Journal of Animal and Veterinary Advances 11 (8): 1230-1236, 2012
ISSN: 1680-5593

Identification, mRNA Expression and Characterization of a Novel *CENPP* Gene from Chinese Banna Mini-Pig Inbred Line (BMI)

[1]Liu Lixian, [2, 3]Wang Pei, [2, 3]Huo Jinlong, [1, 4]Huo Hailong, [1]Zhao Yue,
[1]Wang Rui and [2, 3]Zeng Yangzhi
[1]Department of Husbandry and Veterinary,
Yunnan Vocational and Technical College of Agriculture,
650212 Kunming, Yunnan, P.R. China
[2]Key Laboratory of Banna Mini-Pig Inbred Line of Yunnan Province,
[3]Faculty of Animal Science and Technology, Yunnan Agricultural University,
650201 Kunming, Yunnan, P.R. China
[4]Faculty of Life Science, Yunnan University, 650091 Kunming, Yunnan, P.R. China

Abstract: The complete CDS sequence of Banna Mini-pig Inbred line (BMI) gene *CENPP* was amplified using the Reverse Transcriptase Polymerase Chain Reaction (RT-PCR) based on the conserved sequence information of the cattle or other mammals and known highly homologous swine ESTs. This novel gene was then deposited into NCBI database and assigned to Accession No.: JF944894. Sequence analysis revealed that the BMI CENPP encodes a protein of 291 amino acids that has high homology with the Centromere Protein P (CENPP) of seven other species-horse (85%), cattle (83%), human (77%), orangutan (76%), monkey (76%), mouse (70%) and rat (68%). The phylogenetic tree analysis revealed BMI CENPP has a closer genetic relationship with the horse CENPP than with those of cattle, human, orangutan, monkey, mouse and rat. Analysis by RT-PCR showed that BMI *CENPP* gene was over-expressed in midbrain, ovary, spleen, lung, skin, muscle, moderately expressed in diencephalon, heart, nerve fiber, stomach, small intestine, large intestine and almost not expressed in other 6 tissues. Two microRNA target sites were predicted in the CDS of BMI CENPP mRNA for further studying this gene in the future. The experiment will establish a foundation for further insight into this swine gene.

Key words: Banna Mini-pig Inbred line (BMI), pig, CENPP, tissue expression, bioinformatics analysis, China

INTRODUCTION

CENPP gene encodes the centromere protein P which is a subunit of a CENPH-CENPI associated centromeric complex that targets CENPA to centromeres and is required for proper kinetochore function and mitotic progression (Okada *et al*., 2006). The centromere of higher eukaryotes was first defined as the primary constriction on mitotic chromosomes (Rattner, 1987; Ris and Witt, 1981) which is essential for faithful chromosome segregation during mitosis and meiosis (Pluta *et al*., 1995). The histone H3 variant CENP-A is a component of centromeric chromatin and defines active centromere regions by forming centromere-specific nucleosomes (Ando *et al*., 2002). Overexpression in this gene may indirectly be associated with aneuploidy which is the hallmark of many human cancers (Tomonaga *et al*., 2003). In particular as a member of CENP-H-I complex, CENPP plays an role in chromosome congression, maintenance of stale chromosome alignment and timely progression from metaphase to anaphase (Okada *et al*., 2006).

Swine are generally considered to be the most ideal biomedical laboratory animals, for their anatomical, physiological and metabolic characteristics are similar to human's. The inbred animals are good enough to be used as experimental animals, owing to their clear genetic background, high homozygosity, stable inheritance and so on. Inbred animals can also make less experimental errors using in biological research than noninbred ones (Wright, 1921; Harris, 1997). In 1980, the Banna Mini-pig Inbred line (BMI) was exploited by Yunnan Agricultural University based on the small-ear pigs at Xishuangbanna, Yunnan province, China. A pair of progenitors was a sow and her son with some degree inbreeding background. Then, the propagation was conducted by means of highly

Corresponding Author: Huo Jinlong, Key Laboratory of Banna Mini-Pig Inbred Line of Yunnan Province,
650201 Kunming, Yunnan, P.R. China

full sibling or parent-offspring inbreeding and each generation underwent the strict selection. As heterozygotic genes were separated and recombined in the process of inbreeding, BMI has already owned six families and eighteen substrains with different phenotypes and genotypes. Due to their consistent genetic background and minor interindividual differences, BMI is considered as an ideal model organism for biological studies (Yu *et al*., 2004; Zeng and Zeng, 2005).

