Identification and expression analysis of cDNA encoding insulin-like growth factor 2 in horses

Kohta KIKUCHI1), Keisuke SASAKI1)†#, Hiroki AKIZAWA1), Hayato TSUKAHARA1), Hanako BAI1), Masashi TAKAHASHI1), Yasuo NAMBO2)##, Hiroshi HATA3) and Manabu KAWAHARA1)

1)Laboratory of Animal Genetics and Reproduction, Research Faculty of Agriculture, Hokkaido University, Hokkaido 060-8589, Japan
2)Equine Science Division, Hidaka Training and Research Center, Japan Racing Association, Hokkaido 057-0171, Japan
3)Field Science Center for Northern Biosphere, Hokkaido University, Hokkaido 060-0811, Japan
† Present: Department of Bioscience, Tokyo University of Agriculture, Tokyo 156-8502, Japan
## Present: Department of Clinical Veterinary Sciences, Obihiro University of Agriculture and Veterinary Medicine, Hokkaido 080-8555, Japan

Abstract. Insulin-like growth factor 2 (IGF2) is responsible for a broad range of physiological processes during fetal development and adulthood, but genomic analyses of IGF2 containing the 5′- and 3′-untranslated regions (UTRs) in equines have been limited. In this study, we characterized the IGF2 mRNA containing the UTRs, and determined its expression pattern in the fetal tissues of horses. The complete equine IGF2 mRNA sequence harboring another exon approximately 2.8 kb upstream from the canonical transcription start site was identified as a new transcript variant. As this upstream exon did not contain the start codon, the amino acid sequence was identical to the canonical variant. Analysis of the deduced amino acid sequence revealed that the protein possessed two major domains, IGF and IGF2-C, and analysis of IGF2 sequence polymorphism in fetal tissues of Hokkaido native horse and Thoroughbreds revealed a single nucleotide polymorphism (T to C transition) at position 398 in Thoroughbreds, which caused an amino acid substitution at position 133 in the IGF2 sequence. Furthermore, the expression pattern of the IGF2 mRNA in the fetal tissues of horses was determined for the first time, and was found to be consistent with those of other species. Taken together, these results suggested that the transcriptional and translational products of the IGF2 gene have conserved functions in the fetal development of mammals, including horses.

Key words: Amino acid sequence, Horse, Insulin-like growth factor 2, Untranslated region

Insulin like growth factor 2 (IGF2) regulates a wide range of important processes, such as cell growth and proliferation, and metabolic activities at the cellular and physiological levels. Moreover, it is also associated with a variety of productive traits, including milk and meat production, and progeny weight [1–6]. Hence, the gene encoding IGF2 is of great interest to animal breeders because of the critical roles performed by this protein [7–9].

The expression pattern of IGF2 is unique in mammals. Although the progeny inherits two sets of chromosomes from its parents through fertilization, IGF2 expresses primarily from the paternal allele, whereas the maternal allele is silenced during fetal development [6, 10]. Furthermore, the expression pattern of IGF2 varies with tissue type. That IGF2 expresses mainly from the paternal allele appears to be universal among species, however [11–13], and therefore, characterization of IGF2 across species may provide further insight into the significance of genomic imprinting in mammals.

Previous studies have isolated and sequenced horse IGF2 cDNA [14, 15], but genetic analyses of the equine IGF2 gene, including examination of the 5′- and 3′-untranslated regions (UTRs) of IGF2 and its protein domains, are scarce. The UTRs, which contain important regulatory elements [16], play important roles in both transcriptional and post-transcriptional regulation of gene expression [17, 18]. Recently, it has been demonstrated that the different 5′-UTRs of mouse IGF2 variants mediated translational control in embryonic stem and neural precursor cells [19]. Moreover, microRNAs that bind to the 3′-UTR of mouse IGF2 have been identified and shown to play an important role in the regulation of IGF2 during placentation [20]. These and other studies highlight the need for further analyses of UTR sequences in order to gain a better understanding of the regulation of IGF2 expression.

