Molecular Cloning, Sequence Characterization and Tissue Transcription Profile of a Novel Gene *PQBP1* from Chinese Banna Mini-Pig Inbred Line (BMI)

¹Jinlong Huo, ²Pei Wang, ¹Hongjiang Wei and ²Yangzhi Zeng
¹Key Laboratory of Banna Mini-Pig Inbred Line of Yunnan Province,
²Faculty of Animal Science and Technology, Yunnan Agricultural University,
650201 Kunming, Yunnan, P.R. China

Abstract: The complete coding sequence of Banna Mini-pig Inbred line (*BMI*) gene PQBP1 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 number JF750401. Sequence analysis revealed that the BMI PQBP1 encodes a protein of 265 amino acids that has high homology with the Polyglutamine Binding Protein1 (PQBP1) of five other species-human (95%), monkey (95%), cattle (95%), mouse (88%) and rat (87%). The phylogenetic tree analysis revealed BMI PQBP1 has a closer genetic relationship with the cattle PQBP1 than with those of human, monkey, mouse and rat. Analysis by RT-PCR showed that *BMI PQBP1* gene was over-expressed in midbrain, ovary, spleen, lung, nerve fiber, skin and muscle, moderately expressed in diencephalon, kidney, heart, stomach, small intestine, large intestine and fat, weakly expressed in lymph node and cerebrum and almost not expressed in liver and pancreas. One microRNA target site was predicted in the CDS of BMI PQBP1 mRNA for further studying this gene in the future. These data serve as a foundation for further insight into this swine gene.

Key words: Banna Mini-pig Inbred line (BMI), pig, PQBP1, tissue expression, bioinformatics analysis, China

INTRODUCTION

Polyglutamine (Q) tract Binding Protein1 (PQBP1) was isolated on the basis of its interaction with polyglutamine tracts and localizes predominantly to the nucleus where it suppresses transcriptional activation by a neuron-specific transcription factor, Brn-2. Its C-terminal domain is highly conserved and binds to a component of the spliceosome (Okazawa *et al.*, 2001, 2002). PQBP1 which also interacts with the polyglutamine tract of triplet repeat disease gene products, binds with a higher affinity to an expanded polyglutamine tract. PQBP1 has several functional domains including hepta and di-amino acid repeat sequences rich in polar residues essential for its interaction with the polyglutamine tract, a WWP/WW domain which binds to proline-rich motifs in other proteins, a putative nuclear localization signal sequence and a C2 domain implicated in Ca^{2+} dependent phospholipid signaling (Waragai *et al.*, 1999). PQBP1 also forms nuclear inclusion bodies which are similar to those nucleated by the protein products of polyglutamine disease genes. PQBP1 is highly conserved in simple animal metazoans and plants (*Caenorhabditis elegans* and Arabidopsis), it may perform a basic function in cells. PQBP1 might be a candidate molecule involved in the pathology of polyglutamine diseases (Okazawa *et al.*, 2001).

PQBP1 was identified as a binding protein to the polyglutamine tract present in various transcription-related factors and causative genes for neurodegenerative disorders (Waragai *et al.*, 2000; Iwamoto *et al.*, 2000). PQBP1 mutation affected males include mental retardation, microcephaly, short stature, spastic paraplegia and midline defects (Kalscheuer *et al.*, 2003).

Interaction between PQBP1 and mutant Ataxin-1 affects transcription and cell death (Okazawa *et al.*, 2002). PQBP1 mediates important cellular functions under physiological and pathological conditions via its interaction with polyglutaminetracts (Waragai *et al.*, 1999).