Based on the description of about the *CENPP* gene, it is necessary to isolate this gene from pig for it is associated with energy metabolism, health and other important biological functions of animals. But until today the porcine CENPP has not been reported yet. The objective of this study was to isolate the full length coding sequence of BMI *CENPP* gene according to the conserved sequence information of cattle or other mammals and highly homologous swine ESTs sequence information, conduct sequence analysis and some necessary function analysis of established nucleotide sequence, finally examine the expression in a range of BMI tissues. These will provide a primary foundation for further research on this porcine gene.

MATERIALS AND METHODS

Samples collection, RNA extraction and first-strand cDNA synthesis: Three matured female BMI were slaughtered for sampling. Fresh tissues (lymph node, midbrain, ovary, diencephalon, cerebrum, liver, kidney, spleen, heart, lung, nerve fiber, stomach, small intestine, large intestine, pancreas, skin, muscle and fat) were instantly frozen in liquid nitrogen and stored at -80°C before use. Total RNA was extracted using the RNAiso Plus (TaKaRa, Dalian) according to the manufacturer's instructions. To remove genomic DNA contamination, total RNA was digested with RNase-free DNase I (TaKaRa, Dalian). The 3 μg of RNA were reverse transcribed with oligo $(dT)_{18}$ primer and M-MLV reverse transcriptase (Invitrogen, USA).

Isolation of the BMI *CENPP* gene: The GenBank CENPP sequences for cattle (Accession No.: NM_001105615), human (Accession No.: NM_001012267) and their highly homologous pig ESTs sequences: EW133717, CX063676, CN157364, CX063924, CV870136, BX665631, CN159311 and BX666162 were used to design a primer pair to amplify the complete coding sequence of CENPP by using Primer Premier 5.0 Software. The primers for BMI *CENPP* gene were: 5'- CAC ACA CCT CTA GGT ACA GC -3' and 5'- AAT TCC GAT AAC GAA CGA GAC T -3'. RT-PCR was performed to isolate the BMI CENPP using the pooled cDNAs from different tissues above. The 25 μL reaction system was: 2.0 μL cDNA (25 ng μL^{-1}), 2.0 μL^{-1} 2.5 mM mixed dNTPs, 2.5 μL 10×Taq DNA polymerase buffer (Mg^{2+} plus), 0.5 μL 10 μM forward primer, 0.5 μL 10 μM reverse primer, 0.25 μL Taq DNA polymerase (5 U μL^{-1}, TaKaRa, Dalian) and 17.25 μL sterile water. The PCR program initially started with 94°C denaturation for 2 min followed by 35 cycles of 94°C/30 sec, 55°C/40 sec, 72°C/1.5 min then 72°C extension for 10 min, finally 4°C to terminate the reaction. After the PCR the gene product was cloned into pMD18-T vector (TaKaRa, Dalian) and sequenced bidirectionally with the commercial fluorometric method. At least five independent clones were sequenced.

Bioinformatics analysis: Sequence analysis of BMI *CENPP* gene was performed using software in NCBI (http://www.ncbi.nlm.nih.gov) and ExPaSy (http://www.expasy.org). The cDNA sequence were predicted using the GenScan Software. The Blastp program and Conserved Domain Architecture Retrieval Tool were used to search for similar proteins and conserved domain, respectively (http://www.ncbi.nlm.nih.gov/Blast). The alignment of the nucleotide sequences and deduced amino acid sequences were computed using ClusterX and the phylogenetic trees were computed using the Mega 4.0 Software with standard parameters. The theoretical Isoelectric point (pI) and Molecular weight (Mw) were predicted using the Compute pI/Mw Tool (http://us.expasy.org/tools/pi_tool.html). The putative signal peptide was predicted using the SignalP 4.0 server. The putative protein subcellular localization was predicted using PSort II (http://psort.hgc.jp/). Transmembrane topology prediction was performed using TMHMM-2.0 server. Secon-dary structures of deduced amino acid sequences were predicted with SOPMA. Web-based microRNA (miRNA) predicting programs were used to locate conserved potential miRNA targets (http://www.mirbase.org/). The 3D structures was predicted based on the existed 3D structures by the amino acids homology modeling on Swiss server (http://swissmodel.expasy.org/).

