



# Molecular cloning, sequence identification, polymorphism and association of the porcine *SPATS2L* gene

D. Yang<sup>1,2</sup> and Y. Liu<sup>2</sup>

<sup>1</sup>Yunnan Key Laboratory of Fertility Regulation and Minority Eugenics, Kunming, China

<sup>2</sup>Key Laboratory of Animal Nutrition and Feed of Yunnan Province, Yunnan Agricultural University, Kunming, China

Correspondence to: Y. Liu (liuyg4567@163.com)

Received: 22 October 2014 – Revised: 5 August 2015 – Accepted: 30 September 2015 – Published: 4 December 2015

**Abstract.** Spermatogenesis-associated, serine-rich 2-like (*SPATS2L*) is an important reproduction-related gene. In this study, the full-length cDNA sequence of the porcine *SPATS2L* gene was cloned through the RACE (rapid amplification of cDNA ends) method. The porcine *SPATS2L* gene encodes a protein of 559 amino acids which shares high homology with the *SPATS2L* proteins of seven species: dog (94%), white-tufted-ear marmoset (91%), human (92%), rhesus monkey (92%), horse (92%), rat (88%) and mouse (88%). This novel porcine gene was assigned to GeneID: 100415809. The phylogenetic analysis revealed that the porcine *SPATS2L* gene has a close genetic relationship with the canine *SPATS2L* gene. PCR-*Pst* I-RFLP was established to detect GU474997:c.1687 C>T substitution of porcine *SPATS2L* gene mRNA, and eight pig breeds displayed obvious genotype and allele frequency differences at this mutation locus. Association of this single-nucleotide polymorphism (SNP) with litter size traits was assessed in Large White ( $n = 100$ ) and Landrace ( $n = 100$ ) pig populations, and results demonstrated that this polymorphic locus was significantly associated with the litter size of all parities in Large White sows and Landrace sows ( $P < 0.01$ ). Therefore, *SPATS2L* gene could be a useful candidate gene in selection for increasing litter size in pigs.

## 1 Introduction

Spermatogenesis-associated, serine-rich 2-like (*SPATS2L*) is an important reproduction-related gene for it has been identified to be associated with the regulating embryonic development and regulating cellular processes ranging from apoptosis, cell cycle and adult stem cell homeostasis to patterning of the early embryo (Thomas and Voss, 2007; Kueh et al., 2011). However, recent research works showed that *SPATS2L* gene is involved in some important biological processes, such as DNA replication, chromatin modification, histone H3 acetylation, histone H4-K12 acetylation, histone H4-K16 acetylation, histone H4-K5 acetylation, histone H4-K8 acetylation, and regulation of transcription (DNA-dependent), and has important functions such as zinc ion binding, transferase activity, sequence-specific DNA-binding transcription factor activity, protein binding, metal ion bind-

ing, histone acetyltransferase activity and acyltransferase activity (Sharma et al., 2000; Burke et al., 2001; Yanagisawa et al., 2002; Wang et al., 2010; Miotto, and Struhl, 2010; Palacios et al., 2010; Wong et al., 2010).

As mentioned above, the *SPATS2L* gene is an important gene which has many biological functions. Until today, the *SPATS2L* gene has been reported in humans, mice, cattle and other animals, but the porcine *SPATS2L* gene has not been reported yet.

The aim of this work was to clone the full-length porcine *SPATS2L* mRNA sequence, search for polymorphism within this gene, and perform association analysis between a gene-tagged single-nucleotide polymorphism (SNP) and litter size in the Large White and Landrace sows.

## 2 Material and methods

### 2.1 Animals and sample preparation

Six adult Large White pigs were slaughtered. Samples of large intestine, spleen, lung, muscle, fat, liver, heart, kidney and ovary were collected, frozen in liquid nitrogen, and then stored at  $-80^{\circ}\text{C}$ . The total RNA was extracted using a total RNA extraction kit (Gibco, USA). These RNA samples were used to perform RACE PCR.

