BF-1507
[1-10]

RESEARCH ARTICLE Indian Journal of Animal Research

Three SINE Insertion Polymorphic Sites were Identified in Insulin-like Growth Factor 2 mRNA-binding Proteins

Mengli Wang[1#], Yao Zheng[1#], Ali Shoaib Moawad[1,2], Cai Chen[1], Chenglin Chi[1], Xiaoyan Wang[1], Chengyi Song[1]

10.18805/IJAR.BF-1507

ABSTRACT

Background: The insulin-like growth factor-2 mRNA-binding proteins 1, 2 and 3 (IGF2BP1, IGF2BP2, IGF2BP3) belong to a conserved family of RNA-binding, oncofetal proteins, these RNA-binding proteins (RBPs) modulate important aspects of cell function during development and in cancer. However, the structural variations of IGF2BPs gene generated by retrotransposon insertion have not yet been reported.

Methods: In this study, the bioinformatic prediction was performed to screen for retrotransposon insertion polymorphisms (RIPs) in IGF2BP genes.Sixteen predicted RIPs in IGF2BP genes were identified and three RIPs caused by the youngest SINEA1 retrotransposons and located in IGF2BP3 introns were verified by PCR.

Result: Polymorphisms of these three RIPs in commercial breeds are poor, but in Chinese native pig breeds, all three RIPs showed abundant polymorphisms. This is consistent with the intensive selection of commercial pigs. In summary, our data suggested that there are at least three RIPs caused by SINE retrotransposons in the IGF2BP3 gene. And they shows different polymorphic distribution in Chinese native and commercial breeds, suggesting that they can be used for population genetic analysis.

Key words: IGF2BP3, Insertion polymorphic, Pig, Population genetic analysis, SINE.

INTRODUCTION

Insulin-like growth factor 2 (IGF2) messenger RNA (mRNA)-binding proteins (IGF2BPs) are RNA-binding proteins (RBPs) that regulate RNA processing at multiple levels, including localization, translation and stability. IGF2BP1, IGF2BP2 and IGF2BP3 (insulin-like growth factor-2 mRNA-binding proteins 1, 2 and 3) are members of a highly conserved protein family that can bind RNA and alter the fate of their transcript targets. Many animals, including the cow (Kirby *et al.* 1996), sheep (Ko *et al.* 1991; Reynolds *et al.* 1997), pig (Hofig *et al.* 1991), horse (Lennard *et al.* 1995), human (Zhou *et al.* 1994) and rodents (DeChiara *et al.* 1990), have an insulin-like growth factor (IGF) system in their uterine or conceptus environment. IGFBPs are important regulators of IGF function, which influence the bioavailability and interaction of IGF-I and IGF-II with cellular receptors (Cohick and Clemmons 1993). IGF2BP1 has been reported to be the main gene responsible for the body size and plumage color of Peking ducks in a recent whole-genome analysis (Zhou *et al.* 2018)ÿand novel 15 and 5 bp insertions/deletions (InDels) within the IGF2BP1 gene were identified and were found to be significantly associated with growth traits of Shaanbei white cashmere goats (Wang *et al.* 2020). While in sheep, nine InDels mutations within IGF2BP1 were identified, three loci were polymorphic and the three InDels were crucial variants correlated with growth traits and could be applied in marker-assisted selection (MAS) in sheep (Liu *et al.* 2021).

Transposable elements account for about 40% of the pig genome, of which retrotransposons account for more than 90% of total transposable elements (Chen *et al.* 2019). Retrotransposons are one of the most important sources of genetic diversity, which can produce structural changes in sequences, including deletion, inversion, displacement and breaking (Cordaux and Batzer 2009). According to comparative genomics, there are at least 8,000 structural changes associated with transposons in the human genome (Xing *et al.* 2009). Zhao *et al* discovered that retrotransposons were responsible for more than half of the sequence deletions in the pig genome (Zhao *et al.* 2016). The phenotypic can be affected directly or indirectly by retrotransposons inserted into animal genes. For example, the characteristics of chicken hen feathers are related to the insertion of a complete avian leukemia virus in the 5'UTR of CYP19A1 (Li *et al.* 2019). The degree of dilution of the dog's coat color will be affected by inserting the SINE retrotransposon into the exon of the PMEL gene

[1]College of Animal Science and Technology, Yangzhou University, Yangzhou 225009, China.
[2]Department of Animal Production, Faculty of Agriculture, Kafrelsheikh University. Kafrelsheikh 33516, Egypt.
[#]Thease authors contributed equally to this work.

Corresponding Author: Chengyi Song, College of Animal Science and Technology, Yangzhou University, Yangzhou 225009, China. Email: cysong@yzu.edu.cn

How to cite this article: Wang, M., Zheng, Y., Moawad, A.S., Chen, C., Chi, C., Wang, X. and Song, C. (2022). Three SINE Insertion Polymorphic Sites were Identified in Insulin-like Growth Factor 2 mRNA-binding Proteins. Indian Journal of Animal Research. DOI: 10.18805/IJAR.BF-1507.