Semi-quantitative RT-PCR: To characterize the *CENPP* gene further, RT-PCR was conducted to determine its expression in 18 BMI tissues. To eliminate the effect of cDNA concentration, researchers selected the housekeeping gene 18S rRNA (NR_002170) as a positive control. The control primers used were: 5'- GGA CAT CTA AGG GCA TCA CAG -3' and 5'- AAT TCC GAT AAC GAA CGA GAC *T* -3'. The BMI CENPP primers which were used to perform the semi-quantitative RT-PCR for tissue expression profile analysis were the same as the

primers for isolation RT-PCR before. The PCR reactions were optimized for a number of cycles to ensure product intensity within the linear phase of amplification.

RESULTS AND DISCUSSION

Cloning and identification of BMI *CENPP* cDNA: An CENPP fragment 920 bp of including the entire 876 bp coding region plus 3 bp of 5'-UTR and 41 bp of 3'-UTR was amplified (Fig. 1). This cDNA nucleotide sequence analysis using the BLAST software at NCBI server revealed that BMI *CENPP* gene was not homologous to any of the known porcine genes and it was then deposited into the GenBank database (Accession No.: JF944894). The sequence prediction was carried out using the GenScan Software and results showed that the 876 bp cDNA sequence represent a single gene which encoded 291 amino acids. The complete CDS and the encoded amino acids were shown in Fig. 2.

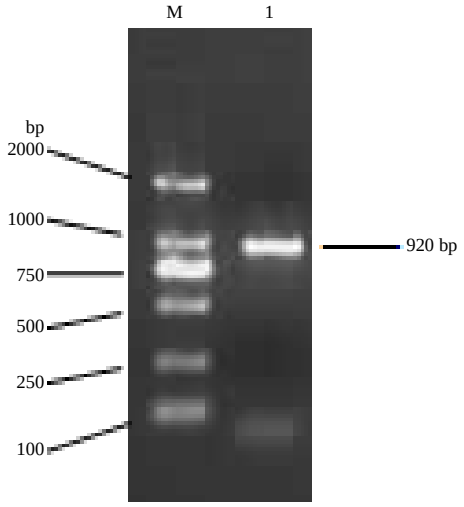

Fig. 1: RT-PCR result for BMI *CENPP* gene; M: DL2000 DNA marker, 1: PCR product

Physical and chemical characteristics of BMI CENPP: The theoretical Isoelectric point (pI) and Molecular weight (Mw) of CENPP were computed using the compute pI/Mw tool. The theoretical pI and the molecular weight of BMI CENPP are 5.43 and 33579.29, respectively. The result from SignalP revealed that there is has no putative signal peptide in BMI CENPP and indicated that it was probably a non-secretory protein (Petersen *et al*., 2011). The potential protein subcellular localization prediction by Reinhardt's method showed that BMI CENPP was probably located in the nuclear with up to 52.2% probability (Nakai and Horton, 1999). Using a hidden Markov Model Algorithm, transmembrane topology prediction made by TMHMM program (Moller *et al*., 2001) showed that BMI CENPP was not a potential transmembrane protein.

Prediction and analysis of structures and conserved domains of BMI CENPP: Proteins often contained several domains, each of which had their own evolutionary origins and functions. Examined using the Conserved Domain Architecture Retrieval Tool of Blast at the NCBI server (http://www.ncbi.nlm.nih.gov/BLAST) indicated that BMI CENPP had not putative conserved domains The putative protein was also analyzed using prosite (http://expasy.org/prosite/).

