ISSN 1672-9145 **Acta Biochimica et Biophysica Sinica** 2005, **37**(12): 843–850 **CN** 31-1940/Q

cDNA Cloning, Sequence Analysis of the Porcine LIM and Cysteine-rich Domain 1 Gene

Jun WANG, Chang-Yan DENG*, Yuan-Zhu XIONG, Bo ZUO, Lei XING, Feng-E LI, Ming-Gang LEI, Rong ZHENG, and Si-Wen JIANG

Key Laboratory of Swine Genetics and Breeding, Ministry of Agriculture, College of Animal Sciences, Huazhong Agricultural University, Wuhan 430070, China

Abstract LIM domain proteins are important regulators in cell growth, cell fate determination, cell differentiation and remodeling of the cell cytoskeleton by their interaction with various structural proteins, kinases and transcriptional regulators. Using molecular biology combined with *in silico* cloning, we have cloned the complete coding sequence of pig LIM and the cysteine-rich domain 1 gene (*LMCD1*) which encodes a 363 amino acid protein. The estimated molecular weight of the LMCD1 protein is 40,788 Da with a pI of 8.39. It was found to be highly expressed in both skeletal muscle and cardiac muscle. Alignment analysis revealed that the deduced protein sequence shares 86%, 91% and 93% homology with that of its human, mouse and rat counterparts, respectively. The LMCD1 protein was predicted by bioinformatics software to contain a novel cysteine-rich domain in the N-terminal region, two LIM domains in the C-terminal region, nine potential protein kinase C phosphorylation sites, seven casein kinase II phosphorylation sites, a tyrosine kinase phosphorylation site, seven N-glycosylation and N-myristoylation sites and a single potential N-glycosylation site, which is similar to the protein's human counterpart. Phylogenetic tree was constructed by aligning the amino acid sequences of the LIM domain from different species. In addition, four base mutations were detected by comparing the sequences of Large White pigs with those of Chinese Meishan pigs. The G294A mutation site was confirmed by polymerase chain reaction-single-strand conformation polymorphism analysis. Its allele frequencies were studied in five pig breeds.

Key words LIM and cysteine-rich domain 1 (LMCD1); SSCP; PCR; gene expression; data analysis

LIM domain proteins are defined as proteins having a double zinc fingers motif with a consensus amino acid sequence C-X_2-C-X_{16-23}-H-X_2-C-X_2-C-X_{16-23}-C-X_2-C, (where C represents cysteine, and X represents other amino acids) [1]. They are important regulators in cell growth, cell fate determination, cell differentiation and remodeling of the cell cytoskeleton [2]. Bespalova and Burmeister [3] isolated the human complete *LMCD1* coding region and mapped this gene to 3p26-p24 by radiation hybrid mapping. Then the mouse *LMCD1* gene was mapped to the central region of chromosome 6 [3]. The predicted 365-amino acid LMCD1 protein contains a cysteine-rich domain in the N-terminal region and two LIM domains in the C-terminal region. It also has several potential phosphorylation and N-myristoylation sites and a single potential N-glycosylation site. Northern blot analysis detected a major 1.7 kb *LMCD1* transcript in most of the human adult and fetal tissues tested, with highest expression in skeletal muscle. Little or no *LMCD1* expression was found in fetal brain and liver, or in adult brain, liver, thymus, small intestine, or peripheral blood [3]. Recently, Ota *et al*. [4] and Strausberg *et al*. [5] have isolated the human complete *LMCD1* cDNA sequence, which contains six exons and five introns. Furthermore, LIM protein was validated to have action on the control of muscle genes [6]. The presence of LIM domains in the *LMCD1* gene implies it may be involved in skeletal muscle protein-protein interactions [1,7,8].

Received: June 20, 2005 Accepted: October 10, 2005

This work was supported by the grants from the Major State Basic Research Development Program of China (G2000016105), the National High Technology Development Program of China and the Natural Science Foundation of Hubei Province (2005ABA142)

*Corresponding author: Tel, 86-27-87287390; Fax, 86-27-87394184; E-mail, wes2100@webmail.hzau.edu.cn

DOI: 10.1111/j.1745-7270.2005.00113.x

Therefore, the *LMCD1* gene was selected as a candidate gene for pig meat quality traits.