Submitted: 22-02-2022 **Accepted:** 13-05-2022 **Online:** 23-06-2022

(Murphy *et al.* 2018). Diverse mutations (SNPs) within MC1R, TYRP1, ASIP genes have been reported and they may be associted with phenotypes of coat color in wild pigs (Yang *et al.* 2019), furthermore, at least 40 RIPs in coat color genes have been reported and some RIPs may be associated with the coat color differences between different domesticated pig breeds (Du *et al.* 2022).

In summary, transposable elements are an important part of the pig genome and its insertion polymorphism is not only the raw material for molecular markers but also the entry point for studying gene structural variation. Eventhough an important functional gene of the IGF pathway, IGF2BP has no relevant research reports on whether there are structural variations mediated by retrotransposons in the IGF2BP genes in different breeds and their distribution. It is urgent to explore and reveal it to provide a basis for further research about IGF2BPs.

MATERIALS AND METHODS

Experiment material

There were 12 pig breeds involved in this experiment, including Duroc, Landrace, Large White, Sujiang pigs, Banna pigs, Erhualian pigs, Wuzhishan pigs, Bama pigs, Tibetan pigs, Meishan pigs, Fengjing pigs and Wild boars. Ear tissues were collected in parallel to agricultural procedures *(i.e.,* pulling in ear tags). All collected samples were stored at -80°C and detailed information is shown in Table S1. This experiment was carried out from September 2020 to August 2021 in the Laboratory of in College of Animal Science and Technology, Yangzhou University.

Primer design and synthesis

Primers were designed by Oligo7 software according to Fig 1A. For a SINE insertion site, the primers were designed in the flanking regions at both ends of the SINE insertion position. Only one larger band could be got for the homozygous SINE insertion (+/+) individuals and only one smaller band for homozygous SINE deletion (-/-) individuals and two bands will be got for the heterozygotes (+/+) (Fig 1B). The LINE and endogenous retrovirus (ERV) inserts involved in this article are less than 2000bp and the primers are also designed according to the SINE insertion detection method. Primers were synthesized by Nanjing Kinco Biotech Co., Ltd (Nanjing, China) and their coordinates in the reference genome, annealing temperature and sequences were supplied in Table S2. The PCR (Kaushik *et al.* 2017) amplification time was determined according to the target fragment size.

RIP annotation of porcine IGF2BP genes

Sequence acquisition of porcine IGF2BP genes and their flanks

The online website Ensembl (http://asia.ensembl.org/index.html) was used to obtain the sequences of the 3 genes of Duroc pig IGF2BP as reference sequences (IGF2BP1: ENSSSCG0000002312; IGF2BP2: ENSSSCG00000011795; IGF2BP3: ENSSSCG00 0000366 95) and the sequence was extended to the 5′flanking region and 3′flanking region by 5000 bp and 3000 bp, respectively. NCBI Blast (https://blast.ncbi.nlm. nih.gov/Blast.cgi?PROG RAM=blastnand PAGE _TYPE= BlastSearchand LINK_ LOC =blasthome) was used to compare the reference sequences with the pig non-reference genomes in the WGS library and obtain the gene sequence fragments of the 3 genes of pig IGF2BP in the non-reference genomes. Finally, the sequence fragments were spliced together according to the reference sequence and a complete sequence from each genome for each gene was obtained.

Retrotransposon annotation

RepeatMasker (Caballero *et al.* 2014) (versions: 4.0.7, -cutoff 250 -nolow) was used to perform retrotransposon annotation on the gene (including flanking region) sequences of all genomes obtained and only retaining the annotations with an alignment score of more than 1000 and size more than 100 bp. Multiple sequence alignment of each gene got from 16 genomes was done by using Clustalx (2.0 version) software, then the structural variations (more than 50 bp) were

Table S1: Detailed information of samples.

Breeds name	Sample type	Source	Samples quantity
Duroc	Ear tissue	Anhui Academy of Agricultural Sciences	6
Landrace	Ear tissue	A certain pig company in Anhui Province	38
Large white	Ear tissue	Anhui Academy of Agricultural Sciences	38
Sujiang pigs	Ear tissue	Jiangsu Sujiang Breeding Pig Co., Ltd.	38
Banna pigs	Ear tissue	Yunnan Province Xishuangbanna Prefecture Diannan Small Ear Pig Breeding Plant	6
Erhualian pigs	Ear tissue	Sutai Pig Original Breeding Farm in Suzhou, Jiangsu	38
Wuzhishan pigs	Ear tissue	Hainan Wuzhishan Pig Breeding Farm	6
Bama pigs	Ear tissue	Bama Original Fragrant Pig Farming and Animal Husbandry Industry Co., Ltd.	6
Tibetan pigs	Ear tissue	Animal Husbandry Institute of Ganzi City, Yunnan Province	6
Meishan pigs	Ear tissue	Sutai Pig Original Breeding Farm in Suzhou, Jiangsu	6
Fengjing pigs	Ear tissue	Sutai Pig Original Breeding Farm in Suzhou, Jiangsu	38
Wild boars	Ear tissue	Anhui Province	6
Sushan	Ear tissue	Jingjiang Lvye Ecological Science and Technology Park, Jiangsu	32

Table S2: The detail information of sequences used for conservation analysis for the IGF2BPs gene.