Six kinds sites were found which were protein kinase C phosphorylation sites (3-SlR-5, 52-SlK-54, 77-SiR-79, 82-SkK-84, 94-TeK-96, 183-TfK-185, 261-SfR-263), Leucine zipper pattern (4-25), Casein kinase II phosphorylation sites (38-SfqD-41, 45-SdsE-48,52-SlkD-55, 89-TstE-92, 155-SrtE-158, 212-SrpE-215, 245-SalE-248), N-glycosylation sites (80-NYSK-83, 130-NLSS-133), cAMP- and cGMP-dependent protein kinase phosphorylation site (180-RKrT-183) and N- myristoylation sites (200-GAssSY-205, 270-GIeaTL-275). The prediction of secondary structure

```
accATGGACTCCCTTCGCGCTGAGGCCGACGCCTTGGAAGCCGAGATCAAGGCCTTGAGGCGGGCGTGCGTGGAGCTACCGGCGCCGGGTGAAGACACCT 100
   M  D  S  L  R  A  E  A  D  A  L  E  A  E  I  K  A  L  R  R  A  C  V  E  L  P  A  P  G  E  D  T   32
CACCAGTCCGACAATCATTTCAAGATATATACCAATCAGATTCTGAAGGATGGGAGTCATTGAAAGATCTGAGAAGTCATCTTGGACATGTAGAATCAGA 200
S  P  V  R  Q  S  F  Q  D  I  Y  Q  S  D  S  E  G  W  E  S  L  K  D  L  R  S  H  L  G  H  V  E  S  E   66
ACTTCAATTTCTAAGTTCGCTTACTGGCATCAGCATAAGAAATTATTCCAAGAAGACGGAGGACCTAACAAGCACTGAAATGACAGAAAAGAATATAAAG 300
  L  Q  F  L  S  S  L  T  G  I  S  I  R  N  Y  S  K  K  T  E  D  L  T  S  T  E  M  T  E  K  N  I  K   99
AAAGTTCTACAGAGACACAGATTATCAGGAAGTTGCCACATGATTACATTTCAACTTGAATTTCAGATTCTGGAAATCCAGAATAAGGAGAATTTATCTT 400
 K  V  L  Q  R  H  R  L  S  G  S  C  H  M  I  T  F  Q  L  E  F  Q  I  L  E  I  Q  N  K  E  N  L  S  132
CTGTTATTACTGACCTCAGCATAATAATGGAGCCCACCGAATATTCAGAATTAAGTGAATTTGTATCTAGAACAGAAGAAAGAAGAGATCTGTTTATGTT 500
S  V  I  T  D  L  S  I  I  M  E  P  T  E  Y  S  E  L  S  E  F  V  S  R  T  E  E  R  R  D  L  F  M  F 166
TTTCCGAAGCCTACACTTCTTTGTGGAGTGGTGTGAATATCGCAAGCGCACATTTAAACATTTCAAGGAAAAAGTACCCAGAGACTGTGCACCTTGCCAAG 600
  F  R  S  L  H  F  F  V  E  W  C  E  Y  R  K  R  T  F  K  H  F  K  E  K  Y  P  E  T  V  H  L  A  K  199
GGAGCCTCCTCCAGCTACATGGGGATCAGGAGCCACAGCCGACCAGAGTTTGAACTGGTAATTGTTTGGAGGATACAAATAGATGAAGAAGGGAAGGTTT 700
 G  A  S  S  S  Y  M  G  I  R  S  H  S  R  P  E  F  E  L  V  I  V  W  R  I  Q  I  D  E  E  G  K  V  232
TGCCGAGGCTGGATCTTCTCACCAAGGCGCCGCTGTCAGCCCTGGAACTGGACAAGAACAGAGTCATAGAAACTGCTCCTCTCAGCTTCCGAACCCTGCT 800
L  P  R  L  D  L  L  T  K  A  P  L  S  A  L  E  L  D  K  N  R  V  I  E  T  A  P  L  S  F  R  T  L  L 266
GGGTGTGCTGGGCATTGAAGCCACTCTAGAAAGCCTGATAAAATCGCTTTGCACAGAGGAACAAGTTCCCACACAGTGAacaactacacaggagtaagat 900
  G  V  L  G  I  E  A  T  L  E  S  L  I  K  S  L  C  T  E  E  Q  V  P  T  Q                         291
gctgtacctagaggtgtgtg                                                                               920
```

Fig. 2: The complete cDNA sequence and amino acid sequence of the protein encoded by CENPP (GenBank Accession No.: JF944894). ATG: Start codon; TGA: Stop codon and gray highlighted nucleotide sequence, primers