Materials and Methods

Tissue and blood samples

The tissue samples (heart, liver, lung, kidney, spleen, longissimus dorsi muscle, fat, stomach and small intestine) were collected from Large White, Landrace and Chinese local breed Meishan pigs at Jingping Pig Station (Huazhong Agricultural University, Wuhan, China).

The blood samples of Yorkshire, Landrace and Chinese local breed Meishan pigs were collected from Jingpin Pig Station. The Chinese local breeds Exi Hei and Wannanhua pigs were from scattering farms in Enshi Municipality (Enshi, China) and Anhui Province, respectively.

Total RNA extraction and genomic DNA isolation

Total RNA was extracted from different tissues with a Trizol kit (Gibco, Carlsbad, USA). In case where the samples were seriously contaminated with genomic DNA, DNase I treatment on the total RNA was carried out before first-strand cDNA synthesis.

Genomic DNA was isolated from the blood samples of Yorkshire, Landrace, Meishan, Exi Hei and Wannanhua pigs using phenol/chloroform extraction and ethanol precipitation [9].

Primer design

A number of pig expressed sequence tags (ESTs) were initially identified using the cDNA sequence of human *LMCD1* (GenBank accession No. NM_014583) by running a Blast search against the GenBank "EST-others" databases. Two ESTs (GenBank accession No. BF198782 and BF442878) were assembled into one contig with Sequencher 4.14 software. Primer L1 (forward, 5′-AGCCTCTGTCTAAGAAGCAAA-3′; reverse, 5′-CACGGGCTGCTTCTCCTT-3′) was designed by the contig. Primer L2 (forward, 5′-CTCCAAGTATTCCACCCTCACA-3′, locating the contig; reverse, 5′-CCTCAGGATGGCCTTAGCAC-3′, locating the EST) was designed by the contig and the EST (GenBank accession No. C94730). The structure of the amplified sequence covers the coding region and all the exons. Primer L3 (forward, 5′-CCTGGAAGATGATCGGAAAA-3′; reverse, 5′-TGATGGTGTCAAAGGTGGGA-3′) was designed by the porcine *LMCD1* gene sequence (GenBank accession No. AY821789) which we had submitted. L3 was used to validate the correctness of mutation sites obtained by single-strand conformation polymorphism (SSCP) analysis of polymerase chain reaction (PCR) products.

Primers were designed with Primer 5.0 (http://www.premierbiosoft.com).

Reverse transcription (RT)-PCR

Primary cDNA synthesis was processed in a final volume of 25 μl containing 5×reaction buffer (5 μl), 1 μg of total RNA from certain tissue as the template, 0.5 mM of each dNTP, 25 U of RNasin (40 U/μl), 2 μl of 10 μM oligo(dT_{15}) and 200 U of M-MuLV reverse transcriptase (200 U/μl; Promega, USA).

PCR amplification was carried out in a final volume of 25 μl containing standard 1×PCR buffer with Mg^{2+} and 1 U *Taq* polymerase (Biostar International, Toronto, Canada), 0.8 mM of each dNTP, 10 pmol of each primer and 1.0 μl of first-strand cDNA product as the template. The PCR conditions were as follows: denaturation at 94 ºC for 3 min; 94 ºC for 50 s, 57 ºC for 45 s for L1 (or 62 ºC for 50 s for L2), 72 ºC for 50 s, 35 cycles; and an additional extension step at 72 ºC for 10 min.

Cloning of PCR products and sequencing

The PCR products were fractionated on 1.5% (*W*/*V*) agarose gel, and selected bands were purified using a gel extraction kit (Sangon, Shanghai, China). The purified PCR products were ligated into the pGEM-T vector (Promega) and transformed into DH5α competent cells. Bacteria were grown in LB-ampicillin agar. Cloned PCR products were sequenced by Sangon.

mRNA expression analysis

The tissue distribution of pig *LMCD1* mRNA was determined by semiquantitative RT-PCR [10,11]. The house-keeping gene glyceraldehydes-3-phosphate dehydrogenase (*GAPDH*) was used as an internal control on the template level. Primers for *GAPDH* amplification were: forward, 5′-ACCACAGTCCATGCCATCAC-3′; reverse, 5′-TCCACCACCCTGTTGCTGTA-3′. The primer for *LMCD1* was L1.