Gene name	Breeds name	Ensembl ID	Genome version	Chromosome	Start	End	Transcript ID
IGF2BP1	Pig	ENSSSCG00000023128	Sscrofa11.1	Chromosome 12	25191969	25246272	ENSSSCT00000022825.3
	Chacoan peccary	ENSCWAG00000022044	CatWag_v2_BIUU_UCD	PVHT020001179.1	9495091	9549394	ENSCWAT00000031915.1
	Cow	ENSBTAG00000011736	ARS-UCD1.2	Primary_assembly 19	37494583	37548886	ENSBTAT00000015586.5
	Sheep	ENSOARG00000006509	Oar_v3.1	Chromosome 11	37047348	37101651	ENSOART00000007081.1
	Horse	ENSECAG00000009723	EquCab3.0	Primary_assembly 11	25021727	25076030	ENSECAT00000010381.3
	Dog	ENSCAFG00000016907	CanFam3.1	Chromosome 9	25195874	25250177	ENSCAFT00000026763.4
	Human	ENSG00000159217	GRCh38	Chromosome 17	48995482	49049785	ENST00000290341.8
	Mouse	ENSMUSG00000013415	GRCm38	Chromosome 11	95951167	96005470	ENSMUST00000013559.3
IGF2BP2	Pig	ENSSSCG00000011795	Sscrofa11.1	Chromosome 13	1234806241	23653039	ENSSSCT00000035397.3
	Chacoan peccary	ENSCWAG00000018377	CatWag_v2_BIUU_UCD	PVHT021183693.1	41349858	41585892	ENSCWAT00000026332.1
	Cow	ENSBTAG00000007666	ARS-UCD1.2	Primary_assembly 1	81366734	81615320	ENSBTAT00000036032.5
	Sheep	ENSOARG00000020556	Oar_v3.1	Chromosome 1	199583327	199809947	ENSOART00000022394.1
	Horse	ENSECAG00000020685	EquCab3.0	Primary_assembly 19	26182441	26402784	ENSECAT00000072803.1
	Dog	ENSCAFG00000013297	CanFam3.1	Chromosome 34	18346126	18558100	ENSCAFT00000021148.4
	Human	ENSG00000073792	GRCh38	Chromosome 3	185610161	185872345	ENST00000382199.7
	Mouse	ENSMUSG00000033581	GRCm38	Chromosome 16	22013991	22208183	ENSMUST00000100052.11
IGF2BP3	Pig	ENSSSCG00000036695	Sscrofa11.1	Chromosome 9	92099847	92275211	ENSSSCT00000043710.2
	Chacoan peccary	ENSCWAG00000001238	CatWag_v2_BIUU_UCD	PVHT020000173.1	11250528	11425892	ENSCWAT00000001652.1
	Cow	ENSBTAG00000019406	ARS-UCD1.2	Primary_assembly 4	31922092	32097456	ENSBTAT00000025854.6
	Sheep	ENSOARG00000012961	Oar_v3.1	Chromosome 4	31890885	32066249	ENSOART00000014090.1
	Horse	ENSECAG00000016797	EquCab3.0	Primary_assembly 4	55050918	55226282	ENSECAT00000018011.3
	Dog	ENSCAFG00000002766	CanFam3.1	Chromosome 14	36951085	37126449	ENSCAFT00000046043.1
	Human	ENSG00000136231	GRCh38	Chromosome 7	23299691	23475055	ENST00000258729.8
	Mouse	ENSMUSG00000029814	GRCm38	Chromosome 6	49059351	49234715	ENSMUST00000031838.9

counted manually. All statistics are recorded but they may not be accurate due to the uncertainty of sequencings, such as the existence of gap or long sequence N, which had not been counted. Then, the structural variations that overlap with retrotransposons over 80% of the length were identified as predicted retrotransposon insertion polymorphic sites (RIPs) and be further verified by PCR.

Verification for the predicted RIPs in porcine IGF2BPs gene

Two DNA pools were prepared for each breed after DNA was extracted from the above 12 breeds by using the Tiangen Genome Extraction Kit (DP304) kit and the extraction steps were carried out strictly following the instructions. Each pool is made up of equal amounts of DNA from three individuals and the final concentration was adjusted to 40 ng/μL. Using pool DNA as a template, PCR amplification was performed to verify each predicted RIP (Zheng *et al.* 2020) (Fig 2).

