

Molecular characterization, polymorphism and association with reproductive traits of porcine CDK20 gene*

Y.G. Liu¹ and X.H. Xia^{2,3}

*¹Yunnan Agricultural University, College of Animal Science and Technology
Kunming 650201, P.R. China*

*²Kunming University of Science and Technology, Faculty of Life Science and Technology
Kunming 650224, P.R. China*

(Received 2 May 2011; revised version 29 July 2011; accepted 6 December 2011)

ABSTRACT

The complete coding sequence of one pig gene, CDK20, was amplified by RT-PCR. Sequence analysis revealed that the porcine CDK20 gene encodes a protein of 343 amino acids which has high homology with the cyclin-dependent kinase 20 (CDK20) of eight species: cattle (96%), giant panda (96%), mouse (96%), rabbit (96%), rat (96%), dog (95%), human (93%) and Sumatran orangutan (93%). This novel porcine gene was then deposited into the GenBank database and finally assigned to GeneID: 100157041. This gene is structured in eight exons and seven introns as revealed by computer-assisted analysis. Polymorphism analysis revealed that there was a T/C substitution at the position of 96-bp of exon 4 and this led to the amino acid alteration from the Arg residue to the Cys residue. PCR-Alu I-RFLP analysis indicated that eight pig breeds displayed obvious genotype and allele frequency differences at this mutation locus. Association analysis showed that this polymorphic locus was significantly associated with the litter size of all parities in Large White and Landrace sows. Therefore, CDK20 gene could be a useful candidate gene for increasing the litter size of pigs. These data serve as a foundation for further insight into the porcine CDK20 gene.

KEY WORDS: pig, CDK20 gene, SNP, litter size, reproduction

* Supported by Grants from the Natural Science Foundation Key Project of Yunnan Province (No. 2009CC015), National Nature Science Foundation of China (No. 30800810 and No. 30960346) and the Candidates of the Young and Middle Aged Academic and Technical Leaders of Yunnan Province (No. 2009CI055 and No. 2010CI006)

³ Corresponding author: e-mail: oliverxia2000@yahoo.com.cn

INTRODUCTION

Cyclin-dependent kinase 20 (CDK20) is a protein encoded by CDK20 gene containing a kinase domain most closely related to the cyclin-dependent protein kinases. The encoded kinase may activate cyclin-dependent kinase 2 and is involved in cell growth. However, latest studies have shown that CDK20 gene also functions in cell division, multicellular organismal development, protein phosphorylation and is also involved in the biological processes of transferase activity, nucleotide binding, cyclin-dependent protein kinase activity and ATP binding (Liu et al., 2004; Wohlbold et al., 2006; Qiu et al., 2008; Wu et al., 2009; Zimin et al., 2009; An et al., 2010).

As mentioned above, CDK20 gene is an important gene which has many biological functions. Until today, CDK20 gene has been reported in human, mouse, cattle and other animals, but the swine CDK20 gene has not been reported yet.

In the present experiment, we will clone the complete coding sequence of porcine CDK20 gene, and further do sequence analysis and polymorphic analysis. Furthermore, we will examine the swine CDK20 gene as a candidate gene for porcine reproductive traits through association analysis with the litter size.

MATERIAL AND METHODS

Animals and sample preparation

Six Large White pigs were slaughtered. Large intestine, spleen, lung, muscle, fat, liver, heart, kidney and ovary samples were collected, frozen in liquid nitrogen, and then stored at -80°C. The total RNA was extracted using the Total RNA Extraction Kit (Gibco, USA). These RNA samples were used to perform the RT-PCR for cloning the coding sequence.

Ear samples were collected from 795 unrelated animals belonging to eight pig populations presented in Table 1. Genomic DNA isolated from these ear samples would be used to perform the polymorphism analysis.