PCR was carried out in a final volume of 25 μl as above. The conditions for PCR were in the exponential phase of amplification, as judged by the use of different concentrations of cDNA template, which provided a direct correlation between the amount of the amplification products obtained and RNA template abundance in the samples: denaturing at 94 ºC for 4 min; 94 ºC for 50 s, 57 ºC for 45 s and 72 ºC for 50 s, 28 cycles; and an additional extension step for 10 min at 72 ºC. To validate the results,

we repeated the RT-PCR three times.

Prediction and analysis of putative LMCD1 domain

The cDNA sequence prediction was conducted with GenScan software (http://genes.mit.edu/GENSCAN.html). Sequence similarity analysis in GenBank was performed using the Blast 2.1 search tool (http://www.ncbi.nlm.nih.gov/blast/). ClustalW software (http://www.ebi.ac.uk/clustalw/) was used for alignment of multiple sequences. Phylogenetic and molecular evolutionary analyses were conducted using MEGA 3.1 software [12]. To predict the biophysics characteristics of the putative protein of *LMCD1*, software on the ExPASy Proteomics Server (http://au.expasy.org/) was used. The prediction and analysis for the protein structural domain and functional site were performed using Prosite software (http://www.expasy.org/prosite/). The 3-D structure of the putative protein conserved domain was analyzed using the 3-D Conserved Domain Architecture Retrieval Tool of Blast (http://www.ncbi.nlm.nih.gov/blast/).

SSCP analysis [13,14]

L3 amplified DNA product (amplification procedure was the same as L1 except for the anneling temperature was 60 ºC) was mixed with four volumes of formamide loading dye [98% formamide, 0.3% bromophenol blue and xylene cyanol (Sangon), 10 mM EDTA and 7% glycerin (Sunbiotech, Beijing, China)], then denaturized at 98 ºC for 10 min, loaded onto a non-denaturing gel [12% acrylamide/bisacrylamide (29/1, *W/W*), 1×TBE (25 mM Tris, 25 mM boric acid, 0.5 mM EDTA)] containing 5% glycerol (Sunbiotech), and subject to electrophoresis at 4 ºC, 5 W for 10–25 h, depending on the size of the PCR product analyzed and the gel composition. The gels were processed by silver staining.

Results

RT-PCR, cloning of PCR products and sequencing

Amplified cDNA products were fractionated on 1.5% (*W/V*) agarose gel, and clear amplified bands of primer L1 and L2 were obtained. RT-PCR products were cloned into vector pGEM-T and sequenced. Sequencing results showed that the sizes of the L1 and L2 PCR products were 479 bp and 945 bp, respectively (**Fig. 1**).

mRNA expression analysis

Fig. 2 shows that *LMCD1* mRNA was present at high levels in the Longissimus dorsi muscle and heart, and at medium levels in the lung, kidney and spleen. The *LMCD1* mRNA expression level was low in fat and in the stomach, and lower in the liver and small intestine. All expression patterns were, on the whole, in accordance with the results of adult human *LMCD1* mRNA expression [3].

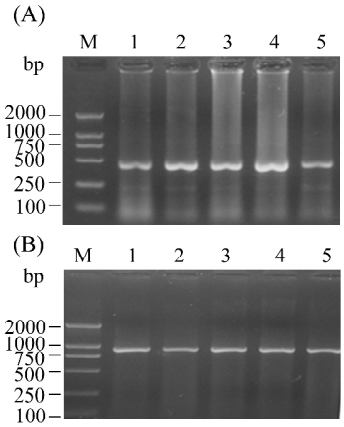

Fig. 1 Reverse transcription-polymerase chain reaction products of each pig breed using primers L1 and L2 respectively
(A) Primer L1 amplified products in various breeds. (B) Primer L2 amplified products in various breeds. M, DL2000 marker; 1 and 2, Meishan; 3 and 4, Large White; 5, Landrace.

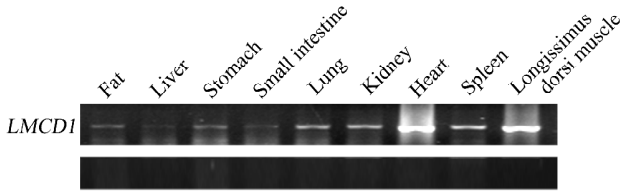

Fig. 2 Reverse transcription-polymerase chain reaction analysis of porcine *LMCD1* expression in different issues
GAPDH, glyceraldehydes-3-phosphate dehydrogenase.