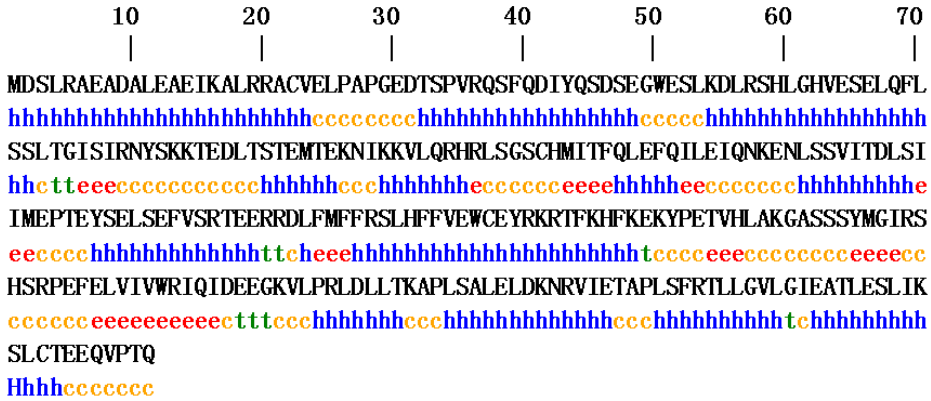

Fig. 3: The secondary structure of the BMI CENPP protein predicted by SOPMA. Hh: Alpha helix; Ee: Extended strand; Tt: Beta turn; Cc: Random coil

by SOPMA indicates that the deduced BMI CENPP consists of alpha helices (165AA), extended strands (33AA), beta turn (9AA) and random coils (84AA) (Fig. 3) using the amino acids homology modeling on swiss server (http://swissmodel.expasy.org/) indicated that BMI CENPP could not identify useful template structures.

Location of potential miRNA targets: MicroRNAs are noncoding single-stranded RNA molecules of 17-24 nucleotides that can regulate gene expression by binding to the coding region of target mRNAs (Bartel, 2004; Zeng *et al*., 2003). Researchers use web-based microRNA (miRNA) predicting programs to locate conserved potential miRNA targets: miRBase (http://www.mirbase.org/). The results showed two *Sus scrofa* microRNAs (ssc-miR-202, ssc-miR-708-3p) were found to have the predicted target sites (499-uuccgaagccuacacuucuuu-519 and 349-caacuugaa-uuucagauucugga-371) in the BMI CENPP CDS sequence, respectively. Further investigation is needed to confirm whether corresponding miRNA molecules can regulate the *CENPP* gene expression in swine.

Analysis of sequence identity and evolutionary relationships of BMI CENPP: The deduced protein sequence of BMI CENPP was submitted to generate BLAST reciprocal best hits and similarity comparison revealed that BMI CENPP protein has high homology with the CENPP proteins of seven other species-horse (85%), cattle (83%), human (77%), orangutan (76%), monkey (76%), mouse (70%) and rat (68%) (Fig. 4). To evaluate the evolutionary relationships of BMI CENPP with other species then we constructed a phylogenetic tree using Mega 4.0 Software on the basis of the CENPP amino acid sequences. The phylogenetic tree analysis revealed that the BMI *CENPP* gene has a closer genetic relationship with the horse, bovine and human CENPP (Fig. 5). Therefore, researchers can use horse, bovine and human as model organisms to study the pig *CENPP* gene or use pig as a model organism to study the horse, bovine and human *CENPP* genes.

mRNA tissue-specific expression profile: To check the relative expression levels of CENPP mRNA in various porcine tissues, semi-quantitative RT-PCR was performed in 18 BMI tissues. The continuously expressed gene, 18S was used and served as an endogenous reference for determination of targeted mRNA profiles. semi-quantification analysis was performed with Quantity One Gray Scanning Software.

Result revealed that BMI *CENPP* gene was over-expressed in midbrain, ovary, spleen, lung, skin and muscle moderately expressed in diencephalon, heart, nerve fiber, stomach, small intestine and large intestine almost not expressed in other 6 tissues (Fig. 6).