Corresponding Author: Yangzhi Zeng, Key Laboratory of Banna Mini-pig Inbred Line of Yunnan Province, 650201 Kunming, Yunnan, P.R. China

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 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 above described about the *PQBP1* gene, it is necessary to isolate this gene from pig for it is associated with metabolism, health and other important biological functions of animals. But until today, the porcine PQBP1 has not been reported yet. The objective of this study was to clone and analyze the full length coding sequence of *BMI PQBP1* 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 determine the tissue transcription profiles of BMI tissues. The data obtained will serve as a basis for understanding 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 snap 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). About 3 µg of RNA were reverse transcribed with oligo $(dT)_{18}$ primer and M-MLV reverse transcriptase (Invitrogen, USA).

Isolation of the *BMI PQBP1* gene: The GenBank PQBP1 sequences for cattle (Accession No.: NM_001046389) and their highly homologous pig ESTs sequences: BW973240, CK459475, CN157731, CN159667, DB817830, DN119942, DN120289 and FS702168 were used to design a primer pair to amplify the complete coding sequence of PQBP1 by using Primer Premier 5.0 Software. The primers for *BMI PQBP1* gene were: 5'- GGT CTG TCT GCT ACC CGC TAT -3' and 5'- GAA GGT TCA GTC CTG CTG CTT -3'. RT-PCR was performed to isolate the BMI PQBP1 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 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 PQBP1* gene was performed using software in NCBI (http://www.ncbi.nlm.nih.gov) and ExPaSy (http://www.expasy.org). The cDNA sequence was predicted using the GenScan Software (http://genes.mit.edu/GENSCAN.html). 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 tree was 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 (http://www.cbs.dtu.dk/services/SignalP/). The putative protein subcellular localization was predicted

J. Anim. Vet. Adv., 11 (10): 1674-1680, 2012

using PSort II (http://psort.hgc.jp/). Transmembrane topology prediction was performed using TMHMM-2.0 server (http://www.cbs.dtu.dk/services/TMHMM-2.0/). Secondary structures of deduced amino acid sequences were predicted with SOPMA (http://npsa-pbil.ibcp.fr/). Web-based microRNA (miRNA) predicting programs were used to locate conserved potential miRNA targets (http://www.mirbase.org/).

Semi-quantitative RT-PCR: To characterize the *PQBP1* 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 PQBP1 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 above. 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 PQBP1 cDNA: A PQBP1 fragment 823 bp of including the entire 798 bp coding region plus 19 bp of 5'-UTR and 6 bp of 3'-UTR was amplified (Fig. 1). This cDNA nucleotide sequence analysis using the BLAST software at NCBI server revealed that *BMI PQBP1* gene was not homologous to any of the known porcine genes and it was then deposited into the GenBank database (Accession No.: JF750401). The sequence prediction was carried out using the GenScan Software and results showed that the 798 bp cDNA sequence represents a single gene which encoded 265 amino acids. The complete CDS and the encoded amino acids were shown in Fig. 2.

Physical and chemical characteristics of BMI PQBP1: The theoretical Isoelectric point (pI) and Molecular weight (Mw) of PQBP1 were computed using the Compute pI/Mw tool. The theoretical pI and the molecular weight of BMI PQBP1 are 5.93 and 30571.39, respectively. The result from SignalP revealed that there is has no putative signal peptide in BMI PQBP1 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 PQBP1 was probably located in the nuclear with up to 73.9% 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 PQBP1 was not a potential transmembrane protein.

Prediction and analysis of structures and conserved domains of BMI PQBP1: 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
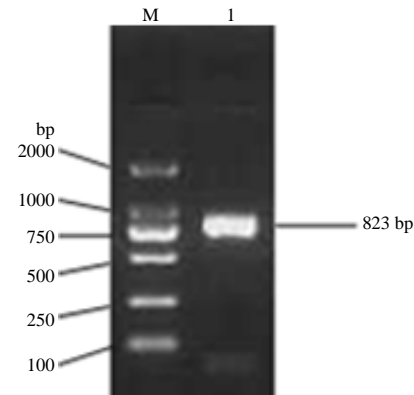