Nucleotide sequence analysis

The sequences amplified by primers L1 and L2 were assembled into a 1216 bp sequence with Sequencher 4.14 software. This sequence was identified and similarity aligned by running a BlastN search against the GenBank "nr" databases. The analysis revealed that the sequence contains the complete coding sequence of *LMCD1*. It was then submitted to the GenBank database with the accession number AY821789. The porcine *LMCD1* nucleotide sequence shared high homology with those of four species

by Blast analysis: mouse (86%), pan troglodytes (91%), rat (82%) and human (89%). Comparative analysis revealed that there was a six-base deletion in the coding sequence of porcine *LMCD1* which results in a two-amino-acid deletion in the porcine protein, compared with that of the human gene at position 825–830 (GenBank accession No. 14277673).

Four putative base substitutions (G294A as shown in **Fig. 3**, C385T as shown in **Fig. 4**, A748G and A1099G) were detected in the exon region by comparing sequences of Large White, Landrace and Meishan pig breeds. A748G substitution changes a codon for Alanine into a codon for Threonine.

Prediction and analysis of protein structural domain and functional site

Similarity comparison for the amino acid sequence of the *LMCD1* gene in four species is shown in **Fig. 5**. The amino acid sequence of porcine LMCD1 shares 86% identity with that of human, 91% with that of mouse and 93% with that of rat. The amino acid sequence of LMCD1 shares significant identity with those of mouse Tes1, mouse Tes2, human LMO6, and human triple-LIM domain protein [3]. Based on the above results, Tes1, Tes2 and several kinds of LIM domain proteins were collected to construct a combined phylogenetic tree by Neighbor-Joining method and Unweighted Pair Group Method with Arithmetic Mean (UPGMA) using MEGA 3.1 software (http://www.megasoftware.net/), as shown in **Fig. 6**. Results revealed that porcine LMCD1 has a closer genetic relationship with mouse and rat LMCD1 than with that of human. All of the LIM domain proteins and the Tes1 and Tes2 genes have a similar location in the phylogenetic tree, only their bootstrap values are different. These results validate the correctness of the current classification of the LIM domain protein.

Primary structure analysis revealed that the molecular weight of the putative LMCD1 protein is 40.788 kDa and its theoretical pI is 8.39. Topology prediction showed there was a novel cysteine-rich domain in the N-terminal region and two idiocratic LIM domain structural sequences in the C-terminal region, but no transmembrane helices

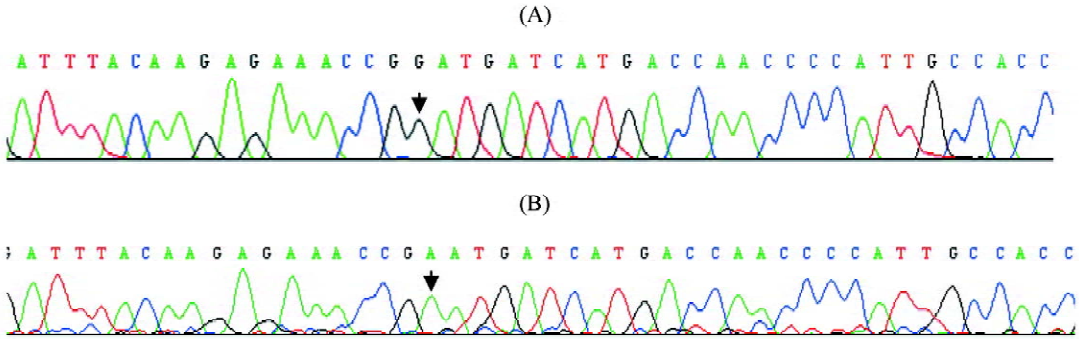

Fig. 3 Comparison of alleles G and A for porcine *LMCD1* gene
(A) Allele G. (B) Allele A.