Comparative genomics determines the relationship of genome structure and function of different species. Researchers have learned a great deal about the function of human genes by examining their counterparts in simpler model organisms such as the mouse and some results have revealed that virtually all (99%) of the protein-coding genes in humans align with homologues in mice and over 80% are clear 1:1 orthologs (Hardison, 2003; Liu *et al*., 2008). This extensive conservation in protein-coding regions implied that the same protein-coding sequences may be expected in different mammals including pig. From the isolation of swine *CENPP* gene, researchers can find

```
BMI-pig     MDSLRAEADALEAEIKALRRACVELPAPGEDTSPVRQSFQDIYQSDSEGWESLKDLRSHL
Horse       ----MAQVNSVEYSRS--------------------KSFQEIYQSDSEGWESSKDLRSHL
Cattle      MDSQRAKARALEAEIAALRRACIETQAPGDDDSQVRKSFQETHQSNSEEWGLSKDLRSHL
Human       MDAELAEVRALQAEIAALRRACEDPPAPWEEKSRVQKSFQAIHQFNLEGWKSSKDLKNQL
Orangutan   MDAELAEVRALQAEIAALRRACEEPPAPWEEKSRVQKSFQAIHQFNLEGWKSSKDLKNQL
Monkey      MDAELAEERVLQAEIAAVRRACENPPAPWEEKSGVQKSFQAIHQFTSEGWKSSKDLKNQL
Mouse       MDSETRELRALEAEVAALQRECRMLQNAGEKASGAWKSFQKISQSDSEEWESLKDLRSQL
Rat         MDPETRELRALEEEVAALQRECRMLQNLEEKTSGAWKSFKKFSQSDSEEWESLKDLRSQL
                  :   :: .                        :**:   *   * *   ***:.:*

BMI-pig     GHVESELQFLSSLTGISIRNYSKKTEDLTSTEMTEKNIKKVLQRHRLSGSCHMITFQLEF
Horse       GHLESEILFLSTLTGIKIRNYSKKTKDLTSTEMTEKSIKKVLQRHRLSGNCHMITFQLEF
Cattle      GSLESELQFLSRLTGISIRNYCKNTEDLTNTEMAEKSIKKVLQRHRLSGNCHMITFQLEF
Human       GHLESELSFLSTLTGINIRNHSKQTEDLTSTEMTEKSIRKVLQRHRLSGNCHMVTFQLEF
Orangutan   GHLESELSFLSTLTGINIRNHSKQTEDLTSTEMTEKSIRKVLQRHRLSGNCHMVTFQLEF
Monkey      GHLESELSFLSTLTGINIRNYSRQTEDLTSTEMAEKSIRKVLQRHRLSGNCHMVTFQLEF
Mouse       EHLRSEHSFLSKLTGFNIRNYS-KMEDIVNTEETEKDTKKVLQKHRLSGNCHMVTFQLEF
Rat         GHLKSELSFLSKLTGFNIRNYS-KMEDIGNTEQTENDTKKVLQRHRLSGNCHMVTFQLEF
              :.**  *** ***:.***:. : :*: .** :*:. :****:*****.*:*:******

BMI-pig     QILEIQNKENLSSVITDLSIIMEPTEYSELSEFVSRTEERRDLFMFFRSLHFFVEWCEYR
Horse       QILEIQNKKSLSSVITDLNIIMEPTEYSELSEFVSRAEERRDLFMFFRSLRFFVEWCEYR
Cattle      QILEIQDKESLSSVITDLSIIMEPTEYSELSEFVSRAEEKRDLLMFFRSLRFFVEWCDYR
Human       QILEIQNKERLSSAVTDLNIIMEPTECSELSEFVSRAEERKDLFMFFRSLHFFVEWFEYR
Orangutan   QILEIQNKERLSSAVTDLNIIMEPTECSELSEFVSRAEERKDLFMFFRSLHFFVEWFEYR
Monkey      QILEIQNKERLSSAVTDLNIIMEPTECSELSEFVSRAEDRKDLFMFFRSLHFFVEWFEYR
Mouse       EVLEMETKEKKSSIITDLSIIMEPTEYSELSEFASRAEEKRDLLMFFRSLHFFVEWCEYR
Rat         EVLEIETKEKKSSIITDLSIIMEPTEYSELSEFASRAEEKRDLLMFFRSLHFFVEWYEYR
            ::**:: *:  ** :***.******* ******.**:*:::**:******:***** :**

BMI-pig     KRTFKHFKEKYPETVHLAKGASSSYMGIRSHSRPEFELVIVWRIQIDEEGKVLPRLDLLT
Horse       KRTFKHFKEKYPEAVHLAEGAASSCMGVRSASQPGFEVVIVWRIQIDDEGKVLPKLDLLT
Cattle      KRTFKHFKEKYPEAVQLSEGASSSCMGIRNPSQPGFELVIVWRIQIDEEGKVLPKLDLLT
Human       KRTFKHLKEKYPDAVYLSEGPSSCSMGIRSASRPGFELVIVWRIQIDEDGKVFPKLDLLT
Orangutan   KSTFKHLKEKYPDAVYLSEGPSSCSMGIRSASQPGFELVIVWRIQIDEDGKVFPKLDLLT
Monkey      KRTFKHFKEKYPDAVYLSEGPSSCSMGIRSASRPGFELVIVWRIQIDEDGKVFPKLDLLT
Mouse       ENTFKHFKEKYPDTVYLLEGTCSHCMEIRSTRQPGFELVIVWKIQIDEEGMVFPKLDLLT
Rat         ENTFKHFKERYPDSVYLSEGTCSHCMEIRNARLPGFELVIVWRIQIDEEGKVFPKLDLLT
            : ****:**:***:: * :*..*  * :*.   * **:*****:******:: *:*:*****

BMI-pig     KAPLSALELDKNRVIETAPLSFRTLLGVLGIEATLESLIKSLCTEEQVPTQ-
Horse       QVPQRVLELDKNRVIETAPLGFRNLLGVLGIEATLESLIKLLCTEEHARSRH
Cattle      KVPLRALELDKNRVIETAPLSFRTLLGVLGIEATLESLIKSFCTEESS----
Human       KVPQRALELDKNRAIETAPLSFRTLVGLLGIEAALESLIKSLCAEENN----
Orangutan   KVPQRALELDKNRAIETAPLSFRTLVGLLGIEAALESLIKSLCAEENN----
Monkey      KVPQRALELDKNRAIEAAPLSFRTLLGVLGIEAALESLIKSLCAEENN----
Mouse       KVPERALGLDKNRVIETAPLSFRSLLGVLGIEAALDSLIRLFSGDNN-----
Rat         KVPEQALEMDKNRVIETAPLSFRSLLGVLGIEAALDSLIRLFNTDK------
            :.*  .* :*****.**:***.**.*:*:*****:*:****: :  ::
```