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

These 100 unrelated Large White sows (Table 1) were all raised in one farm under the same feeding condition. They were all artificially inseminated using semen from Large White boars. These 100 unrelated Landrace sows (Table 1) were all raised in another farm under the same feeding conditions. They were all artificially inseminated using semen from Landrace boars. Both the total number of piglets born (TNB) and the number of piglets born alive (NBA) of these Large White sows and Landrace sows were recorded for at least 1400 litters (at least 7 litter records for each sow). These litter size data and genomic DNA samples were then applied for performing association analysis.

### 2.2 5' and 3' RACE

The 5' and 3' RACE methods were performed to isolate the full-length cDNAs for the porcine *SPATS2L* gene according to the instructions of Clontech SMART<sup>TM</sup> RACE cDNA amplification kit (Clontech, USA). For the porcine *SPATS2L* gene, the gene-specific primers (GSPs) were designed based on one pig EST sequence whose sequence is highly homologous to the coding sequence of the human *SPATS2L* gene: FD640738; 5' RACE GSP: 5'-AGAGCATGATTTGGGCCTTTAGCTG-3'; 3' RACE GSP: 5'-TATGATGAAGAGCTCGGGAAAGCTG-3'. RACE touchdown PCRs were carried out with 5 cycles of  $94^{\circ}\text{C}$  (30 s) and  $72^{\circ}\text{C}$  (3 min), followed by 5 cycles of  $94^{\circ}\text{C}$  (30 s),  $69^{\circ}\text{C}$  (30 s) and  $72^{\circ}\text{C}$  (3 min), and finally with 30 cycles of  $94^{\circ}\text{C}$  (30 s),  $66^{\circ}\text{C}$  (30 s),  $72^{\circ}\text{C}$  (3 min) to terminate reaction. The RACE PCR products were then cloned into a 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.

### 2.3 Sequence analysis

The cDNA sequence analysis was conducted using GenScan software (<http://genes.mit.edu/GENSCAN.html>). The protein analysis was performed using the BLAST tool 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 ( $M_w$ ) of proteins were computed using the Compute pI /  $M_w$  tool ([http://www.expasy.org/tools/pi\\_tool.html](http://www.expasy.org/tools/pi_tool.html)).

### 2.4 PCR-RFLP

DNA samples from the abovementioned pigs (Table 1) were used as a template to perform PCR with primers 5'-GGG ACA AAG ATC CCC CGA-3' (forward primer 2) and 5'-GGC GGG GAT CCA AAA CTG-3' (reverse primer 2). The PCR product size is 258 bp. The 25  $\mu\text{L}$  reaction system contained 2.0  $\mu\text{L}$  of DNA (100 ng), 2.5  $\mu\text{L}$  of 2 mM mixed dNTPs, 2.5  $\mu\text{L}$  of  $10 \times$  Taq DNA polymerase buffer, 2.5  $\mu\text{L}$  of 25 mM  $\text{MgCl}_2$ , 1.0  $\mu\text{L}$  of 20  $\mu\text{M}$  forward primer, 1.0  $\mu\text{L}$  of 20  $\mu\text{M}$  reverse primer, 1.0 units of Taq DNA polymerase (1 U/ $\mu\text{L}$ ) (JINMEI, BIOTECH), and 12.5  $\mu\text{L}$  of sterile water. PCR was run as follows:  $94^{\circ}\text{C}$  for 4 min, followed by 35 cycles of  $94^{\circ}\text{C}$  for 50 s,  $60^{\circ}\text{C}$  for 50 s,  $72^{\circ}\text{C}$  for 1 min, then  $72^{\circ}\text{C}$  extension for 10 min, and finally  $4^{\circ}\text{C}$  to terminate reaction. The 31  $\mu\text{L}$  PCR-RFLP reaction volume consisted of 10  $\mu\text{L}$  of PCR product, 18  $\mu\text{L}$  of sterile water, 1  $\mu\text{L}$  of Pst I (10 U), and 2  $\mu\text{L}$  of  $10 \times$  buffer. The mixture was incubated in an air incubator at  $37^{\circ}\text{C}$  for 4 h, and then the genotypes were analyzed on agarose gel (2.5 %) containing ethidium bromide.

