JARQ 52 (2), 149 - 154 (2018) https://www.jircas.go.jp

Nucleotide Sequences of Porcine α1 and α2 Chains of Type I Collagen cDNA and Their Different Expression Levels in Tissues

Satoshi YANAI[1], Toshimi MATSUMOTO[1,2], Etsumori HARADA[3] and Eiji KOBAYASHI[4]*

[1] JATAFF Institute, Japan Association for Techno-innovation in Agriculture, Forestry and Fisheries (Tsukuba, Ibaraki 305-0854, Japan)
[2] Institute of Crop Science, National Agriculture and Food Research Organization (Tsukuba, Ibaraki 305-8634, Japan)
[3] Laboratory of Veterinary Physiology, Faculty of Agriculture, Tottori University (Tottori, Tottori 680-8553, Japan)
[4] Institute of Livestock and Grassland Science, National Agriculture and Food Research Organization (Tsukuba, Ibaraki 305-0901, Japan)

Abstract

To understand the role of type I collagen in pig health and pork production, we report the complete coding sequences of porcine *COL1A1* and *COL1A2* cDNAs as a first step toward the functional analysis of these genes. The coding sequences of cDNA showed high similarity with those of other mammals (88.6%-95.0% and 86.8%-93.3%, respectively). The deduced amino acid sequences of the porcine COL1A1 and COL1A2 mature proteins showed 91.3%-97.1% and 89.5%-96.0% identity with one of the other mammalian sequences, respectively. Expression analysis showed that these genes were more highly expressed in connective tissues such as skin and back fat than in other tissues. The expression ratio of these two genes is similar in almost all porcine tissues examined, although the *COL1A1* gene in back fat and muscle, and the *COL1A2* gene in testis are more highly expressed than the other gene.

Discipline: Animal industry
Additional key words: cDNA sequence, expression pattern

Introduction

Collagens—the primary components of the extracellular matrix—are the most abundant proteins in humans and animals, and include many members of a complex superfamily (Brown & Timpl 1995, Ricard-Blum 2011). Collagens have Gly-Xaa-Yaa repeats in their primary structure and can be divided into two large subgroups: fibrillar and nonfibrillar collagens (Vuorio & de Crombrugghe 1990, van der Rest & Garrone 1991). Type I collagen is the most abundant and ubiquitously distributed protein in the collagen family. It takes the form of a homotrimer of three α1(I) chains or a heterotrimer comprising two α1(I) chains and one α2(I) chain, which are encoded by the unlinked genes *COL1A1* and *COL1A2*, respectively. The human versions of these genes consist of 51 and 52 exons, respectively (Dalgleish 1997).

Mutations at these two loci result primarily in connective tissue disorders, such as osteoporosis and osteogenesis imperfecta in humans; several connective tissue diseases have also been shown to be related to abnormal collagen metabolism (Dalgleish 1997). A disease affecting connective tissues has been noted as a decisive factor for selection in farm animals. Abnormal collagen metabolism is thought to play an important role in the development of primary inguinal hernia (Rosch et al. 2002). Moreover, leg weakness caused by a connective tissue defect is important for pig health and production (Fan et al. 2009). These disorders may result from defects of collagen molecules that have formed abnormally in cartilage and connective tissue.

In pork production, such meat quality traits as intramuscular fat content (IMF) and tenderness are important for purchasing decisions made by Japanese consumers (Sapp & Knipe 1990). Suzuki et al. have reported a moderate genetic correlation (0.43) between collagen content and IMF (2005). Some studies have also reported that pork collagen

*Corresponding author: e-mail ekoba@affrc.go.jp
Received 27 March 2017; accepted 25 July 2017.