Fig. 4: Alignment of the protein encoded by the BMI CENPP and seven other types of CENPP from horse (XP_001491590), cattle (NP_001099085), human (NP_001012267), orangutan (XP_001146410), monkey (NP_001181730), mouse (NP_079771) and rat (XP_002725285)

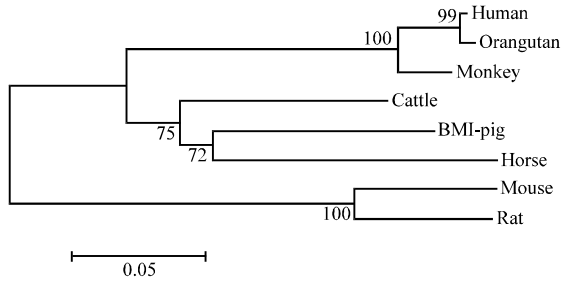

Fig. 5: Phylogenetic tree for CENPP proteins from 8 species

that swine CENPP is highly homologous with CENPP of horse, bovine, human and other mammals. This further validated that comparative genomics method is one useful tool to isolate the unknown genes especially the conserved coding region of genes for pig. From the alignment analyses for swine CENPP protein we also find BMI CENPP protein was not completely identity to horse or other mammals. This implied that BMI CENPP will have some differences in functions to those of horse, cattle, human and other mammals. In this study, we not only cloned the CDS sequences of the BMI *CENPP* gene but also conducted the sequence analysis and tissue

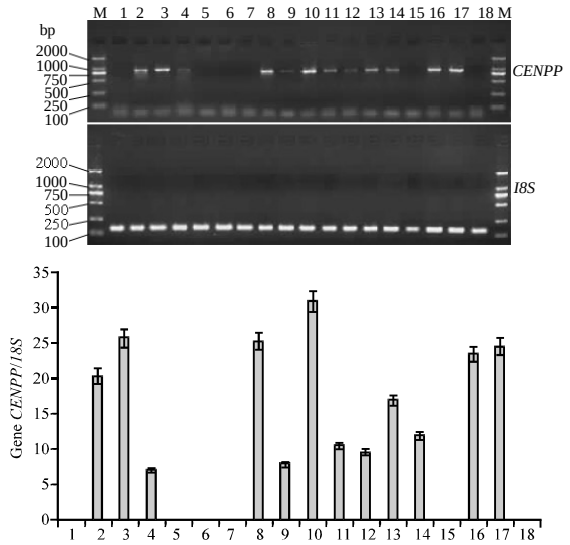

Fig. 6: Tissue expression distribution of BMI *CENPP* gene. The 18S expression is the internal control; Lymph node; 2: Midbrain; 3: Ovary; 4: Diencephalon; 5: Cerebrum; 6: Liver; 7: Kidney; 8: Spleen; 9: Heart; 10: Lung; 11: Nerve fiber; 12: Stomach; 13: Small intestine; 14: Large intestine; 15: Pancreas; 16: Skin; 17: Muscle and 18: Fat

expression profiles analysis. From the tissue expression profile analysis it can be seen that the gene was obviously differentially expressed in various tissues. As the researchers did not study functions at protein levels, there might be many possible reasons for differential expression of this porcine gene. The suitable explanation for this under current conditions is that the biological activities associated with the functions of the gene was demanded to a different extent in different tissues at the same time.