### 2.5 Statistical analysis

The associations between *SPATS2L* genotypes and litter size of Large White ( $n = 100$ ) and Landrace ( $n = 100$ ) sows were evaluated with the general linear model (GLM) procedure of SAS version 8.0. Both additive and dominance effects were also estimated using the REG procedure, where the additive effect was estimated as  $-1$ ,  $0$  and  $1$  for *CC*, *TC* and *TT* genotypes, respectively; the dominance effect was represented as  $1$ ,  $-1$  and  $1$  for *CC*, *TC* and *TT* genotypes, respectively (Zhang et al., 2009). The model was  $Y_{ijkl} = \mu + B_n + P_i + S_j + F_k + G_l + E_{ijkl}$ , where  $Y_{ijkl}$  is the observation of the trait,  $\mu$  is the least-squares means,  $B_n$  is the effect of the  $n$ th boar ( $n = 1, 2$  or  $3$ ),  $P_i$  is the effect of the  $i$ th parity ( $i = 1, 2, 3, 4, 5, 6$  or  $7$  (parity  $\geq 7$ )),  $S_j$  is the effect of the  $j$ th season ( $j = 1, 2, 3$  or  $4$ ; 1, 2, 3 and 4 represent spring, summer, autumn and winter, respectively),  $F_k$  is the effect of the  $k$ th farm ( $k = 1, 2$ ),  $G_l$  is the effect of the  $l$ th genotype ( $l = 1, 2$  or  $3$ ; 1, 2, and 3 represent *CC*, *TC* and *TT* genotype, respectively) and  $E_{ijkl}$  is the random residual (Niu et al., 2009). Hardy-Weinberg equilibrium was tested using the Hardy-Weinberg equilibrium calculator (<http://www.oege.org/software/hwe-mr-calc.shtml>).

**Table 1.** Information of 795 unrelated pigs from eight populations.

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

### 3 Results

#### 3.1 RACE results for the porcine *SPATS2L* gene

For the porcine *SPATS2L* gene, through 5' RACE, one PCR product of 1196 bp was obtained. The 3' RACE product was 1238 bp. These products were then cloned to the *T* vector and sequenced. Taken together, a 2361 bp cDNA complete sequence was finally obtained (Supplement 1).

#### 3.2 Sequence analysis

cDNA nucleotide sequence analysis revealed that this 2361 bp cDNA sequence was not homologous to any of the known pig genes, and it was then deposited in the GenBank database (accession number: GU474997). The sequence prediction was carried out using the GenScan software, and results showed that the 2361 bp cDNA sequence represented one single gene which encoded 559 amino acids (Supplement 2). The theoretical isoelectric point (pI) and molecular weight ( $M_w$ ) of this deduced protein of this pig gene were computed using the Compute pI /  $M_w$  tool. The pI of porcine *SPATS2L* is 9.67. The molecular weight of this putative protein is 61 610.63 Da. This novel porcine gene was assigned to GeneID: 100415809.

BLAST analysis revealed that the porcine *SPATS2L* protein shares high homology with the spermatogenesis associated, serine-rich 2-like (*SPATS2L*) protein of seven species: dog (accession number: XP\_536022; 94%), white-tufted-ear marmoset (accession number: XP\_002749655; 91%), human (accession number: BAG61258; 92%), rhesus monkey (accession number: XP\_001091367; 92%), horse (accession number: XP\_001502845; 92%), rat (accession number: NP\_001014124; 88%) and mouse (accession number: NP\_659131; 88%) (Supplement 3).

Based on the results of the alignment of *SPATS2L* proteins, a phylogenetic tree was constructed using the Dendrogram procedure of ClustalW software (<http://align.genome.jp/>), as shown in Supplement 4.

The phylogenetic analysis revealed that the porcine *SPATS2L* gene has a close genetic relationship with the canine *SPATS2L* gene.