Table 1. Primer sequences of porcine *COL1A1* gene used in this study, target regions, and origin

Primer Name	Primer Sequence (5'-3')	Product Size (bp)	Purpose	Origin
COL1A1-ex1F	CTCCGGCTCCTGCTCCTCTTA	886	PCR & sequencing	human
COL1A1-ex13R	CCATCTGACCAGGAGCTCCAT			human
COL1A1-ex11F	AGCTGGCCTCCCTGGAATGAA	459	PCR & sequencing	human
COL1A1-ex19R	AAGCCAGGAGCACCAGCAATA			human
COL1A1-ex14F	AGAGAGGTCGCCCTGGAGCC	737	PCR & sequencing	human
COL1A1-ex24R	GCCAGTTTTGCCATCAGGACC			human
COL1A1-ex23F	CCCCAAAGGATCTCCTGGTGA	832	PCR & sequencing	human
COL1A1-ex35R	AGCTCCAGTGGGACCAGCAG			human
COL1A1-ex32F	AATGCCTGGTGAACGTGGTGC	788	PCR & sequencing	human
COL1A1-ex42R	TTGCTCCAGAGGGACCTTGTT			human
COL1A1-ex41F	TGCTGGACAGCGTGGTGTGGT	764	PCR & sequencing	human
COL1A1-ex49R	CCTTCTCTTGAGGTGGCTGGG			human
COL1A1-ex48F	CACTGGTGATGCTGGTCCTGT	801	PCR & sequencing	human
COL1A1-ex52R	AGGCGGGAGGTCTTGGTGGTT			human
COL1A1-5'RACE	GGCAGGGCACGGGTTTCCACACG		5'RACE	pig
COL1A1-3'RACE	TCGAGATCCGGGCCGAGGGCAACA		3'RACE	pig

content can be used as an index for determining meat quality (Hamill et al. 2012).

To understand the role of type I collagen in pig health and pork production, we report the complete coding sequences of porcine *COL1A1* and *COL1A2* cDNAs as a first step toward the functional analysis of these genes. We also compared the porcine sequences with those of other animal species, and examined the pattern of gene expression in several tissues.

Materials & methods

1. mRNA preparation

Six castrated male pigs were used in this study. The pigs were produced from three-way crosses (LW × D), fed as per a standard program, and then slaughtered at three months old. At slaughter, nine tissue samples (i.e., heart, kidney, liver, lung, spleen, testis, *longissimus dorsi* skeletal muscle, back fat, skin halfway along the length of the body) were collected. Total RNA was extracted from these tissues using ISOGEN (Nippon Gene, Tokyo, Japan), and then the RNA was reverse-transcribed into cDNA using a ReverTra Ace qPCR RT Kit (Toyobo Bio, Osaka, Japan). Each 20-µl reaction product was diluted with nuclease-free water to a final volume of 40 µl.

2. Nucleotide sequence determination

Polymerase chain reaction (PCR) and direct sequencing were performed using the primers shown in Table 1 and Table 2 for the *COL1A1* and *COL1A2* genes, respectively. Initially, primers designed for the human *COL1A1* and *COL1A2* cDNA sequences were used (accession numbers NM_000088 and NM_000089, respectively). Using the synthesized cDNA as a template, the amplification of gene segments was performed using each primer set with AmpliTaq Gold (Applied BioSystems Japan, Tokyo, Japan). All amplified fragments were purified by excision from 3% agarose gel after electrophoresis. These purified products were used as templates for direct sequencing with the ABI PRISM Dye Terminator Cycle Sequencing Ready Reaction Kit (ThermoFisher, Tokyo, Japan) and the 3130 DNA Sequencer. After the nucleotide sequences of these fragments were determined, 5′- and 3′-noncoding regions of porcine *COL1A1* and *COL1A2* cDNAs were also determined by 5′- and 3′-rapid amplification of cDNA ends (5′- and 3′-RACE) with the primers shown in Table 1 and Table 2, respectively.