Here, we determined the 5′- and 3′-UTRs of equine IGF2 via rapid amplification of cDNA ends (RACE) using Hokkaido native horse fetus, and identified the coding sequence (CDS) of this gene. We then also compared the nucleic acid and predicted amino acid sequences of equine IGF2 with those of other mammals, and examined IGF2
polymorphism in the fetal tissues of Hokkaido native horse and Thoroughbreds. Finally, we performed a quantitative polymerase chain reaction (qPCR) analysis of IGF2 using primary tissues derived from an equine fetus retrieved from an early pregnant mare. Horses are monotocous, in contrast to polytocous species such as mice and pigs; moreover, gestation periods of large livestock, including that of equines, is generally far longer than many other experimental animals, making collection of fetal samples extremely difficult, and as the demand for horsemeat is considerably lower than that of beef, opportunities for obtaining samples from slaughterhouse facilities is limited. In addition, because horse breeding is seasonal, collecting samples of fetuses of synchronized embryonic days is highly labor and cost intensive. Because of these challenges, data on equine genetics in official databases are usually insufficient for determining genetic differences between breeds. In cattle, a single fetal sample has been analyzed for global gene expression using microarrays [21]; thus, although the sample size in the present study is comparatively small, our discovery of a novel splicing variant of equine IGF2 led us to perform comparative analyses between different breeds in an effort to better understand the characteristics of equine IGF2, not only from the perspective of animal breeding but also with respect to the role of this gene in developmental and evolutionary biology.

Materials and Methods

Rapid amplification of cDNA ends (RACE) for determining the full-length IGF2 cDNA

A conceptus (embryonic day: E56) was recovered from a Hokkaido native horse mare that was given a lethal injection following an accidental leg fracture (Fig. 1A). Because the imprinted gene transcripts are enriched in extraembryonic tissues, the yolk sac of the conceptus was homogenized using a BioMasher® (Nippi, Tokyo, Japan), and total RNA was extracted with a Nucleospin® RNA II kit (MACHEREY-NAGEL, Düren, Germany) following the manufacturer’s instructions; RNA was quantitated using a spectrophotometer (NanoDrop NAGEL, Düren, Germany) and stored at –80°C until use. To determine the complete equine IGF2 mRNA sequences, the 5’- and 3’-UTRs were amplified and cloned using a GeneRacer Kit (Invitrogen, Carlsbad, CA, USA), in accordance with the manufacturer’s instructions. The primer sets used for this purpose are shown in Table 1. Based on the known sequence of the annotated equine IGF2 cDNA (GenBank ID: NM_001114539.1), the eqIGF2-114F and GeneRacer 3’ primers were designed for 3’-RACE PCR. For 5’- RACE, the eqIGF2-216R and GeneRacer 5’ primers were used. To further screen more specific amplicons for 5’- and 3’-UTRs, nested PCRs were performed using GeneRacer 5’-nested and GeneRacer 3’-nested primers. The IGF2 CDS was also amplified by GoTaq Polymerase (Promega) using eqGF2 CDS F and P primers designed from within the UTRs, with PCR conditions consisting of 40 cycles of 94°C for 30 sec, 65°C for 1 min, and 72°C for 3 min. The PCR amplicon was ligated into pGEM-T Easy Vector (Promega, Madison, WI, USA). Plasmid DNAs were sequenced with an Applied Biosystems 3130 Genetic Analyzer (Applied Biosystems, FosterCity, CA, USA) and a BigDye Terminator v1.1 Cycle Sequencing Kit (Applied Biosystems). The IGF2 mRNA sequence was then compared to available high-throughput RNA sequencing (RNA-seq) datasets from a fetus at E34 (SRX273058, SRX277447), spermatozoa (SRX154655, SRX154656), testes (SRX277445, SRX273056), and placental villi at 0 day post birth (SRX273055, SRX277444) in the Sequence Read Archives (SRA: https://www.ncbi.nlm.nih.gov/sra) using the Nucleotide BLAST tool.