#### 3.3 Polymorphism

Based on the sequencing results of the porcine *SPATS2L* gene, one *T–C* mutation was found at the position of 1687 bp of the mRNA sequence. This led to mutation of one *Pst* I restriction site. This was confirmed by PCR-*Pst* I-RFLP (Supplement 5).

Subsequently, PCR-*Pst* I-RFLP was performed using the DNA from 795 unrelated animals belonging to eight pig populations: Large White pig, Landrace pig, Saba pig, Tibetan pig, Mingguang small-ear pig, Diannan small-ear pig, Wujin pig and Baoshan pig. Results revealed that frequency of the *T* allele is predominantly higher than the frequency of the *C* allele in all eight pig populations. The frequency of *C* allele in the two exotic pig breeds – Large White pig (0.075) and Landrace pig (0.135) – is higher than that in the other six Yunnan local pig breeds: Saba pig (0.010), Tibetan pig (0.011), Mingguang small-ear pig (0.020), Diannan small-ear pig (0.040), Wujin pig (0.030) and Baoshan pig (0.025) (Table 2). The *CC* genotype was not detected in any of the six Yunnan local pig breeds and was only found in a small number of animals of *CC* genotype in Large White pigs and Landrace pigs. This indicates that Yunnan local pig breeds and exotic pig breeds displayed obvious genotype and allele frequency differences at this *T–C* mutation locus. The chi-square Hardy–Weinberg equilibrium test showed that six Yunnan local pig breeds all fit with the Hardy–Weinberg equilibrium ( $P > 0.05$ ) except two exotic pig breeds (Large White pig and Landrace pig) ( $P < 0.05$ ).

For the litter size of the first parity, no significant difference was found among the animals of the three genotypes in the Large White sows and Landrace sows. For the litter size of all parities, in the Large White sows, those with the *CC* genotype had an additional 0.677 piglets born compared to the *TT* animals ( $P < 0.01$ ) and 0.745 piglets born compared to the *TC* animals ( $P < 0.01$ ). In addition, for the litter size

**Table 2.** Allele frequency of the *SPATS2L* gene in eight pig breeds.

Breed	Number of pigs	Genotype			Allele frequency		Hardy–Weinberg equilibrium $\chi^2$
		<i>TT</i>	<i>TC</i>	<i>CC</i>	<i>T</i>	<i>C</i>	
Large White	100	88	7	5	0.925	0.075	30.25( $P < 0.05$ )
Landrace	100	82	9	9	0.875	0.135	37.78( $P < 0.05$ )
Saba	100	98	2	0	0.990	0.010	0.01( $P > 0.05$ )
Tibetan	95	93	2	0	0.989	0.011	0.01( $P > 0.05$ )
Mingguang small-ear	100	96	4	0	0.980	0.020	0.04( $P > 0.05$ )
Diannan small-ear	100	92	8	0	0.960	0.040	0.17( $P > 0.05$ )
Wujin	100	94	6	0	0.970	0.030	0.10( $P > 0.05$ )
Baoshan	100	95	5	0	0.975	0.025	0.07( $P > 0.05$ )