Table 1. The information on 795 unrelated pigs from eight populations

Breed	Sampling location	Sample size		
		total	male	female
Large White	Guangdong Province	100	0	100
Landrace	Guangdong Province	100	0	100
Saba	Dongchuan county of Yunnan Province	100	50	50
Tibetan	Xianggelila county of Yunnan Province	95	50	45
Mingguang small-ear	Tengchong county of Yunnan Province	100	50	50
Diannan small-ear	Banna state of Yunnan Province	100	50	50
Wujin pig	Qujing city of Yunnan Province	100	50	50
Baoshan	Baoshan city of Yunnan Province	100	50	50

Both the total number of piglets born (TNB) and the number of piglets born alive (NBA) of 100 Large White sows and 100 Landrace sows listed in Table 1 were recorded for 700 litters. The litter size traits data and genomic DNA of these pigs would be used to perform association analysis.

Isolation of the coding sequence for the porcine CDK20 gene

RT-PCR was performed to isolate the coding sequence for the porcine CDK20 gene using the cDNAs from different tissues above. The 25 μ l reaction system was: 2.0 μ l cDNA (100 ng/ μ l), 2.5 μ l 2 mM mixed dNTPs, 2.5 μ l 10 \times Taq DNA polymerase buffer, 2.5 μ l 25 mM MgCl₂, 2.0 μ l 10 μ M forward primer1, 2.0 μ l 10 μ M reverse primer1, 2.0 units of Taq DNA polymerase (1U/1 μ l), and 9.5 μ l sterile water. The primers for porcine CDK20 gene isolation were designed based on the conserved coding sequence information from human and mouse CDK20 genes and their highly homologous pig EST sequences: CV874425 and CJ031871. The PCR primers were: 5'- ATG GAG CAG TAC TGC ATC -3' (forward primer1), 5'-TCA GCC CTC CAG GAT GAA -3' (reverse primer1). The PCR programme initially started with a 94°C denaturation for 4 min, followed by 35 cycles of 94°C/50s, 55°C/50s, 72°C/50s, then 72°C extension for 10 min, finally 4°C to terminate the reaction.

The PCR products were then cloned into pMD18-T vector (TaKaRa, Dalian, China) and sequenced bidirectionally with the commercial fluorometric method (SHENGGONG, Shanghai, China). At least five independent clones were sequenced for each PCR product.

Sequence analysis

The gene analysis for cDNA sequence was conducted using GenScan software (<http://genes.mit.edu/GENSCAN.html>). The protein prediction and analysis were performed using the Conserved Domain Architecture Retrieval Tool of BLAST at the National Center for Biotechnology Information (NCBI) server (<http://www.ncbi.nlm.nih.gov/BLAST>) and the ClustalW software (<http://align.genome.jp/>). The theoretical isoelectric point (pI) and molecular weight (Mw) of proteins was computed using the Compute pI/Mw Tool http://www.expasy.org/tools/pi_tool.html.

PCR-RFLP

The DNA from the above pigs (Table 1) was used as a template to perform PCR with primers: 5'- AC ACC AGC TCC GGG AGC A -3' (forward primer2), 5'- CCT CGG GCC AGC TCA AGA -3' (reverse primer2). The PCR product size

is 204-bp. The 25 μ l reaction system was: 2.0 μ l DNA (100 ng), 2.5 μ l 2 mM mixed dNTPs, 2.5 μ l 10 \times Taq DNA polymerase buffer, 2.5 μ l 25 mM MgCl₂, 1.0 μ l 20 μ M forward primer, 1.0 μ l 20 μ M reverse primer, 1.0 units of Taq DNA polymerase (1 U/1 μ l) (JINMEI, BIOTECH), and 12.5 μ l sterile water. PCR was run as follows: 94°C for 4 min, followed by 35 cycles of 94°C for 50 s, 61°C for 50 s, 72°C for 1 min, then 72°C extension for 10 min, finally 4°C to terminate reaction. The 31 μ l PCR-RFLP reaction volume was: 10 μ l PCR product, 18 μ l sterile water, 1 μ l Alu I (10 U), 2 μ l 10 \times buffer. The mixture was incubated in an air incubator at 37°C for 4 h, and then the genotypes were analysed on the agarose gel (2.5%) containing ethidium bromide.