Comparison of equine IGF2 nucleotide and predicted amino acid sequences between breeds and species

Equine IGF2 CDS was compared between Hokkaido native horse and Thoroughbreds. The cDNAs from the livers of two Thoroughbred fetuses (E79 and E88) were prepared in the same manner as that described above for the Hokkaido native horse fetus. We also compared the predicted amino acid sequences of the equine IGF2 between the two breeds. Protein domain analysis was performed based on two public collections of conserved domain models, namely Pfam (http://pfam.xfam.org/) [22] and SMART (http://smart.embl-heidelberg.de/) [23]. To investigate similarities in the IGF2 CDSs and in the predicted amino acid sequences among different species, the determined equine IGF2 sequence was compared to that of humans (GenBank ID: NM_001127598.2), mice (GenBank ID: NM_0105143.3), rats (GenBank ID: NM_031511.2), cattle (GenBank ID: NM_1747087.3), and pigs (GenBank ID: NM_213883.2). The IGF2 amino acid sequences were also compared among the species using ClustalW algorithm (http://www.genome.jp/tools-bin/clustalw) [24].

qPCR analysis in horse fetal tissues

Total RNA was extracted from different tissues using an RNeasy Mini Kit (QIAGEN, Tokyo, Japan), including the neck, yolk sac, kidney, intestine, heart, leg, lung, umbilical cord, liver, and brain of the Hokkaido native horse fetus, following the procedures described above. Each cDNA was synthesized using ReverTra Ace qPCR RT Master Mix (TOYOBO, Osaka, Japan), after which qRT-PCR was performed using a LightCycler480 (Roche Applied Science, Penzberg, Germany). The reaction mixtures were prepared by adding THUNDERBIRD SYBR qPCR Mix (TOYOBO) at a final concentration of 0.5 μM for each primer. Thermal cycling conditions consisted of one cycle at 95°C for 30 sec (denaturation), followed by 45 cycles at 95°C for 10 sec (denaturation), 61°C for 15 sec (primer annealing), and 72°C for 30 sec (extension). Relative mRNA abundance was calculated by the ΔΔCt method, with GAPDH as the reference gene. To improve reproducibility of the results, the examined sample from the same cDNA source was analyzed in triplicate for each PCR.

Statistical analysis

The expression levels of IGF2 in the equine fetal tissues were compared to the means of the IGF2 expression level in the yolk sac using Student’s t-test. All analyses were performed with StatView software (Abacus Concepts, Berkeley, CA, USA), and a value of P < 0.01 was considered significant.