3. Homology analysis with other animals

A survey to identify sequence similarity to one of the other animals was performed with the BLAST program (http://www.ncbi.nlm.nih.gov/BLAST). Homology analyses were performed with corresponding sequences of 10 other animals: human (*Homo sapiens*), cattle (*Bos taurus*), horse (*Equus caballus*), sheep (*Ovis aries*), goat (*Capra hircus*), dog (*Canis lupus*), cat (*Felis catus*), mouse (*Mus musculus*), rat (*Rattus norvegicus*), and hamster (*Mesocricetus auratus*). All of the sequences were aligned using the ClustalW program implemented in BioEdit version 7.0.5.2. The phylogenetic tree among the species was constructed using ClustalX (1.83) software.

4. Quantitative real-time RT-PCR analysis

In this study, the expression levels of two α chain genes of type I collagen were analyzed in nine porcine tissues. Quantitative real-time RT-PCR was performed using

Table 2. Primer sequences of porcine *COL1A2* gene used in this study, target regions, and origin

Primer Name	Primer Sequence (5'-3')	Product Size (bp)	Purpose	Origin
COL1A2-ex1F	TACGCGGACTTTGTTGCTGCT	904	PCR & sequencing	human
COL1A2-ex18R	ACCAGTAAGGCCGTTTGCTCC			human
COL1A2-ex17F	TGGTAACGCTGGTCCTACTGG	723	PCR & sequencing	human
COL1A2-ex26R	CCTTTATCACCGTTTTTGCCAG			human
COL1A2-ex25F	CAAGAGGAGAGCCTGGCAACA	798	PCR & sequencing	human
COL1A2-ex37R	CGTTTTCACCCTTAGGCCCTT			human
COL1A2-ex36F	CCTGCTGGCCCCAATGGATTT	588	PCR & sequencing	human
COL1A2-ex42R	ACTACCCACAGCACCAGGAGG			human
COL1A2-ex41F	TTGGTGCTCCTGGTATTCTGG	849	PCR & sequencing	human
COL1A2-ex49R	GAAGGGTCTCAATCTGGTTGTT			human
COL1A2-ex49F	TACGATGGAGACTTCTACAGGG	724	PCR & sequencing	human
COL1A2-ex52R	CAGCACCACCGATGTCCAAAG			human
COL1A2-5'RACE	TCATCACCATCTCTGCCTGGTGG		5'RACE	pig
COL1A2-3'RACE	ACTTGTTGCCGAGGGCAACAGCAGA		3'RACE	pig

Table 3. Primer sequences used for gene expression

Primer Name	Primer Sequence (5'-3')	Product Size (bp)
COL1A1-F	AGTGTGAGGCCACGCATGAGC	404
COL1A1-R	GGTTTCCTGGTCGGTGGGTGA	
COL1A2-F	CTACTGGTGAAACCTGCATTCG	434
COL1A2-R	AAGGATAGGCAGGCGAGATGG	
ACTB-F	CAAGAGGAGAGCCTGGCAACA	798
ACTB-R	CGTTTTCACCCTTAGGCCCTT	
GAPDH-F	CTACTGGTGAAACCTGCATTCG	434
GAPDH-R	AAGGATAGGCAGGCGAGATGG	

StepOnePlus™ Systems (Applied Biosystems, Foster City, CA, USA) in a 20-µl reaction volume containing 2 µl of cDNA, 0.5 µl of each of the forward and reverse primers (Table 3), 7 µl of nuclease-free water, and 10 µl of Fast SYBR Green PCR Master Mix (Applied Biosystems Japan, Tokyo, Japan). In each set of PCR reactions, duplicate cDNA samples were run to confirm the reproducibility of the real-time RT-PCR results. Universal thermal cycling parameters (initial step of 20 s at 95°C, followed by 45 cycles of 3 s at 95°C, 10 s at 60°C, and 20 s at 72°C) were used to quantify the expression of each gene. After the end of the last cycle, a melting curve program was run. A standard curve was generated for both the target gene and endogenous control gene in every PCR run using serial tenfold dilutions of treated cDNAs. The quality of the PCR products was confirmed by melting curve analysis and agarose gel electrophoresis. Finally, quantitative analysis was performed with the standard method, and the results are reported as the relative expression after normalization of the transcript amounts to those of the *ACTB* and *GAPDH* genes. To compare relative expression levels of collagen genes among the nine tissues, we used two genes as internal control. Furthermore, we calculated the expression in all tissues relative to that of the *COL1A1* gene in the lung, which was the lowest expression level identified.