Conservation analysis of porcine IGF2BP genes

The Fasta format file of the corresponding gene sequence of each gene in 7 different species (cow, sheep, dog, horse, human, mouse. See Table S2 for detailed species information) and the annotation information of each porcine IGF2BP gene was downloaded from Ensembl. Then the conservation analysis was done using the online program of mVISTA (http://genome.lbl.gov/vista/mvista/submit.shtml). The gene sequence of the pig genome was used as a reference sequence for conservative analysis and a conservative peak map was generated in Fig S2.

Population genetic analysis

Six pig breeds information of the samples is shown in Table S1. HWE and Polymorphic information content (PIC) analysis. The genotype and allele frequencies were calculated and Hardy-Weinberg equilibrium (Jadhav *et al.* 2020) was tested using the chi-square test in the Popgene32 software (Yeh *et al.* 1999). PIC was calculated according to the formula:

$$PIC = 1 - \sum_{i=1}^{n} p_i^2 - \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} 2p_i^2 p_j^2$$

Linkage disequilibrium for four RIPs in GHR genes was performed by Haploview (Barrett *et al.* 2005).

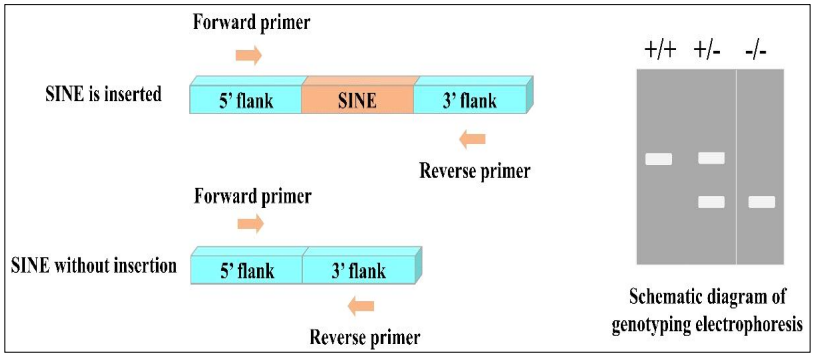

Fig 1: Primers design methods for RIPs verification (A) and Schematic diagram of genotyping electrophoresis (B).

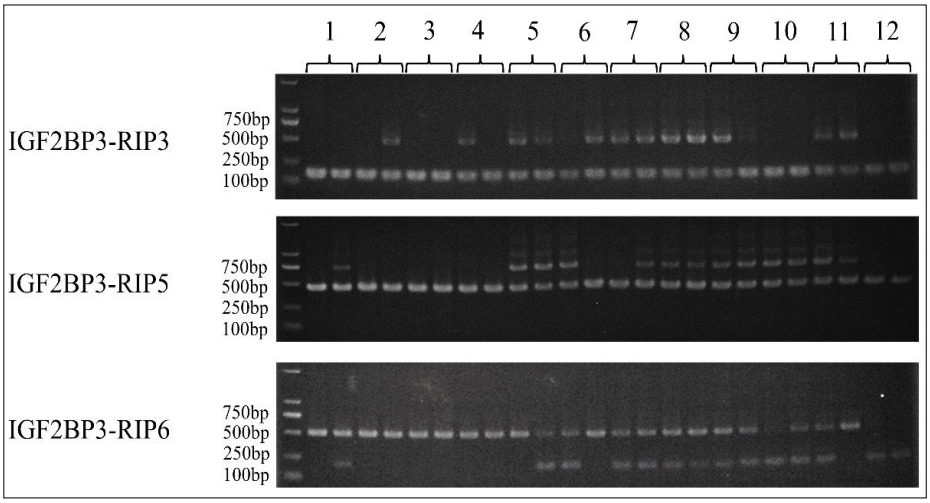

Fig 2: Polymorphism detection results of the three RIPs by PCR with 12 DNA pools. The names of the pool DNA are 1: Duroc, 2: Landrace, 3: Large White, 4: Sujiang pigs, 5: Banna pigs, 6: Erhualian pigs, 7: Wuzhishan pigs, 8: Bama pigs, 9: Tibetan pigs, 10: Meishan pigs, 11: Fengjing pigs, 12: Wild boars.