Results

Identification of a new transcript variant of the equine IGF2 mRNA

The IGF2 mRNA sequence containing the UTRs was determined using RACE-PCR, and was deposited in the DNA Data Bank of Japan
Fig. 1. Determination of the full-length equine *IGF2* mRNA sequence. A: the Hokkaido native horse fetus used in the analyses. The possible embryonic day of conceptus was evaluated as day 56, based on the fetal crown-rump-length. Bar = 10 mm. B: The full-length *IGF2* mRNA was 815 bp, and contained a 5′-untranslated region (UTR) of 64 bp, a coding sequence (CDS) of 546 bp, and a 3′-UTR of 205 bp. The UTR sequences are shown in lower-case letters and red font; uppercase letters indicate the CDS. C: Structural comparison of the identified mRNA with sequences available in public database repositories. The identified complete *IGF2* mRNA sequence had an additional exon located ~2.8 kb upstream from the canonical transcription start site (TSS). An alternative TSS determined by 5′ RACE was also located in the upstream UTR. The upstream 5′-UTR consisted of 58 bp, and did not contain the translation start site (ATG). White and black boxes indicate intron and exon region of *IGF2*, respectively.
The complete IGF2 mRNA was 815 bp, with a 5ʹ-UTR of 64 bp, a CDS of 546 bp, and a 3ʹ-UTR of 205 bp. The 5ʹ- and 3ʹ-UTR sequences of the IGF2 mRNA are shown in Fig. 1B. In addition, the CDS of the equine IGF2 was also identified. The IGF2 mRNA sequence was 55 bp longer than the 5ʹ-UTR sequence deposited previously (GenBank ID: NM_001114539.1). Compared to the equine genomic sequence (GenBank ID: AH006597), the 58 bp sequence identified in the present study was located approximately 2.8 kbp upstream of the canonical transcriptional start site (TSS) (Fig. 1C). Thus, as the nucleotide sequence beginning with “GT” and ending in “AG” is generally regarded as an intron, the 58 bp sequence identified here was derived from the exon coding IGF2 mRNA differently than in the sequence deposited previously. Furthermore, the determined complete equine IGF2 mRNA sequence was compared to the RNA-seq datasets obtained from Thoroughbred fetal tissue, spermatozoa, testes, and placental villi. Consequently, almost every short read was unmatched in the upstream exon, suggesting that the equine IGF2 mRNA sequence identified in the present study represents a novel alternative TSS in horses.

### Table 1. Primer sets used in this study

<table>
<thead>
<tr>
<th>Analysis</th>
<th>Name</th>
<th>Primer sequence (5'-3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>RACE</td>
<td>GeneRacer 5'</td>
<td>CGACTGGACAGGACAGCTGA</td>
</tr>
<tr>
<td></td>
<td>GeneRacer 3'</td>
<td>GCTGTCAACGATAGCTAGTAACG</td>
</tr>
<tr>
<td></td>
<td>GeneRacer 5' Nested</td>
<td>GAGACATGACATGGACTGAAGGAGTA</td>
</tr>
<tr>
<td></td>
<td>GeneRacer 3' Nested</td>
<td>CGCTAGTAAAGGCATGACAGTG</td>
</tr>
<tr>
<td>CDS sequencing</td>
<td>eqIGF2 CDS F</td>
<td>TGTTCGGTTTGGCATACCGCAAG</td>
</tr>
<tr>
<td></td>
<td>eqIGF2 CDS R</td>
<td>GACATGAGGAGTTGGGGGCAA</td>
</tr>
<tr>
<td>qPCR</td>
<td>eqIGF2-103bp-F</td>
<td>GAGATACCCCTCCAGTTTGTCG</td>
</tr>
<tr>
<td></td>
<td>eqIGF2-103bp-R</td>
<td>GAAACAGCACTCTTCCACGAT</td>
</tr>
<tr>
<td></td>
<td>GAPDH 135bp F</td>
<td>CGACACCTTTGTCAG</td>
</tr>
<tr>
<td></td>
<td>GAPDH 135bp R</td>
<td>TCCCTCTTTGCTGGGTGAT</td>
</tr>
</tbody>
</table>

### Fig. 2. Polymorphism analysis using IGF2 sequences from three equine fetuses and the protein structure of equine IGF2. A: Nucleotide and amino acid substitution of equine IGF2 determined in this study. Nucleotide position indicates the number from the transcription start site. The sequences corresponding to the protein domains are aligned under the horse sequence. B: A schematic showing the nucleotide sequence (upper) and protein domains (lower), IIGF and IGF2_C, predicted from the Pfam and SMART databases. Asterisks represent the nucleotide and amino acid substitution site in the equine IGF2 gene, as shown in A.
an amino acid substitution (Leu → Pro) at position 133 (133 aa) of the IGF2 amino acid sequence. Analysis of the IGF2 sequence using two public collections of the conserved domain models Pfam and SMART revealed that the equine IGF2 protein possessed two protein domains, IlGF (CDD ID: cd04368) and IGF2_C (CDD ID: pfam08365), indicating that the detected SNP was present in the IGF2_C domain (Fig. 2B).