Results and discussion

The 4741-bp and 4836-bp cDNA sequences of the porcine *COL1A1* and *COL1A2* genes, the products of which constitute porcine type I collagen, were obtained by PCR amplification and RACE. These sequences were expected to contain the complete coding regions of the human genes, namely, 4401 bp (1466 amino acids) and 4101 bp (1366 amino acids), respectively. The nucleotide sequences reported in this study are registered in the DDBJ/EMBL/GenBank nucleotide sequence databases with accession numbers LC223106 and LC223107, respectively (sequence data not shown). For the *COL1A1* gene, several partial sequences had been registered in a DNA database (e.g., AF201723.1), but the complete coding sequence of the pig gene had not. As the annotation in DNA databases was incomplete, the sequence information provided by this study should be useful for further work. In contrast, the complete coding sequence of the *COL1A2* gene had already been registered in a DNA database (NM_001243655, Uenishi et al.

Table 4. Nucleotide and amino acid sequence similarity of porcine *COL1A1* and *COL1A2* genes to other mammalian species

	COL1A1		GenBank	COL1A2		GenBank
	Nucleotide	Amino acid	Accession No.	Nucleotide	Amino acid	Accession No.
Human	93.8%	96.8%	NM_0000088.3	92.1%	93.8%	NM_000089.3
Cattle	95.0%	97.1%	NM_001034039.2	92.8%	95.1%	NM_00174520.2
Goat	94.0%	95.8%	XM_005693673.1	92.9%	95.1%	XM_005678936.1
Sheep	93.1%	94.0%	XM_004012773.1	93.0%	95.2%	XM_004007726.1
Horse	94.3%*	98.6%*	XM_005597481.1	93.3%	94.3%	XM_001492939.1
Dog	93.1%	96.7%	NM_001003090.1	92.8%	96.0%	NM_001003187.1
Cat	93.0%	96.8%	XM_003996699.1	93.2%	95.7%	XM_003982764.1
Mouse	89.2%	91.5%	NM_007742.3	86.8%	89.5%	NM_007743.2
Rat	89.6%	91.3%	NM_053304.1	87.1%	90.2%	NM_053356.1
Hamster	88.6%	91.4%	XM_005075850.1	87.8%	90.1%	XM_005082899.1

*: similarity with partial sequence

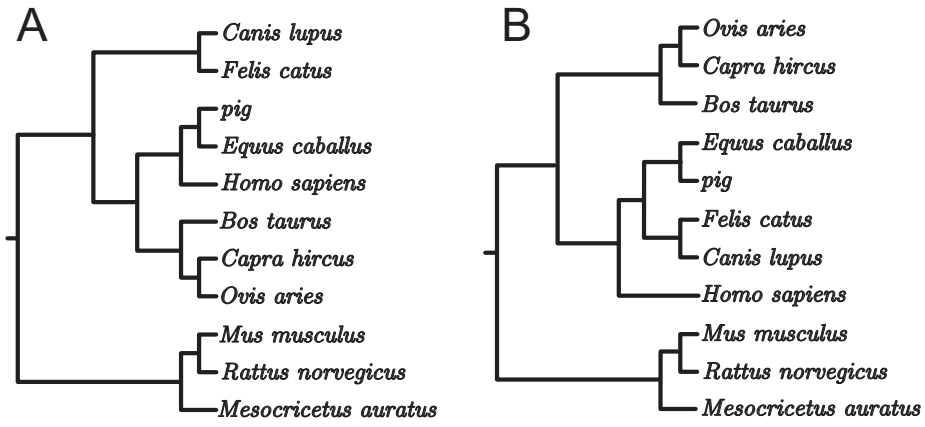

Fig. 1. Phylogenetic trees constructed based on sequences of the *COL1A1* (A) and *COL1A2* (B) genes in pig and nine other animals with scientific names indicated