Statistical analysis

The relationships between CDK20 genotypes and litter size traits of Large White (n=100) and Landrace (n=100) sows were evaluated with the general linear model (GLM) procedure of SAS version 8.1 (SAS Institute Inc. Cary, NC, 2000). Both additive and dominant effects were also estimated using REG procedure, where the additive effect was estimated as -1, 0 and 1 for CC, CT and TT genotype, respectively; and the dominant effect represented as 1, -1 and 1 for CC, CT and TT genotype, respectively (Zhang et al., 2009). The model was:

$$Y_{ijkl} = \mu + P_i + S_j + F_k + G_l + e_{ijkl}$$

where: Y_{ijkl} - the observation of the trait, μ - the least square means, P_i - the effect of i^{th} parity ($i=1, 2, 3, 4, 5, 6, 7$ (parity ≥ 7)), S_j - the effect of j^{th} season, F_k - the effect of k^{th} farm ($k=1, 2$), G_l - the effect of l^{th} genotype ($l=1-3$), e_{ijkl} - the random residual (Niu et al., 2009).

RESULTS

Isolation of the coding sequence for the porcine CDK20 gene

For pig CDK20 gene, through RT-PCR, one PCR product of 1041-bp was obtained (Figure 1).

Sequence analysis

The cDNA nucleotide sequence analysis using the BLAST software revealed that this gene was not homologous to any of the known pig genes and it was then deposited into the GenBank database (Accession number: GU373708). The

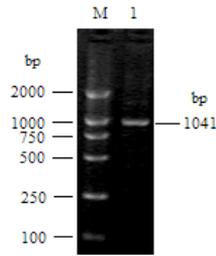


Figure 1. RT-PCR result for pig CDK20 gene. M - DL2000 DNA markers; 1 - PCR product for pig CDK20 gene

sequence prediction was carried out using the GenScan software and result showed that the 1041-bp cDNA sequence represented one single gene which encoded 346 amino acids. The theoretical isoelectric point (pI) and molecular weight (Mw) of these deduced proteins of this pig gene were computed using the Compute pI/Mw Tool. The pI of pig CDK20 is 6.05. The molecular weight of this putative protein is 38449.45. This novel porcine gene was finally assigned to GeneID: 100157041.

BLAST analysis revealed that the pig CDK20 protein has high homology with the cyclin-dependent kinase 20 (CDK20) of eight species: cattle (accession number: NP_001092403; 96%), giant panda (accession number: XP_002919653; 96%), mouse (accession number: NP_444410; 96%), rabbit (accession number: XP_002708330; 96%), rat (accession number: NP_001020923; 96%), dog (accession number: XP_533555; 95%), human (accession number: NP_001034892; 93%) and Sumatran orangutan (accession number: NP_001126328; 93%).

To obtain the genomic DNA of CDK20, the publicly available pig genome database at the NCBI Pig Genome Resources (<http://www.ncbi.nlm.nih.gov/projects/genome/guide/pig/>) was screened using the full-length cDNA sequence of CDK20 as a seed. A bacterial artificial chromosome (BAC) clone (*Sus scrofa* chromosome 14 clone CH242-21112, GenBank accession no. CU571022) which encompasses entire CDK20 gene was identified by BLASTGen analysis. The pig CDK20 gene (nucleotides 8,531-15,223 in the *Sus scrofa* chromosome 14 clone CH242-21112) is 7,693 bp in length and consists of eight exons. All exon–intron splice junction sequences conform to the GT-AG rule (Figure 2).



Figure 2. The genomic sequence organization representing the ORF of the pig CDK20 gene

Polymorphism

Based on the sequence of pig CDK20 gene, primers (forward primer 2 and reverse primer 2) were used to amplify the DNA of Large White and Landrace. The products were then cloned into PMD18-T vector and sequenced bidirectionally with the commercial fluorometric method. One T/C mutation was found at the position of 96-bp of the exon 4. This led to mutation of one Alu I restriction site and the amino acid alteration from the Arg residue to the Cys residue. This was confirmed by PCR-Alu I-RFLP (Figure 3).