Next, the complete equine IGF2 mRNA sequences (from Equus caballus) was compared with those of humans (Homo sapiens), mice (Mus musculus), rats (Rattus norvegicus), cattle (Bos taurus), and pigs (Sus scrofa). Comparisons of the complete mRNA sequences (Supplementary Table 1: online only) revealed that the equine IGF2 was most similar to that of Sus scrofa (85.5%) and least similar to that of Bos taurus (82.0%). However, when only CDS were compared, the equine IGF2 was determined to be most similar to that of Homo sapiens (90.3%) and least similar to that of Mus musculus (84.7%). Comparison of the individual domains IlGF and IGF2_C revealed the number of species-specific amino acid variations in the IGF1 domain to be markedly smaller than in the IGF2_C domain (Fig. 3A). Furthermore, comparisons of the predicted IGF2 amino acid sequences among the different species showed that the equine IGF2 amino acid sequence (from Equus caballus) was more similar to the sequences of the livestock Sus scrofa and Bos taurus, and Homo sapiens, than to those of the rodents Mus musculus and Rattus norvegicus (Fig. 3B).

Expression dynamics of IGF2 mRNA among the fetal tissues

To explore IGF2 expression dynamics during fetal development in horses, we conducted quantitative gene expression analyses in 10 different tissues, consisting of the yolk sac, neck, kidney, intestine, heart, leg, lung, umbilical cord, liver, and brain, which were collected from a Hokkaido native horse fetus (Fig. 4). A criterion for IGF2 expression level was interpreted based on the level of expression in the yolk sac, as most imprinted genes are expressed and play critical roles in extraembryonic tissues [25–27]. Of the 10 tissues that were examined, expression levels of IGF2 were found to be significantly higher (P < 0.01) in the neck, leg, umbilical cord, and liver than in the yolk sac, whereas expression levels in the kidney, intestine, heart, and brain were significantly lower than those in the yolk sac (P < 0.01) (Fig. 4).

Discussion

In this study, we first identified the complete equine IGF2 mRNA sequences, which harbor longer 5'-UTR than did the previously annotated sequence. The newly identified 5'-UTR was predicted to be located in a different exon, composed of 58 nucleotides. It has been reported that equine IGF2 has at least three tissue-specific promoters driving the different transcripts, depending on the respective TSS [14]. The sequence of the IGF2 transcript identified here did not correspond with the sequences annotated previously. In humans and mice, a variety of IGF2 transcripts with differing 5'-UTRs have been identified in a tissue-specific manner [28, 29]. We identified a single variant from the yolk sac, but other variants may yet to be discovered in the embryonic tissues.

Generally, IGF2 possesses two major domains, those of IGF1 and IGF2_C. Based on the predicted amino acid sequence, we determined that these domains were also present in equine IGF2, with the IGF1 domain formed by the amino acids from position 30 to 84, and the IGF2_C domain formed by the amino acids from position 112 to 166 (Fig. 2B). The IGF1 domain is common in the family of proteins that includes insulin, relaxin, and IGFs, which represents evolutionarily related active peptides (Pfam). As shown in Fig. 3A, the species-specific amino acid variations in the IGF1 domain were smaller than those in the IGF2_C domain, indicating that the amino acid sequence in the IGF1 domain was well-conserved among the different species; as such, we would expect that the equine IGF2 possesses homologous physiological functions. During fetal development, IGF2 plays critical roles in fetal and placental growth, in the exchange of material between the fetus and the mother via the placenta, and in the regulation of hematopoiesis in the liver [30–32]. However, these critical physiological functions have been elucidated in only a limited number of species, such as mice, pigs, and humans; for horses, only one study has shown that IGF2 expression increased in the preovulatory follicles of adult mares [33].