2004). As the two porcine gene sequences of type I collagen determined in this study correspond to the entire translated region in the human genes, both are considered to provide a foundation for future molecular and genetic research on porcine type I collagen.

Table 4 shows the results of comparing nucleotide sequences for two α chain genes of type I collagen between pig and 10 other animals. The porcine α1 chain of type I collagen cDNA had 93.8% and 89.2% identity with the human and mouse *COL1A1* nucleotide sequences (accession numbers NM_000088.3 and NM_007742.3), respectively. Similarly, the porcine α2 chain of type I collagen cDNA had 92.1% and 86.8% identity with the human and mouse *COL1A2* sequences (accession numbers NM_000089.3 and NM_007743.3), respectively. The large-scale porcine full-length cDNA sequences and draft sequence have already been reported (Uenishi et al. 2007, 2012, Groenen et al. 2012). Compared with our results, it was shown that homologies in the collagen genes are higher than those in the other genes because of its important function. In addition, homology analysis of both α chain genes for type I collagen was performed among pig and the other species using ClustalX (1.83) software. In the phylogenetic tree constructed for pig and the other nine species in the *COL1A1* genes, the farm animals, namely, pig, cattle, goat, sheep, and horse, are classified into a single clade (Fig. 1A). The *COL1A2* genes in the farm and companion animals were categorized into a different clade with rodents, such as mouse, rat, and hamster (Fig. 1B). The deduced amino acid sequences of porcine COL1A1 and COL1A2 showed that all functional domains seem to be conserved in those sequences (Table 4). Moreover, given the high homology with those of other animal species, the deduced amino acid sequences of porcine COL1A1 and COL1A2 are expected to be similar in terms of gene expression control and function. Among those of the farm and companion animals, the porcine gene sequences at the DNA level are highly homologous to cattle, sheep, and goat, which raises interesting questions about the

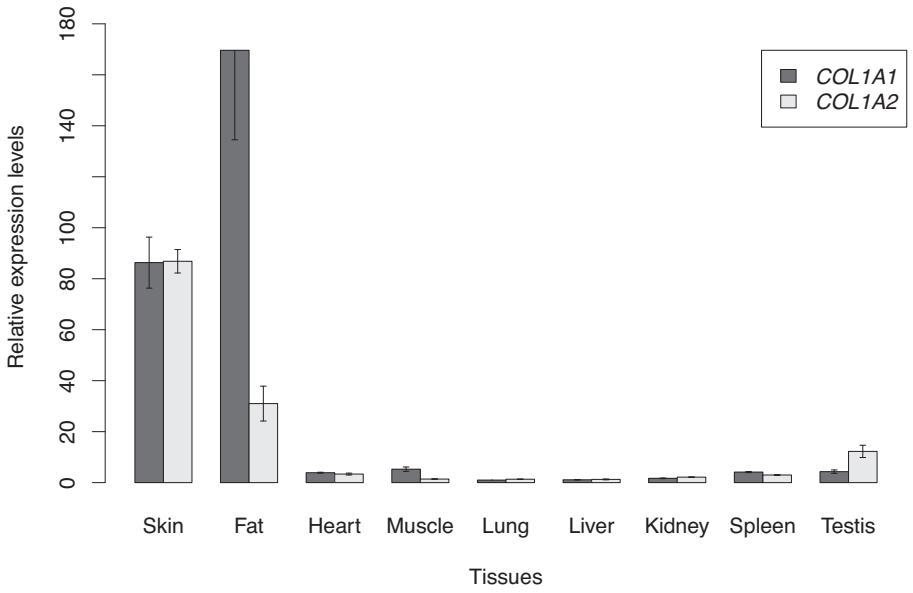

Fig. 2. Relative expression levels with standard deviations of the *COL1A1* and *COL1A2* genes in nine tissues of six young male pigs
Quantitative analysis of expression levels of the *COL1A1* and *COL1A2* genes by real-time PCR. Each expression level is shown as a value relative to the lung *COL1A1* gene expression level, which is designated as 1.

evolution of livestock species.