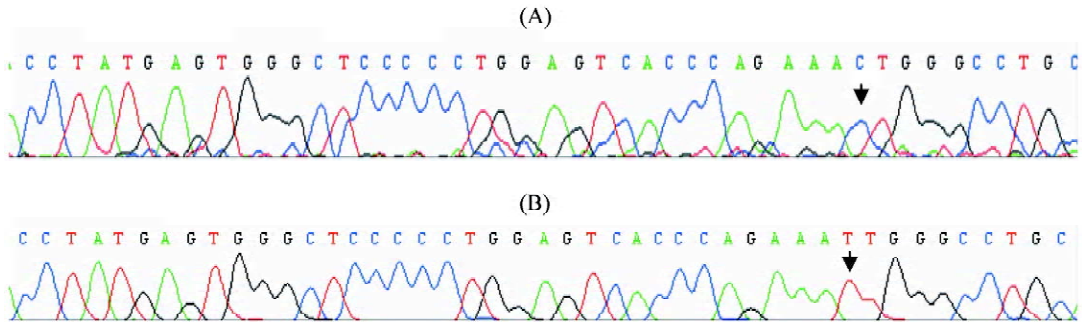

Fig. 4 Comparison of alleles C and T for porcine *LMCD1* gene
(A) Allele C. (B) Allele T.

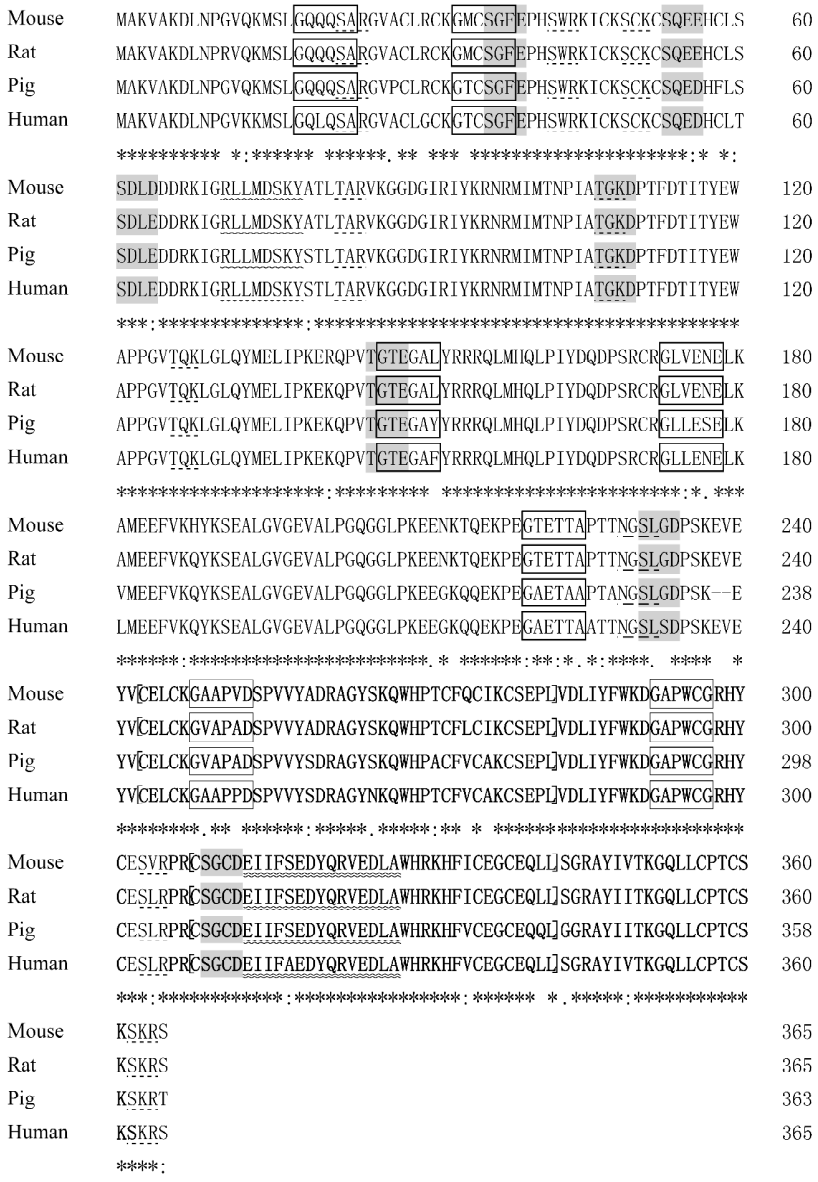

Fig. 5 Alignment of the protein encoded by *LMCD1* in mouse, rat, pig and human

The putative boundaries of known structural domains, including the LIM domain, LIM zinc-binding domains, N-myristoylation sites, N-glycosylation site and several phosphorylation sites. Bold text, LIM domains; square brackets, LIM zinc-binding domains; boxed text, N-myristoylation sites for LMCD1; dotted underline, protein kinase C phosphorylation sites; grey highlight, casein kinase II phosphorylation sites; wavy underline, tyrosine kinase phosphorylation site; dashed underline, N-glycosylation site; wavy double underline, tyrosine sulfation site. *, identical residues identified by the ClustalW program; . and :, similar residues identified by the ClustalW program.