RESULTS AND DISCUSSION

The sequence of the IGF2BPs gene and their flanks (including the 5kb 5′flanking and 3kb 3′flanking region) from the pig reference genome and 15 non-reference genomes were obtained by Blast and spliced, the genomic coordinates of the analyzed IGF2BPs genes and their flanking sequences were summarized in Table S3 and then the length of IGF2BPs genes are counted and shown in Table 1. For the IGF2BP3 gene in Göttingen pigÿit showed a long gap in the end/start after the alignment, so the effective sequence of Göttingen pig was only 55445bp. These three genes show different lengths in different genomes, indicating that there are structural variations in genes. And in Tibetan pigs, these three genes all show a relatively longer length, suggesting that they have more insertion mutations, which are different from other breeds. According to the multiple sequence alignment results obtained by Clustalx, the structural variation in each gene was counted. In IGF2BP1, IGF2BP2 and IGF2BP3 genes and their flanks, 17, 29 and 22, SVs were identified. Then, the RepeatMasker results were used for identifying the structure variation caused by retrotransposons. The results are shown in Table S3. Among these SVs, 4, 6 and 6 SVs were thought to be caused by retrotransposons which be named predicted RIPs. To confirm the mutant sequence is a retrotransposon, we cloned and sequenced the insertion allele and deletion allele sequence for the above three RIPs. As shown in Fig S1, sequencing results show that the sequence of

Table 1: The length of IGF2BP genes (including 5kb 5′ and 3kb 3′ flank regions) in the 16 genomes.

Genome name	IGF2BP1	IGF2BP2	IGF2BP3
Duroc (Sscrofa11.1)	54304 bp	172416 bp	175365 bp
Landrace	54642 bp	172713 bp	175383 bp
Large white	52863 bp	172065 bp	175753 bp
Bamei pig	54230 bp	172629 bp	175629 bp
Berkshire	54405 bp	172341 bp	175451 bp
Cross-breed	54318 bp	172324 bp	175268 bp
Wuzhishan pig	54300 bp	175137 bp	175770 bp
Bama pigs	54666 bp	172415 bp	174967 bp
Pietrain	53011 bp	172082 bp	176124 bp
Meishan pig	56197 bp	172199 bp	175315 bp
Rongchang pig	54275 bp	172303 bp	175704 bp
Hampshire	54453 bp	172709 bp	175424 bp
Jinhua pig	54345 bp	172455 bp	174787 bp
Tibetan pig	58654 bp	175620 bp	176529 bp
Gottingen pig	54412 bp	169085 bp	55445 bp
Göttingen mini pig	53166 bp	170088 bp	173952 bp

Table 2: Information of the three RIPs in IGF2BP3 gene.

RIPs name	Mutation type	Chromosome	Begin	End	Retrotransposon type	Orientation relative to gene	Gene structure
IGF2BP3-RIP3	Insertion	Chr9	92302063	92302064	SINEA1	Antisense	Intron10
IGF2BP3-RIP5	Insertion	Chr9	92373841	92373842	SINEA1	Sense	Intron3
IGF2BP3-RIP6	Deletion	Chr9	92356617	92357316	SINEA1	Sense	Intron3

Table 3: Genotype statistics of three RIPs in 6 pig populations.

RIPs name	Breeds name	Number of individuals	Genotype frequency		Allele frequency			Hardy Weinberg Balance Test/P	PIC
			+/+	+/-	-/-	+	-		
IGF2BP3-RIP3	Large White	32	0.00	0.00	1.00	0.00	1.00	NA	0.000
	Fengjing	32	0.31	0.38	0.31	0.50	0.50	0.157	0.578
	Sushan	32	0.00	0.09	0.91	0.04	0.96	0.781	0.085
	Erhualian	32	0.22	0.56	0.22	0.50	0.50	0.480	0.375
	Sujiang	32	0.00	0.22	0.78	0.11	0.89	0.487	0.176
	Landrace	32	0.03	0.22	0.75	0.14	0.86	0.591	0.147
IGF2BP3-RIP5	Large White	32	0.00	0.00	1.00	0.00	1.00	NA	0.000
	Fengjing	32	0.31	0.38	0.31	0.50	0.50	0.157	0.375
	Sushan	32	0.00	0.06	0.94	0.03	0.97	0.855	0.059
	Erhualian	32	0.03	0.31	0.66	0.18	0.82	0.885	0.258
	Sujiang	32	0.00	0.00	1.00	0.00	1.00	NA	0.000
	Landrace	32	0.00	0.00	1.00	0.00	1.00	NA	0.000
IGF2BP3-RIP6	Large White	32	0.00	0.00	1.00	0.00	1.00	NA	0.000
	Fengjing	32	0.41	0.34	0.25	0.58	0.42	0.095	0.379
	Sushan	32	0.94	0.06	0.00	0.97	0.03	0.855	0.059
	Erhualian	32	0.66	0.31	0.03	0.82	0.18	0.885	0.258
	Sujiang	32	0.78	0.19	0.03	0.88	0.12	0.419	0.195
	Landrace	32	0.00	0.00	1.00	0.00	1.00	NA	0.000

Table S3: Detail information of all predicted RIPs in the IGF2BP genes and its flanking regions.