To examine the expression pattern of the *COL1A1* and *COL1A2* genes in porcine tissues, multiplex quantitative real-time RT-PCR was performed with internal standards *GAPDH* and *ACTB* using the porcine-specific primer sets. Upon comparing the expression levels of both genes in the nine tissues, the relative expression level was extremely high in skin and adipose tissue for both genes (Fig. 2). Because these tissues mainly constitute the animal body, collagen gene expression seemed to be more prominent than other tissues. In other tissues, various genes exert their functions, such as many genes encoding enzymes, and this exertion of function is considered linked to the relatively low expression of collagen genes.

From the results of comparing the nine tissues, no differences between the expression ratio of *COL1A1* and *COL1A2* genes were observed in about half of the tissues examined, although expression of the *COL1A1* gene was high in back fat and muscle, whereas expression of the *COL1A2* gene was high in testis (Fig. 2). This may be meaningful as these tissues have specific functions among the constituents of the body. The different expression ratio of the *COL1A1* and *COL1A2* genes in muscle and back fat from those in other tissues may be related to the pork quality in meat production. Zhang et al. reported different changes of the expression levels in both genes due to various environmental conditions and treatments (2007). The difference in the expression balance of the *COL1A1* and *COL1A2* genes as observed in muscle tissue suggested that the selection of type I collagen α chains in adipose tissue is temporarily altered in the early stage of muscle development, but not at the mature stage.

In this study, the expression levels of the *COL1A1* and *COL1A2* genes in young pig tissues were compared. A report on a study about porcine cultured cells asserted that the expression level of the *COL1A1* gene decreases with age (Martin et al. 1990), while another report described that the balance of *COL1A1* and *COL1A2* in testis changes with the stage of development in mice (He et al. 2005). In addition, in mice with knockout of the myostatin gene, the influence is different between *COL1A1* and *COL1A2* genes (Welle et al. 2009); therefore, the myostatin gene may be important for maintaining the balance in expression between the *COL1A1* and *COL1A2* genes. These results suggest that the *COL1A1* and *COL1A2* genes are controlled by different mechanisms.

Many studies have also been conducted on the relation between expression of the type I collagen gene and pork quality (McBryan et al. 2010). In particular, some reports have described an association between the tenderness of pork and collagen content (Lepetit 2007, Lepetit 2008, Hamill et al. 2012). Since the balance of gene expression in muscle and adipose tissue is different from that in other tissues, an association between the expression ratio of these genes and intramuscular fat mass is also expected. Especially in muscle tissue, the balance of gene expression may affect fat

accumulation and muscle stiffness. Therefore, it is also necessary to examine the relation between the balance of gene expression and stage change or meat quality. Because transcriptome analysis using microarrays (Damon et al. 2012, Hamill et al. 2013) and sequencing (Lim et al. 2016) are not available to study pork quality, it is considered necessary to analyze by focusing on the expression ratio of the two collagen genes studied here. Therefore, further research is needed on the control of the expression of both genes.

Acknowledgement

We would like to thank Mrs. Saori Tsukutani and Mrs. Noriko Sejima for their technical assistance.

References

Brown, J. C. & Timpl, R. (1995) The collagen superfamily. *International Archives of Allergy and Immunol*ogy, **107**, 484-490.

Dalgleish, R. (1997) The human type I collagen mutation database. *Nucleic Acids Research*, **25**, 181-187.