Fig. 1: RT-PCR result for *BMI PQBP1* gene; M: DL2000 DNA marker, 1: PCR product
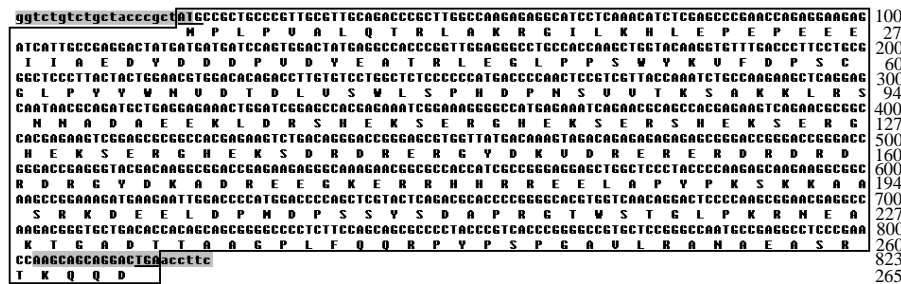

Fig. 2: The complete cDNA sequence and amino acid sequence of the protein encoded by PQBP1 (GenBank Accession Number: JF750401). ATG, start codon; TGA, stop codon; capital letters in pane, complete CDS and amino acid sequence; gray highlighted nucleotide sequence, primers

1676

that BMI PQBP1 contains one separated conserved domain with 2 conserved Trp (W) residues from 47-78 amino acid residues (also known as the WWP or rsp5 domain, binds proline-rich polypeptides, Fig. 3). The putative protein was also analyzed using prosite (http://expasy.org/prosite/). Four kinds sites were found which were Casein kinase II phosphorylation sites (43-TrlE-46, 77-SphD-80, 138-SdrD-141, 195-SrkD-198, 208-SysD-211, 229-TgaD-232); N-myristoylation sites (47-GLppSW-52, 215-GTwsTG-220, 230-GAdtTA-235); Protein kinase C phosphorylation sites (88-SaK-90, 110-SeR-112, 117-SeR-119, 124-SeR-126, 131-SeR-133, 138-SdR-140, 190-SkK-192, 195-SrK-197); Tyrosine kinase phosphorylation sites (140-Rdr. Erg.Y-146, 159-Rdr. Drg.Y-165, 180-Rre. ElapY-187). The prediction of secondary structure by SOPMA indicates that the deduced BMI PQBP1 consists of alpha helices (73AA), extended strands (22AA), beta turn (9AA) and random coils (161AA) (Fig. 4).

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 one *Sus scrofa* microRNA (ssc-miR-335) was found to have the predicted target sites (294-ucaggagcaauaacgcaga-312) in the BMI PQBP1 CDS sequence. Further investigation is needed to confirm whether corresponding miRNA molecules can regulate the *PQBP1* gene expression in swine.

Analysis of sequence identity and evolutionary relationships of BMI PQBP1: The deduced protein sequence of BMI PQBP1 was submitted to generate BLAST reciprocal best hits and similarity comparison revealed that BMI PQBP1 protein has high homology with the PQBP1 proteins of five other species-human (95%), monkey (95%), cattle (95%), mouse (88%) and rat (87%) (Fig. 5). To evaluate the evolutionary relationships of BMI
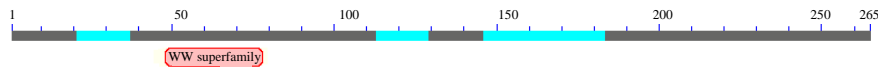

Fig. 3: The putative domains of the protein encoded by BMI PQBP1
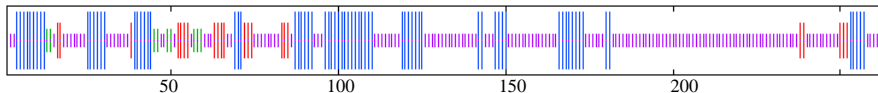

Fig. 4: The secondary structure of the BMI PQBP1 protein predicted by SOPMA. Helices, extended strands, beta turns and random coils are indicated, respectively with the longest, the second longest, the second shortest and the shortest vertical lines