Gene name	Predicted RIPs	Breeds	Status	Chromosome	Start	End
IGF2BP1	IGF2BP1-RIP1	Tibetan	Insertion	Chr 12	25219923	25219924
	IGF2BP1-RIP2	Bamei	Deletion	Chr 12	25208622	25208860
	IGF2BP1-RIP3	Bamei	Deletion	Chr 12	25227835	25228061
	IGF2BP1-RIP4	Bamei	Deletion	Chr 12	25241824	25241932
IGF2BP2	IGF2BP2-RIP1	Goettingen	Deletion	Chr 13	123668625	123668687
	IGF2BP2-RIP2	Meishan Bamei	Deletion	Chr 13	123710046	123710267
	IGF2BP2-RIP3	Jinhua	Insertion	Chr 13	123710528	123710529
	IGF2BP2-RIP4	Pietrain	Deletion	Chr 13	123735943	123737353
	IGF2BP2-RIP5	Hampshire berkshire	Deletion	Chr 13	123753092	123753326
	IGF2BP2-RIP6	Landrace	Insertion	Chr 13	123756298	123756299
	IGF2BP2-RIP7	Meishan	Deletion	Chr 13	123788981	123789072
IGF2BP3	IGF2BP3-RIP1	Wuzhishan	Insertion	Chr 9	92125778	92125779
	IGF2BP3-RIP2	Goettingen	Insertion	Chr 9	92130886	92130887
	IGF2BP3-RIP3	Wuzhishan	Insertion	Chr 9	92134699	92134700
	IGF2BP3-RIP5	Rongchang	Insertion	Chr 9	92206477	92206478
	IGF2BP3-RIP6	Bama	Deletion	Chr 9	92189253	92189952
	IGF2BP3-RIP7	Wuzhishan	Deletion	Chr 9	92222906	92223708

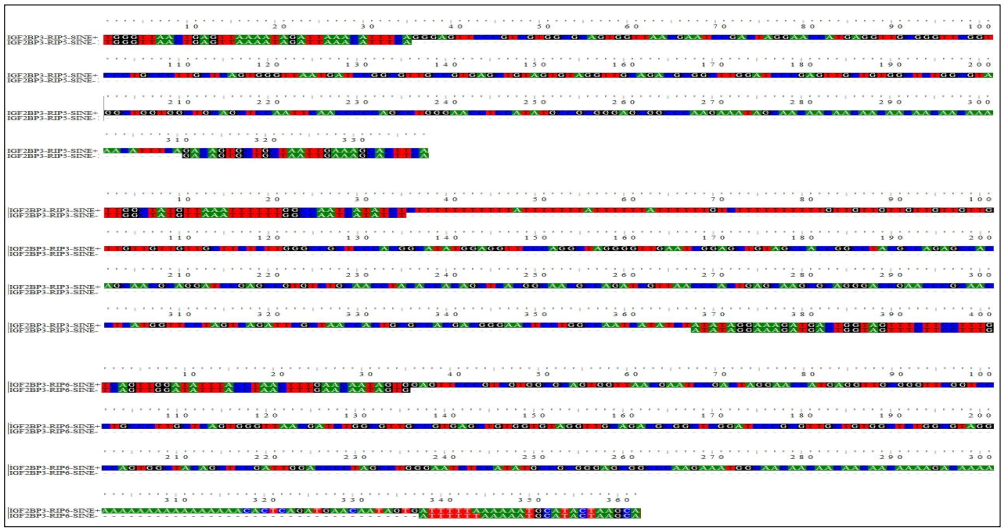

Fig S1: The alignment results for sequences of RIPs obtained by cloning and sequencing.

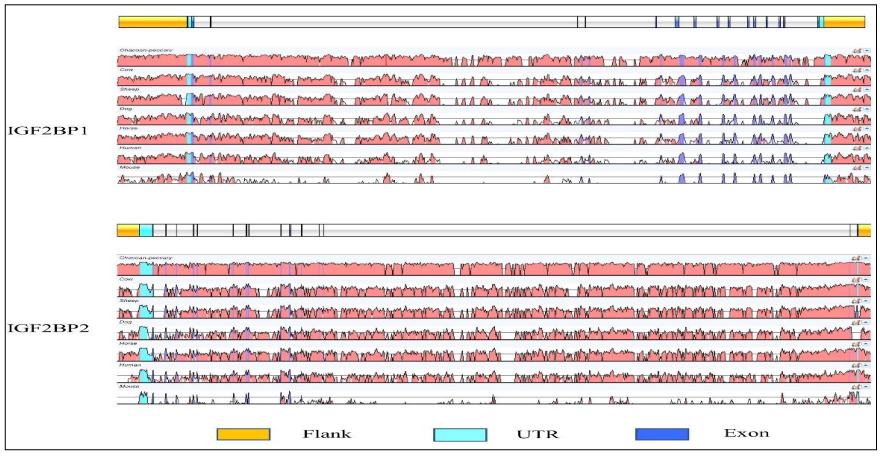

Fig S2: Schematic diagram of gene structure, RIP distribution and conservative analysis results of IGF2BP1 and IGF2BP2 genes. Yellow: flank region, Light blue: UTR, Dark blue: CDS.