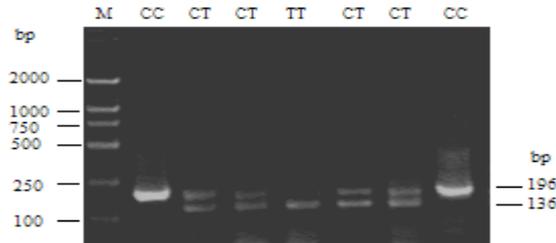


Figure 3. The polymorphism analysis of pig CDK20 gene by PCR-Alu I-RFLP. CC: 196 bp; CT: 196 bp+136 bp+60 bp; TT: 136 bp+60 bp

Subsequently, PCR-Alu I-RFLP were performed using the DNA from 795 unrelated animals belonging to eight pig populations including Large White, Landrace, Saba, Tibetan, Mingguang small-ear, Diannan small-ear, Wujin and Baoshan pigs. Results revealed that the frequency of T allele in the two European pig breeds: Large White (0.425) and Landrace (0.450) is higher than that in other six Yunnan local pig breeds: Saba (0.340), Tibetan (0.320), Mingguang small-ear (0.275), Diannan small-ear (0.330), Wujin (0.405) and Baoshan pigs (0.370). The two European pig breeds: Large White and Landrace have more animals of genotype CT and TT. This indicated that Yunnan local pig breeds and European pig breeds displayed obvious genotype and allele frequency differences at this C/T mutation locus.

Table 2. Allele frequency and genotype of Alu I polymorphic locus in different pig breeds

Breed	Number of pigs	Genotype			Allele frequency	
		CC	CT	TT	C	T
Large White	100	26	63	11	0.575	0.425
Landrace	100	23	64	13	0.550	0.450
Saba	100	39	54	7	0.660	0.340
Tibetan pig	95	45	46	4	0.680	0.320
Mingguang small-ear	100	49	47	4	0.725	0.275
Diannan small-ear	100	39	56	5	0.670	0.330
Wujin	100	26	67	7	0.595	0.405
Baoshan	100	33	60	7	0.630	0.370

Table 3. Association between pig CDK20 gene PCR-Alu I -RFLP genotypes and litter size traits

Breed	Traits	Genotype, mean \pm SE			Genetic effects, mean \pm SE		
		CC	CT	TT	additive	dominant	
Large White	1 st parity	N	26	63	11		
		TNB	10.629 \pm 0.337	11.342 \pm 0.280	11.047 \pm 0.379	0.209 \pm 0.181	0.504 \pm 0.188
	All parities	NBA	9.064 \pm 0.316	9.057 \pm 0.279	9.119 \pm 0.367	0.027 \pm 0.176	-0.034 \pm 0.209
		N	26	63	11		
Landrace	1 st parity	TNB	11.206 \pm 0.288 ^A	11.571 \pm 0.295	12.243 \pm 0.329 ^B	0.518 \pm 0.176	-0.149 \pm 0.216
		NBA	9.714 \pm 0.367 ^A	9.957 \pm 0.315	10.787 \pm 0.415 ^B	0.536 \pm 0.209	-0.293 \pm 0.181
	All parities	N	23	64	13		
		TNB	11.090 \pm 0.345	11.426 \pm 0.348	11.725 \pm 0.448	0.317 \pm 0.215	0.018 \pm 0.177
Landrace	1 st parity	NBA	9.600 \pm 0.38514	9.764 \pm 0.264	10.050 \pm 0.363	0.225 \pm 0.198	-0.061 \pm 0.214
		N	23	64	13		
	All parities	TNB	11.152 \pm 0.315 ^A	11.681 \pm 0.244	12.164 \pm 0.312 ^B	0.506 \pm 0.244	0.023 \pm 0.181
		NBA	10.277 \pm 0.214 ^a	10.560 \pm 0.319	11.052 \pm 0.369 ^b	0.387 \pm 0.274	-0.104 \pm 0.221

N - number of investigated litters; TNB - total number of piglets born; NBA - the number of piglets born alive; least square mean values with different letters are significantly different: small letter - P<0.05; capital letter - P<0.01

Association of the SNP and litter size was assessed in two populations (purebred Large White and purebred Landrace sows). Statistical analysis demonstrated that for the first parity, no significant difference was found between animals of the three genotypes in the experimental purebred Large White sows and in the experimental purebred Landrace sows. For all parities, in the purebred Large White sows, those with the TT genotype had an additional 1.037 piglets born compared to the CC animals ($P < 0.01$) and an additional 1.073 piglets born alive compared to the CC animals ($P < 0.01$). In addition, for all parities, in the purebred Landrace sows, TT animals had 1.012 more piglets born ($P < 0.01$) and 0.775 more piglets born alive ($P < 0.05$) than the CC animals.