[15,16]. Nine potential protein kinase C phosphorylation sites, seven casein kinase II phosphorylation sites, a tyrosine kinase phosphorylation site, seven N-glycosylation and N-myristoylation sites and a single potential N-glycosylation site were also found by prediction (**Fig. 6**), similar to the structure of human LMCD1 [3]. Based on the single potential N-glycosylation site in the protein, it can be inferred that the LMCD1 protein may be a glycoprotein.

The putative domain of the protein encoded by porcine LMCD1 and the 3-D structure of the conserved domain of this putative protein are shown in **Figs. 7** and **8**, respectively. From these two figures, we can find two LIM zinc-binding domains and a PET domain that is

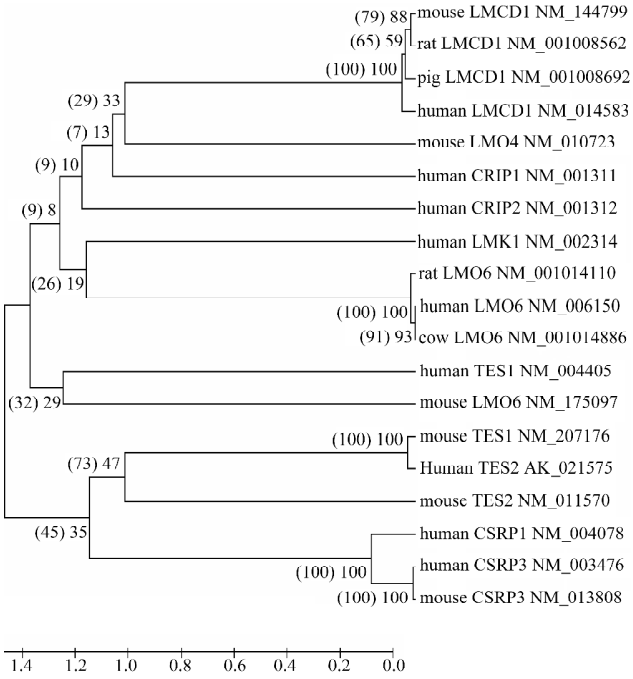

Fig. 6 Phylogenetic tree for several LIM domain proteins

A Neighbor-Joining and Unweighted Pair Group Method with Arithmetic mean (UPGMA) tree was constructed using MEGA 3.1 software. Numbers out of the parentheses are the UPGMA bootstrap values; numbers in parentheses are the neighbor-joining bootstrap values.

Fig. 7 Putative domain of the protein encoded by porcine *LMCD1*

suggested to be involved in protein-protein interactions, which further validates the correctness of the current classification of the putative LMCD1 protein.

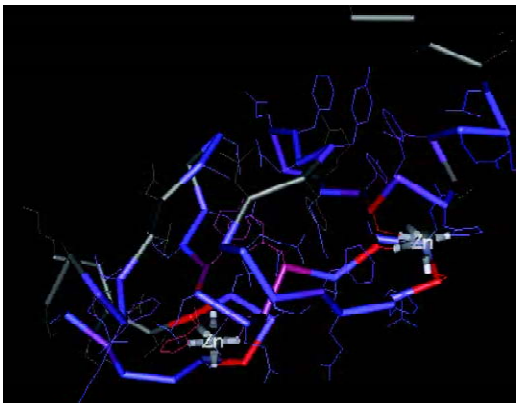

Fig. 8 3-D structure of the putative conserved domain of porcine LMCD1 analyzed by 3-D conserved domain architecture retrieval tool of blast

SSCP analysis

The correctness of the G294A substitute was confirmed by PCR-SSCP. The size of the primer L3 PCR product was 164 bp. The PCR-SSCP results are shown in **Fig. 9**. The distribution of the polymorphism in five different pig breeds is given in **Table 1**. We can conclude that only allele G was found in Large White and Landrace pigs. χ^2 analysis of three genotypes in different pig populations showed that the frequency of genotype was significantly different (χ^2=128.1200>$\chi^2_{0.01(8)}$, P<0.01) in Large White, Landrace, Meishan, Exi Hei and Wannanhua pigs.