the variation is indeed a retrotransposon which further proves that the three SVs were caused by retrotransposons. Further analysis of the retrotransposon of these RIPs, it is found that these three RIPs are all mediated by the SINEA1 retrotransposon, which was the youngest SINE in the pig genome (Chen *et al*. 2019) and distributed in the intron of the gene and the retrotransposon in IGF2BP3-RIP3 shows opposite direction as the gene (Fig 3) and in IGF2BP3-RIP5, IGF2BP3-RIP6 show same direction to the gene (Table 2). The RIPs are all distributed in relatively weakly conserved regions. The potential transcripts of the IG2BP3 gene sequence were predicted by Genescan and the results showed that there is a transcript in the complementary strand of the IGF2BP3 gene that is affected by whether there is a SINE in the IGF2BP3-RIP3 site. When there is a SINE presented, the third exon will be fused with a 119bp SINE sequence, resulting in the length of the third exon to be extended by 99 bp Fig S3. To further evaluate the distribution of these three RIPs in different breeds. Two commercial pig breeds (Large_White, Landrace), two native Chinese breeds (Fengjing, Erhualian) and two crossbreeds in Jiangsu province (Sujiang, Sushan) were chosen and PCR experiments were performed to detect the distribution of three RIPs in 32 individuals of each breed. The results are shown in Table 3 and Fig S4. In general, the polymorphisms of these three RIPs in commercial breeds are poor and they have been purified. Only IGF2BP3-RIP3 has a polymorphism in Landrace. The polymorphism informative contents in commercial pig breeds are lower than that in China native pig breeds, indicating that the genetic diversity of China native pigs is relatively rich, which is consistent with the analysis of SNPs and SSRs (Conson *et al.* 2018) and supports that molecular markers are reliable and effective for population genetic analysis.

Table S4: Primer information of all predicted RIPs.

Primer Name	Primer sequence	Coordinates	Annealing temperature (°C)
IGF2BP1-RIP1-F	TTTTGTCTTTTCTAGGGTCGC	chr12:25219623-25219969	60
IGF2BP1-RIP1-R	CTGAGAACCTCCATATGCCACA		
IGF2BP1-RIP2-F	TTAACGAATCCGACTAGGAAC	chr12:25208575-25208974	61
IGF2BP1-RIP2-R	CCATGAGCTGTGGTGT		
IGF2BP1-RIP3-F	GCCACCACCTCCTTTACCCTG	chr12:25227729-25228317	62
IGF2BP1-RIP3-R	AGCAAACTAGTCCTCACCAAAGCG		
IGF2BP1-RIP4-F	TACTACACAGCAGGGTGATCCA	chr12:25241734-25242045	58
IGF2BP1-RIP4-R	CAGTGGTTAACGAATCCGACT		
IGF2BP2-RIP1-F	TAGTTCCATCCATGTTGCTG	chr13:123626338-12362519	56
IGF2BP2-RIP1-R	AAAAGACGTGAACCCGTA		
IGF2BP2-RIP2-F	CGCAGTCCCACTCTTCGT	chr13:123584707-123585129	60
IGF2BP2-RIP2-R	TGATAAGCCCAAGCTCCGAT		
IGF2BP2-RIP3-F	GATAACTTGATCCCCTTGCT	chr13:123584477-123584603	58
IGF2BP2-RIP3-R	CTGGCTTCCTCACCCAT		
IGF2BP2-RIP4-F	AAAACTCATATGCTAGAAGGGGTA	chr13:123557687-123559264	54
IGF2BP2-RIP4-R	ATTTCATTCTTTTCCATGTGGC		
IGF2BP2-RIP5-F	CAGTGGTTAATGAATCCGACT	chr13:123541669-123542172	58
IGF2BP2-RIP5-R	AGCTCTGTAAAGCCTATGTGT		
IGF2BP2-RIP6-F	ATTTGGATCAAGGGACCAAC	chr13:123538696-123538820	58
IGF2BP2-RIP6-R	AGGCAATATTGACCTCAACGA		
IGF2BP2-RIP7-F	GATCCAGCGTTGCCAT	chr13:123505913-123506170	59
IGF2BP2-RIP7-R	AGAGCCATATATAAATGAGTTGACA		
IGF2BP3-RIP1-F	TGGCACATCCTTATCCAAC	chr9:92251246-92251498	57
IGF2BP3-RIP1-R	TTTTGACTGCTTGATAGGCAC		
IGF2BP3-RIP2-F	GCTTATCCACTCCAAAGGCAAT	chr9:92246352-92246750	58
IGF2BP3-RIP2-R	AACAAATCCAGCTAGGAACCA		
IGF2BP3-RIP3-F	GTTTAGAGCCTTTTATCCCTT	chr9:92242286-92242438	54
IGF2BP3-RIP3-R	ATGACTAATATCTCCAAGCCTA		
IGF2BP3-RIP5-F	CCCGTGTCATAATTTACCCAT	chr9:92170219-92170672	56
IGF2BP3-RIP5-R	CAACTGCAACCTATACTCCA		
IGF2BP3-RIP6-F	TTTCCCACTAATATATGAGACTG	chr9:921787258-92187754	54
IGF2BP3-RIP6-R	ATGAACAAAACTGATAGCCAA		
IGF2BP3-RIP7-F	CCACCAGCCTACGCCAGA	chr9:92153483-92153594	56
IGF2BP3-RIP7-R	GGTCCCTGCCCTTGCTCA		