DISCUSSION

In the current study, we first get the complete coding sequence of the porcine CDK20 gene. CDK20 gene had been reported to be an important reproduction related gene for it has been identified to be associated with the cell division, multicellular organismal development and protein phosphorylation (Liu et al., 2004; Wohlbold et al., 2006; Qiu et al., 2008; Wu et al., 2009; Zimin et al., 2009; An et al., 2010). Thus, this provides a molecular basis for association analysis of this gene DNA polymorphism with litter size traits of pig.

With the development of modern bioinformatics and establishment of specific pig NCBI EST database along with different convenient analysis tools, it is much easier for researchers to find the useful ESTs which are highly homologous to the coding sequence of human genes. Based on these pig EST sequences, we can obtain the complete coding sequences of some pig genes through some modern experimental methods such as RT-PCR. From the clone and sequence analysis of pig CDK20 gene, it could be seen that this is an effective method to isolate some novel pig genes.

The pig industry is now actively using this information to improve the pig production by marker assisted selection (MAS). So it is of utmost importance to find more candidate genes or DNA molecular markers in order to improve pig production. From association analysis results, it could be found that the polymorphism of the Alu I locus of porcine CDK20 gene can significantly affect litter size. The TT genotype animals obviously have better litter size of all parities than the CC animals in purebred Large White and Landrace sows. The pig industry can select and retain more TT animals to improve the reproductive performance of sows in pig production. We also noticed that only small Large White and Landrace populations had been used to perform the association analysis, and this, whether it will affect the association analysis results or not, should be validated in the future study using a larger size of samples.

CONCLUSIONS

In conclusion, we first isolated the pig CDK20 gene and performed necessary sequence analysis, polymorphic analysis and association analysis. This established the primary foundation for further insight into this novel pig gene.

REFERENCES

- An X., Ng S.S., Xie D. et al., 2010. Functional characterisation of cell cycle-related kinase (CCRK) in colorectal cancer carcinogenesis. *Eur. J. Cancer* 46, 1752-1761
- Liu Y., Wu C., Galaktionov K., 2004. p42, a novel cyclin-dependent kinase-activating kinase in mammalian cells. *J. Biol. Chem.* 279, 4507-4514
- Niu B.Y., Ye L.Z., Li F.E., Deng C.Y., Jiang S.W., Lei M.G., Xiong Y.Z., 2009. Identification of polymorphism and association analysis with reproductive traits in the porcine RNF4 gene. *Anim. Reprod. Sci.* 110, 283-292
- Qiu H., Dai H., Jain K., Shah R., Hong C., Pain J., Tian B., Vatner D.E., Vatner S.F., DePre C., 2008. Characterization of a novel cardiac isoform of the cell cycle-related kinase that is regulated during heart failure. *J. Biol. Chem.* 283, 22157-22165
- Wohlbold L., Laroche S., Liao J.C., Livshits G., Singer J., Shokat K.M., Fisher R.P., 2006. The cyclin-dependent kinase (CDK) family member PNQALRE/CCRK supports cell proliferation but has no intrinsic CDK-activating kinase (CAK) activity. *Cell Cycle* 5, 546-554
- Wu G.Q., Xie D., Yang G.F. et al., 2009. Cell cycle-related kinase supports ovarian carcinoma cell proliferation via regulation of cyclin D1 and is a predictor of outcome in patients with ovarian carcinoma. *Int. J. Cancer* 125, 2631-2642
- Zhang Y.H., Mei S.Q., Peng X.W. et al., 2009. Molecular characterization and SNPs analysis of the porcine Deleted in Azoospermia Like (pDAZL) gene. *Anim. Reprod. Sci.* 112, 415-422
- Zimin A.V., Delcher A.L., Florea L. et al., 2009. A whole-genome assembly of the domestic cow, *Bos taurus*. *Genome Biol.* 10, R42 (Abstr.)