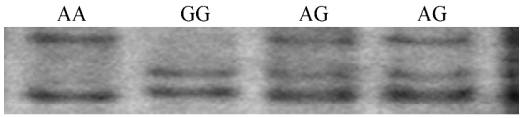

Fig. 9 Results of polymerase chain reaction-single-strand conformation polymorphism analysis of porcine *LMCD1* fragment

Table 1 Distribution of polymerase chain reaction-single-strand conformation polymorphism genotype and allele frequencies in five pig breeds

Breed	Number	Genotype			A allele frequency	G allele frequency
		AA	AG	GG		
Landrace	48	0	0	48	0.00%	100.00%
Large White	28	0	0	28	0.00%	100.00%
Meishan	45	21	22	2	71.11%	28.89%
Exi Hei	45	2	12	31	17.78%	82.22%
Wannanhua	13	0	2	11	7.69%	92.31%

Discussion

Muscle LIM protein is composed of two neighboring LIM domains and a glycine-rich domain [6,17]. It is an important regulator in the development of skeletal muscle and cardiac muscle. According to the current classification of LIM domain proteins, LMCD1 belongs to the group 3 proteins, which contain one or more LIM domains at the C-terminal region [3]. This gene family is characterized by idiocratic LIM domains and a conserved cysteine-rich motif mostly expressed in musculature [3,8].

When a correct gene has been identified, there may be more than one polymorphism within that gene. Polymorphism in the coding sequence that does not change the amino acid (synonymous mutation) is unlikely to have an effect on phenotype [18,19]. Two clues are used to predict whether or not a non-synonymous polymorphism will affect phenotype. First, a mutation that causes a radical change in the amino acid is more likely to affect the properties of the protein than a conservative amino acid substitution. Second, amino acids that are conserved across species are likely to be needed for protein function, so mutations that change them are likely to affect phenotype [19,20]. Because of SNPs (single nucleotide polymorphisms) have high density and stability in genomes. They were widely used in the identification of a functional gene and location of quantitative trait loci (QTL) SNP occurring in the coding region of a gene (cSNP) maybe belong to QTL to affect the level of gene expression and the protein structure, so taking cSNP as a marker would be more available for maker-assisted selection [14,21]. In our study, three cSNPs were detected by aligning the cDNA sequences of different porcine breeds, and the correctness of the G294A substitute was validated by PCR-SSCP. In addition, the distribution of three genotypes, based on the G294A substitute, was studied in five pig breeds. The results showed that allele A did not exist in Large White and Landrace pigs, which may have been affected by long-term breeding and selection or the limited number of animals in this study. It may also be that allele A has a special function that affects porcine meat quality. Further studies must be conducted to confirm whether this site can be regarded as a molecular marker or not.

Bioinformatics analysis showed the LMCD1 protein was highly conserved among the different species in this study. Some potential functional sites predicted in pig LMCD1 were also found in human LMCD1. These results offer some evidence to further understand the function of porcine LMCD1. Two means that constructing phylogenetic tree have different substitution models resulted in dissimilar bootstrap values, all LIM domain protein genes have a similar location in the two trees, which further validates the correctness of the classification of the *LMCD1* gene.

This is the first report on the *LMCD1* gene in pig. We have obtained its complete coding sequence; we will continue our research to obtain its full-length sequence, then carry out a functional analysis. Like other LIM domain proteins with predominant expression in skeletal muscle, the LMCD1 protein might be involved in protein-protein interactions during muscle development and remodeling [1,3,8], so further research based on these primary results is needed.

Acknowledgements

We would like to thank the staff at Jingping Pig Station (Huazhong Agricultural University) and teachers and graduate students at the Key Laboratory of Swine Genetics and Breeding, Ministry of Agriculture for managing and slaughtering experimental animals.