**Table 3.** Association analyses between PCR-*Pst* I-RFLP genotypes and litter size of the porcine *SPATS2L* gene.

Breed	Traits	Genotype (mean $\pm$ standard error)			Genetic effects (mean $\pm$ standard error)		
		<i>TT</i>	<i>TC</i>	<i>CC</i>	Additive	Dominant	
Large White	First parity	N	88	7	5		
		TNB	10.325 $\pm$ 0.350	10.312 $\pm$ 0.412	10.278 $\pm$ 0.342	-0.024 $\pm$ 0.018	0.011 $\pm$ 0.010
		NBA	9.105 $\pm$ 0.442	9.101 $\pm$ 0.383	9.094 $\pm$ 0.447	-0.006 $\pm$ 0.023	0.002 $\pm$ 0.003
	All parities	N	88	7	5		
		TNB	11.573 $\pm$ 0.433 <sup>A</sup>	11.505 $\pm$ 0.513 <sup>A</sup>	12.250 $\pm$ 0.415 <sup>B</sup>	0.338 $\pm$ 0.112	-0.407 $\pm$ 0.355
		NBA	10.759 $\pm$ 0.372	10.810 $\pm$ 0.505	10.930 $\pm$ 0.408	0.086 $\pm$ 0.109	-0.035 $\pm$ 0.210
Landrace	First parity	N	82	9	9		
		TNB	10.124 $\pm$ 0.537	10.538 $\pm$ 0.589	10.302 $\pm$ 0.520	0.089 $\pm$ 0.032	0.325 $\pm$ 0.342
		NBA	9.290 $\pm$ 0.489	9.575 $\pm$ 0.528	9.328 $\pm$ 0.523	0.019 $\pm$ 0.009	0.266 $\pm$ 0.321
	All parities	N	82	9	9		
		TNB	11.448 $\pm$ 0.374 <sup>A</sup>	11.335 $\pm$ 0.257 <sup>A</sup>	12.129 $\pm$ 0.423 <sup>B</sup>	0.341 $\pm$ 0.327	-0.454 $\pm$ 0.528
		NBA	10.852 $\pm$ 0.410	10.875 $\pm$ 0.326	11.001 $\pm$ 0.631	0.075 $\pm$ 0.051	-0.052 $\pm$ 0.118

N: number of investigated litters. Least-squares mean values with different letters are significantly different:  $P < 0.01$ .

of all parities, in the purebred Landrace sows, those with the *CC* genotype had an additional 0.681 piglets born compared to the *TT* animals ( $P < 0.01$ ) and 0.794 piglets born compared to the *TC* animals ( $P < 0.01$ ) (Table 3).

#### 4 Discussion

Through sequence analysis, it can be seen that the encoding protein of the porcine *SPATS2L* gene is highly homologous with *SPATS2L* proteins of dogs, humans, mice and other mammals. This implies that the *SPATS2L* genes were highly conserved in some mammals and the porcine *SPATS2L* gene might have similar functions to the *SPATS2L* genes of dogs, humans, mice and other mammals. It can be also found that the porcine *SPATS2L* protein does not show complete identity to humans, mice or other mammals. This suggests that the porcine *SPATS2L* gene might have some differences in functions to those of dogs, humans, mice or other mammals. From phylogenetic analysis we found that the porcine

*SPATS2L* gene has a close genetic relationship with the canine *SPATS2L* gene; this implies that we can use dogs as a model organism to study the porcine *SPATS2L* gene.

In the current study, we firstly obtained the full length of *SPATS2L* gene cDNA from pigs by RACE-PCR. Involvement of the *SPATS2L* gene in the reproduction process was a molecular basis for association analysis of this gene DNA polymorphism with litter size of pig (Thomas and Voss, 2007; Kueh et al., 2011). Based on association analysis of the SNP and litter size, it could be found that the polymorphism (GU474997:c.1687  $C > T$ ) of the porcine *SPATS2L* gene can significantly affect litter size. The *TT* genotype animals obviously have better litter size for all parities than the *CC* and *TC* animals both in purebred Large White and purebred Landrace sows. This indicates that this polymorphic locus of the porcine *SPATS2L* gene is a valuable marker that deserves to be applied to the marker assistant selection (MAS) in pig breeding. Therefore, the *SPATS2L* gene could be a useful candidate gene in selection for increasing litter size in pigs.

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 a low number of *CC* genotype animals were found in the Large White and Landrace populations. The chi-square Hardy–Weinberg equilibrium test also showed that Large White and Landrace populations did not fit with the Hardy–Weinberg equilibrium ( $P < 0.05$ ). This, whether it affected the association analysis results or not, should be validated in a future study using a larger sample size.

In conclusion, we first isolated the porcine *SPATS2L* gene and performed necessary sequence analysis, polymorphism analysis and association analysis. This established the primary foundation for further insight into this novel pig gene.