To the best of our knowledge, our study represents the first attempt at determining IGF2 mRNA expression levels in multiple fetal tissues of horses. Expression of IGF2 in liver, neck, umbilical cord, and leg tissues was significantly greater than that in other tissues, which was consistent with the expression pattern observed in mice fetal tissues [34]. As with mice, prenatal IGF2 expression was considerably greater than postnatal expression in skeletal muscle and liver tissues of pigs and cattle [35–37]. Such results may reflect the indispensability of IGF2 expression for fetal viability and normal body growth and development, given the vital roles IGF2 plays in fetal and hematopoietic stem cell development [30, 38]. The transcription level of IGF2 during fetal development in mice was shown to affect mRNA expression levels of various angiogenic factors, including Vegf, Flt1, Flt4, Flk1, Ang1, Ang2, Tie1, and Tie2 [39]. Thus, along with structural homology, the expression pattern in the fetal tissues indicated functional significance of IGF2 in the fetal development of horses, as well as in other species. Although it is very difficult to collect sufficient horse fetus samples of the same embryonic days, analyses involving larger sample sizes would provide more precise and in-depth information about the dynamics of IGF2 expression in various tissues, as in the case of mice.

Polymorphism was detected within the IGF2_C domain. Commonly, IGF2 is initially synthesized as a proIGF2 containing the IGF2_C domain. This precursor, proIGF2, undergoes post-translational processing by proprotein convertase enzymes, resulting in variants lacking the IGF2_C domain [40]. The cleavage sites of proprotein convertases on IGF2 are highly conserved among species [41], and were also determined in the present study (Fig. 3A), suggesting that similar IGF2 variants in which the IGF2_C domain is present or absent also exist in horses. Although the IGF2 variants, including proIGF2, have been shown to circulate in fetal and neonatal rats, their specific biological roles in fetal development are not entirely clear. [41]. To further explore the significance of the polymorphism identified here, it will be necessary to clarify the details of the variants present in the equine fetus and assess the physiological functions of proIGF2 containing the IGF2_C domain.

Genome-wide comparisons of horse and human chromosomes show
Fig. 3. Comparison of IGF2 amino acid sequences among different species. A: Predicted IGF2 amino acid sequences in Homo sapiens (Homo), Sus scrofa (Sus), Bos taurus (Bos), Mus musculus (Mus), Rattus norvegicus (Rattus), and Equus caballus (Equus). All sequences contained the same cleavage sites for proprotein convertases [41], which are shown in italic and bold font. B: Comparison of the IGF2 amino acid sequences in humans, mice, rats, cattle, and pigs using the ClustalW algorithm.
strong synteny between these species, and thus equines can serve as model systems for numerous human ailments relating to infertility, inflammatory diseases, and muscle disorders [42, 43]. However, interspecies comparison of the predicted amino acid sequences of IGF2 revealed that the equine sequence was more dissimilar to the human sequence than it was to the sequences of other species (Fig. 3B). Likewise, comparison of nucleic acids containing the UTR sequences indicated that the equine sequence was not closely homologous with the human sequence (Supplementary Table 1). This might be due to lower rates of conservation in the UTR resulting from evolutionary constraints. The length of the 5ʹ-UTR plays important roles in the regulation of IGF2 at the translation level [44–47]. Moreover, the 3ʹ-UTR sequence could also influence mRNA stability and translatability in the regulation of IGF2 roles in the regulation of embryonic growth and lysosomal targeting by the imprinted Igf2/Mpr gene. Genes Dev 1997; 11: 1596–1604. [Medline] [CrossRef]


9. Clark DL, Clark DJ, Bever JE, Dilger AC. Increased prenatal IGF2 expression due to the porcine intron3-G3072A mutation may be responsible for increased muscle mass. J Anim Sci 2015; 93: 2546–2558. [Medline] [CrossRef]


12. Dindot SV, Farin PW, Farin CE, Romano J, Walker S, Long C, Piedrahita JA. Epi-