References

1 Morgan MJ, Madgwick AJ. The LIM proteins FHL1 and FHL3 are expressed differently in skeletal muscle. Biochem Biophys Res Commun 1999, 255: 245–250

2 Li HY, Ng EK, Lee SM, Kotaka M, Tsui SK, Lee CY, Fung KP *et al*. Protein–protein interaction of FHL3 with FHL2 and visualization of their interaction by green fluorescent proteins (GFP) two-fusion fluorescence resonance energy transfer (FRET). J Cell Biochem 2001, 80: 293–303

3 Bespalova IN, Burmeister M. Identification of a novel LIM domain gene, LMCD1, and chromosomal localization in human and mouse. Genomics 2000, 63: 69–74

4 Ota T, Suzuki Y, Nishikawa T, Otsuki T, Sugiyama T, Irie R, Wakamatsu A *et al*. Complete sequencing and characterization of 21,243 full-length human cDNAs. Nat Genet 2004, 36: 40–45

5 Strausberg RL, Feingold EA, Grouse LH, Derge JG, Klausner RD, Collins FS, Wagner L. Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences. Proc Natl Acad Sci USA 2002, 99: 16899–16903

6 Jiang YL, Li N, Wu CX. Studies on the molecular biology of myogeneis. Chin J Agric Biotechol 1999, 7: 201–204

7 Morgan MJ, Madgwick AJ. Slim defines a novel family of LIM-proteins expressed in skeletal muscle. Biochem Biophys Res Commun 1996, 225: 632–638

8 Brown S, Biben C, Ooms LM, Maimone M, McGrath MJ, Gurung R, Harvey RP *et al*. The cardiac expression of striated muscle LIM protein 1 (SLIM1) is restricted to the outflow tract of the developing heart. J Mol Cell Cardiol 1999, 31: 837–843

9 Sambrook J, Fritsch EF, Maniatis T. Molecular Cloning: A Laboratory Manual. 2nd ed. New York: Cold Spring Harbor Laboratory Press 1989

10 Kousteni S, Tura-Kockar F, Ramji DP. Sequence and expression analysis of a novel *Xenopus laevis* cDNA that encodes a protein similar to bacterial and chloroplast ribosomal protein L24. Gene 1999, 235: 13–18
11 Liu YG, Xiong YZ, Deng CY. Isolation, sequence analysis and expression profile of a novel swine gene differentially expressed in the Longissimus dorsi muscle tissues from Landrace×Large White cross-combination. Acta Biochim Biophys Sin 2005, 37: 186–191
12 Kumar S, Tamura K, Nei M. MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief Bioinform 2004, 5: 150–163
13 Watterson SA, Wilson SM, Yates MD, Drobniewski FA. Comparison of three molecular assays for rapid detection of rifampin resistance in *Mycobacterium tuberculosis*. J Clin Microbiol 1998, 36: 1969–1973
14 Tu RJ, Deng CY, Xiong YZ. Sequencing analysis on partial translated region of uncoupling protein-3 gene and single nucleotide polymorphism associations with the porcine carcass and meat quality traits. Acta Genetica Sinica 2004, 31: 807–812
15 Tusnády GE, Simon I. Principles governing amino acid composition of integral membrane proteins: Applications to topology prediction. J Mol Biol 1998, 283: 489–506
16 Tusnády GE, Simon I. The HMMTOP transmembrane topology prediction server. Bioinformatics 2001, 17: 849–850
17 Yao X, Perez-Alvarado GC, Louis HA, Pomies P, Hatt C, Summers MF, Beckerle MC. Solution structure of the chicken cysteine-rich protein, CRP1, a double-LIM protein implicated in muscle differentiation. Biochemistry 1999, 38: 5701–5713
18 Urrutia AO, Hurst LD. Codon usage bias covaries with expression breadth and the rate of synonymous evolution in humans, but this is not evidence for selection. Genetics 2001, 159: 1191–1199
19 Goddard ME. Animal breeding in the (post-) genomic era. Animal Science 2003, 76: 353–365
20 Ng PC, Henikoff S. Predicting deleterious amino acid substitutions. Genome Res 2001, 11: 863–874
21 Zhang EP, Geng SM, Liu LL. Basic principles and statistic methods of QTL marker with SNPs. Animal Biotechnology Bulletin 2002, 8: 365–369

Edited by
Rong-Jia ZHOU