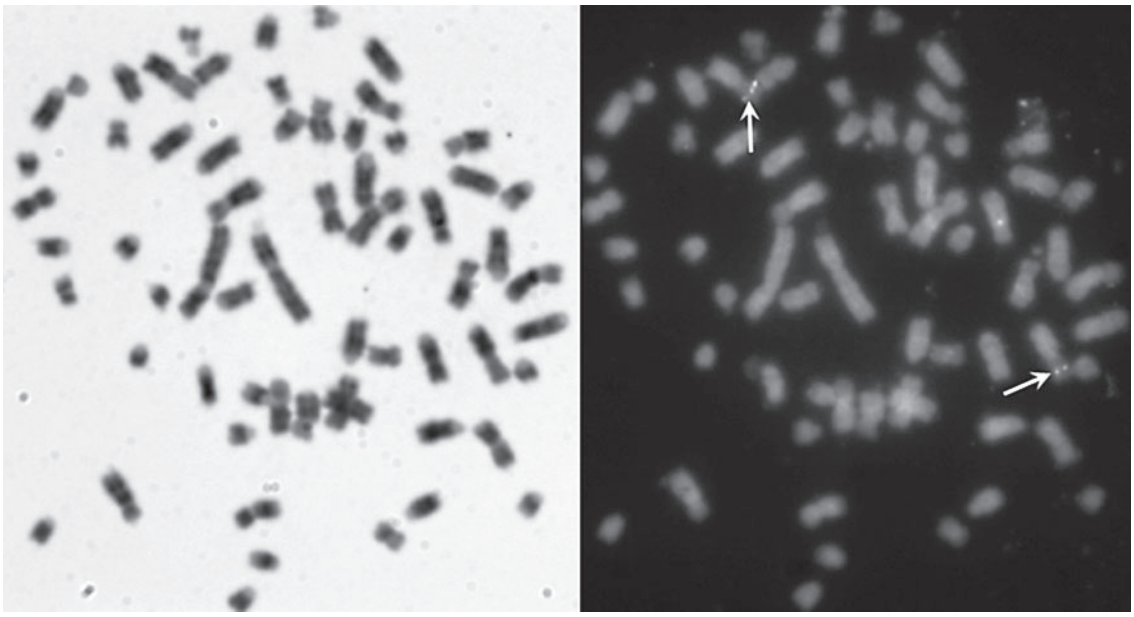

Fig. 3. Chromosomal assignment of the equine *COL9A2* gene by FISH analysis. The digoxigenin labeled BAC clone CH241-67C2 containing the equine *COL9A2* gene was hybridized to GTG-banded metaphase chromosomes of a normal horse. Signal detection was performed using a Rhodamine detection kit. The chromosomes were counterstained with DAPI and embedded in Vectashield mounting medium. Double signals on both ECA2 chromosomes are indicated by arrows.

12. The linked microsatellite marker was previously mapped on ECA2, 47.8 cR from the beginning of *RBBP4* on ECARH02b of the RH_{5000} map (Chowdhary et al., 2003). On the latest male comprehensive linkage map of horse autosomes (Penedo et al., 2005) COR037 was located at 51.3 cM, and at 26.9 cM on the previously published sex-averaged genetic linkage map of the horse autosomes (Swinburne et al., 2006). The assignment of the equine *COL9A2* gene on ECA2 is in good agreement with established syntenies among human chromosome 1p, murine chromosome 4 and equine chromosome 2p (Chowdhary et al., 2003; Swinburne et al., 2006).

Association analysis

All three SNPs genotyped were in Hardy-Weinberg equilibrium. The test statistics for association and linkage disequilibrium analyses with fetlock and hock osteochondrosis were not significant ($P = 0.17$ to 0.97).

Conclusions

In conclusion, our results provide the complete annotated genomic sequence of the equine *COL9A2* gene. We identified 32 SNPs within equine *COL9A2* among them eleven SNPs which seem suitable for association studies. Consistent with the human-horse comparative map the localization of the equine BAC CH241-67C2 containing *COL9A2* confirmed syntenic correspondences between equine chromosome 2p, human chromosome 1p and murine chromosome 4, respectively. The present study provides detailed information towards new polymorphic SNP markers and comparative mapping of the equine genome. An association study with Hanoverian warmblood horses including three to five paternal half-sib families segregating for osteochondrosis could neither show significant linkage disequilibrium nor significant test results for genotypic and allelic association or haplotype trait association.

Acknowledgements

The authors thank Dorothee Müller, Heike Klippert-Hasberg and Stefan Neander for expert technical assistance. Sequences reported in this paper have been deposited in the GenBank database (Accession nos. AM072944, AM072945, AM072946, AM182497, AM182498).