Fig. 5: Alignment of the protein encoded by the BMI-pig PQBP1 and five other types of PQBPA from human (NP_005701), monkey (NP_001181187), cattle (NP_001039854), mouse (AAH51673) and rat (NP_001013979)

PQBP1 with other species then we constructed a phylogenetic tree using DNAstar, cluster, MEGA and DNAMAN Softwares on the basis of the PQBP1 amino acid sequences. The phylogenetic tree analysis revealed that the *BMI PQBP1* gene has a closer genetic relationship with the cattle *PQBP1* gene than with those of human, monkey, mouse and rat (Fig. 6). Therefore, researchers can use cattle as model organisms to study the pig *PQBP1* gene or use pig as a model organism to study the cattle *PQBP1* gene.

mRNA tissue-specific expression profile: To check the relative expression levels of PQBP1 mRNA in various porcine tissues and semi-quantitative RT-PCR was performed in 18 BMI tissues mentioned above. 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 PQBP1* gene over-expressed in midbrain, ovary, spleen, lung, nerve fiber, skin and muscle, moderately expressed in diencephalon, kidney, heart, stomach, small intestine, large intestine and fat, weakly expressed in lymph node and cerebrum and almost not expressed in liver and pancreas (Fig. 7). 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
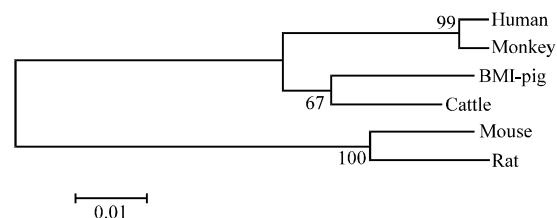

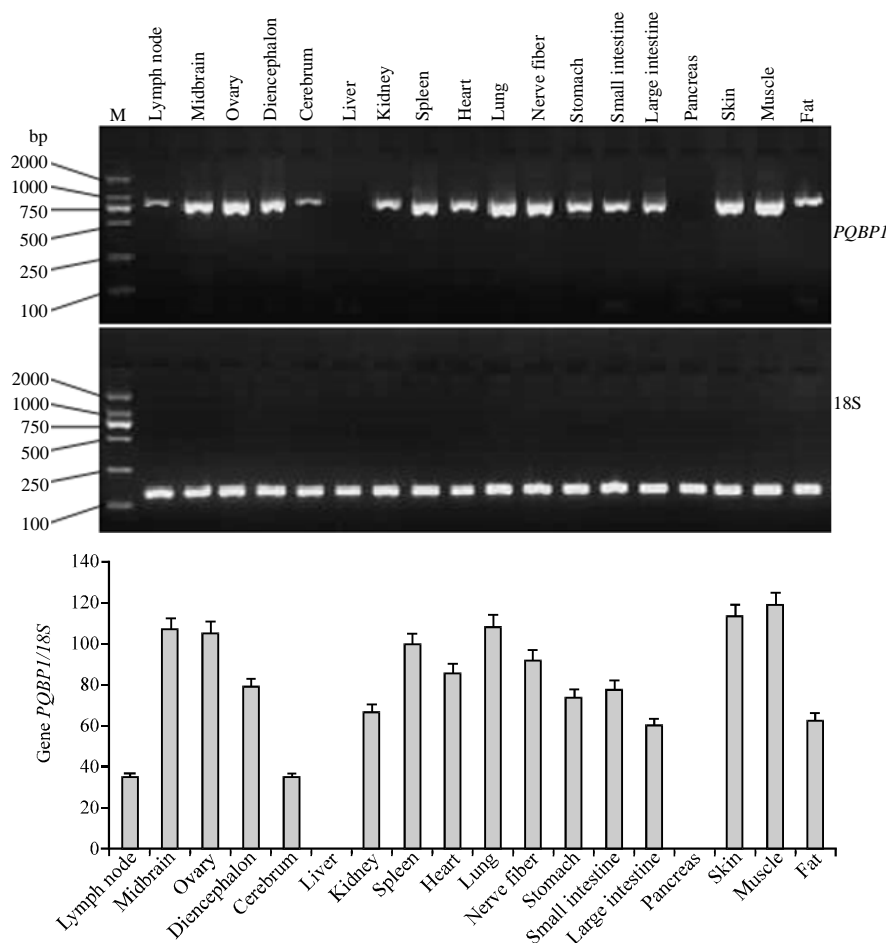Fig. 6: Phylogenetic tree for PQBP1 protein from 6 species