Damon, M. et al. (2012) Comparison of muscle transcriptome between pigs with divergent meat quality phenotypes identifies genes related to muscle metabolism and structure. *PLoS One*, **7**, e33763.

Fan, B. et al. (2009) Large-scale association study for structural soundness and leg locomotion traits in the pig. *Genetics Selection Evolution*, **41**, 14.

Groenen, M. A. M. et al. (2012) Analysis of pig genomes provide insight into porcine demography and evolution. *Nature*, **491**, 393-398.

Hamill, R. M. et al. (2012) Functional analysis of muscle gene expression profiles associated with tenderness and intramuscular fat content in pork. *Meat Science*, **92**, 440-450.

Hamill, R. M. et al. (2013) Transcriptome analysis of porcine M. semimembranosus divergent in intramuscular fat as a consequence of dietary protein restriction. *BMC Genomics*, **14**, 453.

He, Z. et al. (2005) Expression of Col1a1, Col1a2 and procollagen I in germ cells of immature and adult mouse testis. *Reproduction*, **130**, 333-341.

Lepetit, J. (2007) A theoretical approach of the relationships between collagen content, collagen cross-links and meat tenderness. *Meat Science*, **76**, 147-159.

Lepetit, J. (2008) Collagen contribution to meat toughness: Theoretical aspects. *Meat Science*, **80**, 960-967.

Lim, K. S. et al. (2017) Identification of differentially expressed genes in longissimus muscle of pigs with high and low intramuscular fat content using RNA sequencing. *Animal Genetics*, **48**, 166-174.

Martin, M. et al. (1990) Fibronectin and collagen gene expression during in vitro ageing of pig skin fibroblasts. *Experimental Cell Research*, **191**, 8-13.

McBryan, J. et al. (2010) Identification of suitable reference genes for gene expression analysis of pork meat quality and analysis of candidate genes associated with the trait drip loss. *Meat Science*, **86**, 436-439.

Recard-Blum, S. (2011) The collagen family. *Cold Spring Harbor Perspectives in Biology*, **3**, a004978.

Rosch, R. et al. (2002) A role for the collagen I/III and MMP-1/-13 genes in primary inguinal hernia? *BMC Medical Genetics*, **3**, 2.

Sapp, S. G. and Knipe, C. L. (1990) Japanese consumer preferences for processed pork. *Agribusiness*, **6**, 387-400.

Suzuki, K. et al. (2005) Genetic parameter estimates of meat quality traits in Duroc pigs selected for average daily gain, longissimus muscle area, backfat thickness, and intramuscular fat content. *Journal of Animal Science*, **83**, 2058-2065.

Uenishi, H. et al. (2004) PEDE (Pig EST Data Explorer): construction of a database for ESTs derived from porcine full-length cDNA libraries. *Nucleic Acids Research*, **32**, D484-D488.

Uenishi, H. et al. (2007) PEDE (Pig EST Data Explorer) has been expanded into Pig Expression Data Explorer, including 10147 porcine full-length cDNA sequences. *Nucleic Acids Research*, **35**, D650-D653.

Uenishi, H. et al. (2012) Large-scale sequencing based on full-length-enriched cDNA libraries in pigs: contribution to annotation of the pig genome draft sequence. *BMC Genomics*, **13**, 581.

Van der Rest, M. & Garrone, R. (1991) Collagen family of proteins. *The FASEB Journal*, **5**, 2814-2823.

Vuorio, E. & de Crombrugghe, B. (1990) The family of collagen genes. *Annual Review of Biochemistry*, **59**, 837-872.

Welle, S. et al. (2009) Skeletal muscle gene expression after myostatin knockout in mature mice. *Physiology Genomics*, **38**, 342-350.

Zhang, J. et al. (2007) Differential gene expression profile in pig adipose tissue treated with/without clenbuterol. *BMC Genomics*, **8**, 433.