References

Annunen S, Paassilta P, Lohiniva J, Perala M, Pihlajamaa T, et al: An allele of *COL9A2* associated with intervertebral disc disease. Science 285: 409–412 (1999).

Bowling AT, Breen M, Chowdhary BP, Hirota K, Lear T, et al: International system for cytogenetic nomenclature of the domestic horse. Report of the Third International Committee for the Standardization of the Domestic Horse Karyotype, Davis, CA, USA, 1996. Chromosome Res 5:433–443 (1997).

Chowdhary BP, Raudsepp T, Honeycutt D, Owens EK, Piumi F, et al: Construction of a 5000 (rad) whole-genome radiation hybrid panel in the horse and generation of a comprehensive and comparative map for ECA11. Mamm Genome 13:89–94 (2002).

Chowdhary BP, Raudsepp T, Kata SR, Goh G, Millon LV, et al: The first-generation whole-genome radiation hybrid map in the horse identifies conserved segments in human and mouse genomes. Genome Res 13:742–751 (2003).

Fiedler J, Stove J, Heber F, Brenner RE: Clinical phenotype and molecular diagnosis of multiple epiphyseal dysplasia with relative hip sparing during childhood (EDM2). Am J Med Genet 112:144–153 (2002).

Gardiner-Garden M, Frommer M: CpG islands in vertebrate genomes. J Mol Biol 196:261–282 (1987).

Holden P, Canty EG, Mortier GR, Zabel B, Spranger J, et al: Identification of novel pro-alpha2(IX) collagen gene mutations in two families with distinctive oligo-epiphyseal forms of multiple epiphyseal dysplasia. Am J Hum Genet 65:31–38 (1999).

Kroll A, Hertsch B, Höppner S: Entwicklung osteochondraler Veränderungen in den Fessel- und Talokruralgelenken im Röntgenbild beim Fohlen. Pferdeheilkunde 17:489–500 (2001).

KWPN: The frequency and heredity of navicular disease, sesamoidosis, fetlock joint arthrosis, bone spavin, osteochondrosis of the hock. A radiographic progeny study. KWPN (Koninklijke Vereniging Warmbloed Paardenstamboek Nederland), Zeist (1994).

Muragaki Y, Mariman EC, van Beersum SE, Perala M, van Mourik JB, et al: A mutation in *COL9A2* causes multiple epiphyseal dysplasia (EDM2). Ann NY Acad Sci 785:303–306 (1996).

Olsson SE: Osteochondrosis in domestic mammals. Introduction. Acta Radiol Suppl 358:9–14 (1978).

Penedo MC, Millon LV, Bernoco D, Bailey E, Binns M, et al: International Equine Gene Mapping Workshop Report: a comprehensive linkage map constructed with data from new markers and by merging four mapping resources. Cytogenet Genome Res 111:5–15 (2005).

Perala M, Hanninen M, Hastbacka J, Elima K, Vuorio E: Molecular cloning of the human alpha 2(IX) collagen cDNA and assignment of the human *COL9A2* gene to chromosome 1. FEBS Lett 319:177–180 (1993).

Pieramati C, Pepe M, Silvestrelli M, Bolla A: Heritability estimation of osteochondrosis dissecans in Maremmano horses. Livestock Prod Sci 79:249–255 (2003).

Slonim D, Kruglyak L, Stein L, Lander E: Building human genome maps with radiation hybrids. J Comput Biol 4:487–504 (1997).

Stock KF, Hamann H, Distl O: Prevalence of osseous fragments in distal and proximal interphalangeal, metacarpo- and metatarsophalangeal and tarsocrural joints of Hanoverian Warmblood horses. J Vet Med A 52:388–394 (2005a).

Stock KF, Hamann H, Distl O: Estimation of genetic parameters for the prevalence of osseous fragments in limb joints of Hanoverian Warmblood horses. J Anim Breed Genet 122:271–280 (2005b).

Swinburne JE, Boursnell M, Hill G, Pettitt L, Allen T, et al: Single linkage group per chromosome genetic linkage map for the horse, based on two three-generation, full-sibling, crossbred horse reference families. Genomics 87:1–29 (2006).

Unger SL, Briggs MD, Holden P, Zabel B, Ala-Kokko L, et al: Multiple epiphyseal dysplasia: radiographic abnormalities correlated with genotype. Pediatr Radiol 31:10–18 (2001).

Warman ML, McCarthy MT, Perala M, Vuorio E, Knoll JH, et al: The genes encoding alpha 2(IX) collagen (COL9A2) map to human chromosome 1p32.3→p33 and mouse chromosome 4. Genomics 23:158–162 (1994).

Willms F, Röhe R, Kalm E: Genetische Analyse von Merkmalskomplexen in der Reitpferdezucht unter Berücksichtigung von Gliedmaßenveränderungen. Züchtungskunde 71:330–345 (1999).