Fig. 7: Tissue transcription profile of *BMI PQBP1* gene. The 18S expression is the internal control

results have revealed that virtually all (99%) of the protein-coding genes in humans align with homologues in mice and >80% are clear 1:1 orthologs (Hardison, 2003; Liu *et al.*, 2008; Xi *et al.*, 2011). 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 *PQBP1* gene, researchers can find that swine PQBP1 is highly homologous with PQBP1 of cattle, human, monkey mouse, rat 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 PQBP1 protein we also find BMI PQBP1 protein was not completely identity to cattle or other mammals. This implied that BMI PQBP1 will have some differences in functions to those of cattle, human and other mammals. In this study, we not only cloned the CDS sequences of the *BMI PQBP1* gene but also conducted the sequence analysis and tissue 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 were demanded to a different extent in different tissues at the same time.

CONCLUSION

In this study, the research firstly isolated *BMI PQBP1* gene and performed necessary functional analysis and tissue expression profile analysis. Furthermore, one miRNA was found to have the corresponding target sites in the coding sequence of BMI PQBP1 by theoretical prediction. The cDNA clone, sequence information and function analyses of *BMI PQBP1* gene will be extremely important in elucidating the essential physiological function of PQBP1 protein using BMI and other miniature swines as experimental animal models in the future.

ACKNOWLEDGEMENTS

REFERENCES

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.

Iwamoto, K., Y.T. Huang and S. Ueda, 2000. Genomic organization and alternative transcripts of the human PQBP-1gene. Gene, 259: 69-73.

Kalscheuer, V.M., K. Freude, L. Musante, L.R. Jensen and H.G. Yntema *et al.*, 2003. Mutations in the polyglutamine binding protein 1 gene cause X-linked mental retardation. Nat. Genet., 35: 313-315.

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 sorting signals in proteins and predicting their subcellular localization. Trends Biochem. Sci., 24: 34-35.

Okazawa, H., M. Sudol and T. Rich, 2001. PQBP-1 (Np/PQ): A polyglutamine tract-binding and nuclear inclusion-forming protein. Brain Res. Bull., 56: 273-280.

Okazawa, H., T. Rich, A. Chang, X. Lin and M. Waragai *et al.*, 2002. Interaction between mutant ataxin-1 and PQBP-1 affects transcription and cell death. Neuron, 34: 701-713.

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.

Waragai, M., C. H. Lammers, S. Takeuchi, I. Imafuku and Y. Udagawa *et al.*, 1999. PQBP-1, a novel polyglutamine tract-binding protein, inhibits transcription activation by Brn-2 and affects cell survival. Human Mol. Genet., 6: 977-987.

Waragai, M., E. Junn, M. Kajikawa, S. Takeuchi and I. Kanazawa *et al.*, 2000. PQBP-1/Npw38, a nuclear protein binding to the polyglutamine tract, interacts with U5-15kD/dim1p via the carboxyl-terminal domain. Biochem. Biophys. Res. Commun., 273: 592-595.

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

Xi, D.M., Y.D. He, Y.K. Sun, X. Gou and S.L. Yang *et al.*, 2011. Molecular cloning, sequence identification and tissue expression profile of three novel genes *Sfxn1, Snai2* and Cno from Black-boned sheep (*Ovis aries*). Mol. Biol. Rep., 38: 1883-1887.

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.