CONCLUSION

In this study, the research first isolated BMI *CENPP* gene and performed necessary functional analysis and tissue expression profile analysis. Furthermore, two miRNAs were found to have the corresponding target sites in the coding sequence of BMI CENPP by theoretical prediction. The cDNA clone, sequence information and function analysis of BMI *CENPP* gene will be extremely important in elucidating the essential physiological function of CENPP protein using BMI and other miniature swines as experimental animal models in the future.

ACKNOWLEDGEMENTS

The researchers Liu Lixian and Wang Pei contributed equally to this research and should be considered co-first researchers. This research was supported by the National Natural Science Foundation of China (Grant No. 31160439) and Natural Science Foundation of Yunnan province, China (Grant No. 2010ZC241, 2010ZC243).

REFERENCES

Ando, S., H. Yang, N. Nozaki, T. Okazaki and K. Yoda, 2002. CENP-A, -B and -C chromatin complex that contains the I-type α-satellite array constitutes the prekinetochore in HeLa cells. Mol. Cell Biol., 22: 2229-2241.

Bartel, D.P., 2004. MicroRNAs: Genomics, biogenesis, mechanism and function. Cell, 116: 281-297.

Hardison, R.C., 2003. Comparative genomics. PLoS Biol., 1: e58-e58.

Harris, I., 1997. Variables in animal based research part 1: Phenotypic variability in experimental animals. Anzccart News, 10: 1-7.

Liu, G.Y., S.Z. Gao, C.R. Ge and X. Zhang, 2008. Molecular characterization of the encoding regions and tissue expression analyses for three novel porcine genes-HNRPA1, YIPF5 and UB2D2. Mol. Biol. Rep., 35: 519-526.

Moller, S, M.D.R. Croning and R. Apweiler, 2001. Evaluation of methods for the prediction of membrane spanning regions. Bioinformatics, 17: 646-653.

Nakai, K. and P. Horton, 1999. PSORT: A program for detecting the sorting signals of proteins and predicting their subcellular localization. Trends Biochem. Sci., 24: 34-36.

Okada, M., I.M. Cheeseman, T. Hori, K. Okawa and I.X. McLeod *et al*., 2006. The CENP-H-I complex is required for the efficient incorporation of newly synthesized CENP-A into centromeres. Nat. Cell Biol., 8: 446-457.

Petersen, T.N., S. Brunak, G. von Heijne and H. Nielsen, 2011. SignalP 4.0: Discriminating signal peptides from transmembrane regions. Natur. Methods, 8: 785-786.

Pluta, A.F., A.M. Mackay, A.M. Ainszein, I.G. Goldberg and W.C. Earnshaw, 1995. The centromere: Hub of chromosomal activities. Science, 270: 1591-1594.

Rattner, J.R., 1987. The organization of the mammalian kinetochore: A scanning electron microscope study. Chromosoma, 95: 175-181.

Ris, H. and P.L. Witt, 1981. Structure of the mammalian kinetochore. Chromosoma, 82: 153-170.

Tomonaga, T., K. Matsushita, S. Yamaguchi, T. Oohashi and H. Shimada *et al*., 2003. Cancer Res. Overexpression and mistargeting of centromere protein-A in human primary colorectal. Cancer Res., 63: 3511-3516.

Wright, S., 1921. Systems of mating. I. The biometric relations between parent and offspring. Genetics, 6: 111-123.

Yu, P., L. Zhang, S. Li, Y. Li and J. Cheng *et al*., 2004. Screening and analysis of porcine endogenous retrovirus in Chinese banna minipig inbred line. Transplant. Proc., 36: 2485-2487.

Zeng, R. and Y.Z. Zeng, 2005. Molecular cloning and characterization of SLA-DR genes in the 133-family of the banna mini-pig inbred line. Anim. Genet., 36: 267-269.

Zeng, Y., R. Yi and B.R. Cullen, 2003. MicroRNAs and small interfering RNAs can inhibit mRNA expression by similar mechanisms. Proc. Natl. Acad. Sci., 100: 9779-9784.