**The Supplement related to this article is available online at doi:10.5194/aab-1-445-2015-supplement.**

**Acknowledgements.** This work was supported by grants from the Open Project of Yunnan Key Laboratory of Fertility Regulation and Minority Eugenics and the Program for Innovative Research Team (in Science and Technology) in University of Yunnan Province (no. 2014-22).

Edited by: K. Wimmers

Reviewed by: two anonymous referees

## References

- Burke, T. W., Cook, J. G., Asano, M., and Nevins, J. R.: Replication factors MCM2 and ORC1 interact with the histone acetyltransferase HBO1, *J. Biol. Chem.*, 276, 15397–15408, 2001.
- Kueh, A. J., Dixon, M. P., Voss, A. K., and Thomas, T.: HBO1 Is Required for H3K14 Acetylation and Normal Transcriptional Activity during Embryonic Development, *Mol. Cell Biol.*, 31, 845–860, 2011.
- Miotto, B. and Struhl, K.: HBO1 Histone Acetylase Activity Is Essential for DNA Replication Licensing and Inhibited by Geminin, *Mol. Cell*, 37, 57–66, 2010.
- Niu, B. Y., Ye, L. Z., Li, F. E., Deng, C. Y., Jiang, S. W., Lei, M. G., and Xiong, Y. Z.: Identification of polymorphism and association analysis with reproductive traits in the porcine RNF4 gene, *Anim. Reprod. Sci.*, 110, 283–292, 2009.
- Palacios, A., Moreno, A., Oliveira, B. L., Rivera, T., Prieto, J., Garcia, P., Fernandez-Fernandez, M. R., Bernado, P., Palmero, I., and Blanco, F. J.: The Dimeric Structure and the Bivalent Recognition of H3K4me3 by the Tumor Suppressor ING4 Suggests a Mechanism for Enhanced Targeting of the HBO1 Complex to Chromatin, *J. Mol. Biol.*, 396, 1117–1127, 2010.
- Sharma, M., Zarnegar, M., Li, X. Y., Lim, B., and Sun, Z. J.: Androgen receptor interacts with a novel MYST protein, HBO1, *J. Biol. Chem.*, 275, 35200–35208, 2000.
- Thomas, T. and Voss, A. K.: The diverse biological roles of MYST histone acetyltransferase family proteins, *Cell Cycle*, 6, 696–704, 2007.
- Wang, W. Z., Liu, H. O., Wu, Y. H., Hong, Y., Yang, J. W., Liu, Y. H., Wu, W. B., Zhou, L., Sun, L. L., Xu, J. J., Yun, X. J., and Gu, J. X.: Estrogen receptor alpha (ER alpha) mediates 17 beta-estradiol (E2)-activated expression of HBO1, *J. Exp. Clin. Canc. Res.*, 29, 140, 2010.
- Wong, P. G., Glozak, M. A., Cao, T. V., Vaziri, C., Seto, E., and Alexandrow, M. G.: Chromatin unfolding by Cdt1 regulates MCM loading via opposing functions of HBO1 and HDAC11-geminin, *Cell Cycle*, 9, 4351–4363, 2010.
- Yanagisawa, J., Kitagawa, H., Yanagida, M., Wada, O., Ogawa, S., Nakagomi, M., Oishi, H., Yamamoto, Y., Nagasawa, H., McMahon, S. B., Cole, M. D., Tora, L., Takahashi, N., and Kato, S.: Nuclear receptor function requires a TFTC-type histone acetyltransferase complex (retracted article. see vol. 54, p. 536, 2014), *Mol. Cell*, 9, 553–562, 2002.
- Zhang, Y. H., Mei, S. Q., Peng, X. W., Niu, B. Y., Ren, Z. Q., Zuo, B., Xu, D. Q., Lei, M. G., Zheng, R., Jiang, S. W., Deng, C. Y., Xiong, Y. Z., and Li, F. E.: Molecular characterization and SNPs analysis of the porcine Deleted in AZoospermia Like (pDAZL) gene, *Anim. Reprod. Sci.*, 112, 415–422, 2009.