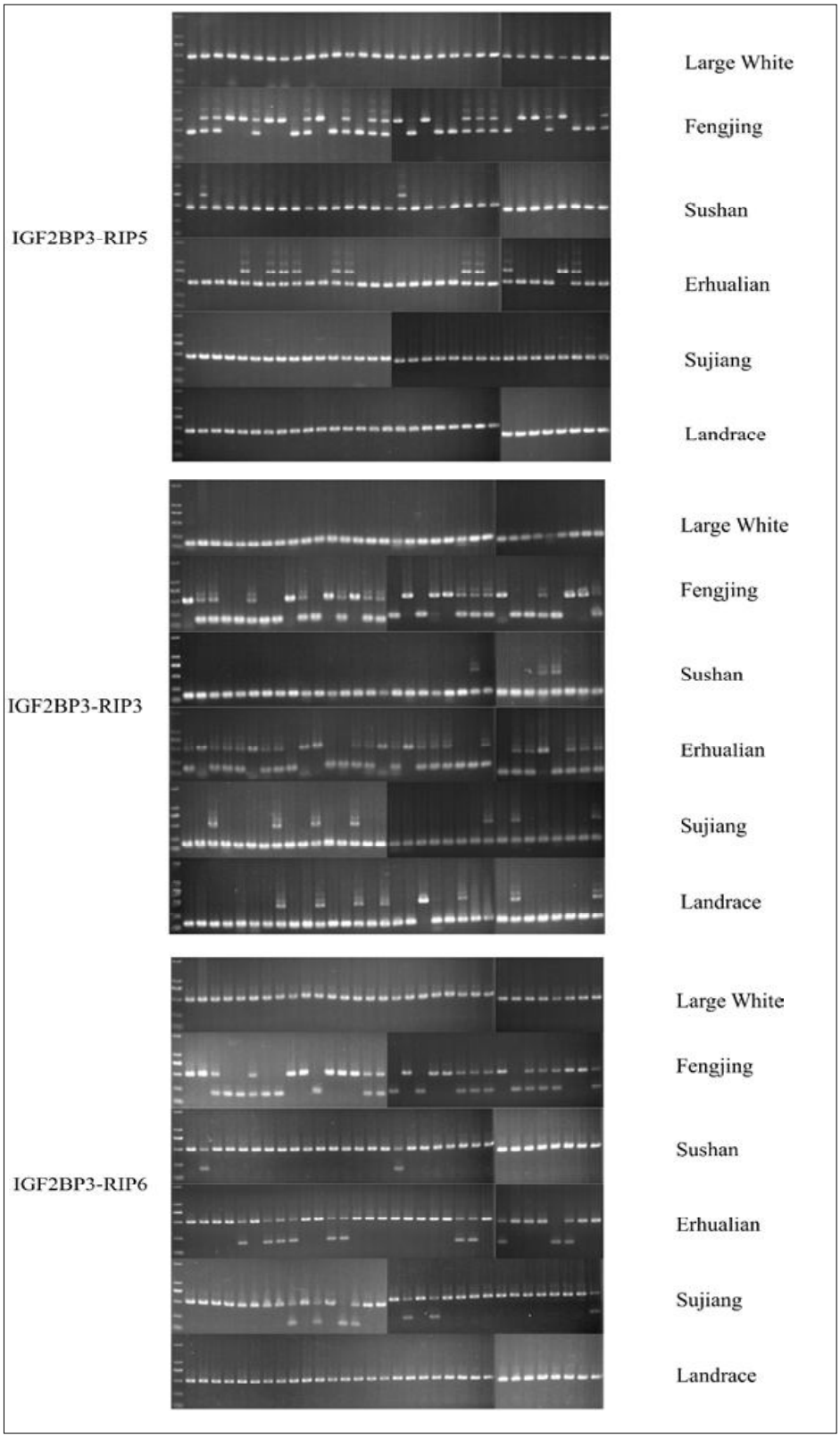

Fig S3: PCR identification and characteristic of RIPs in IGF2BP3. A: RIPs identification in 6 pig breeds by PCR; B: location of retrotransposon insertion in IGF2BP3; C: characteristic of RIP in IGF2BP3.

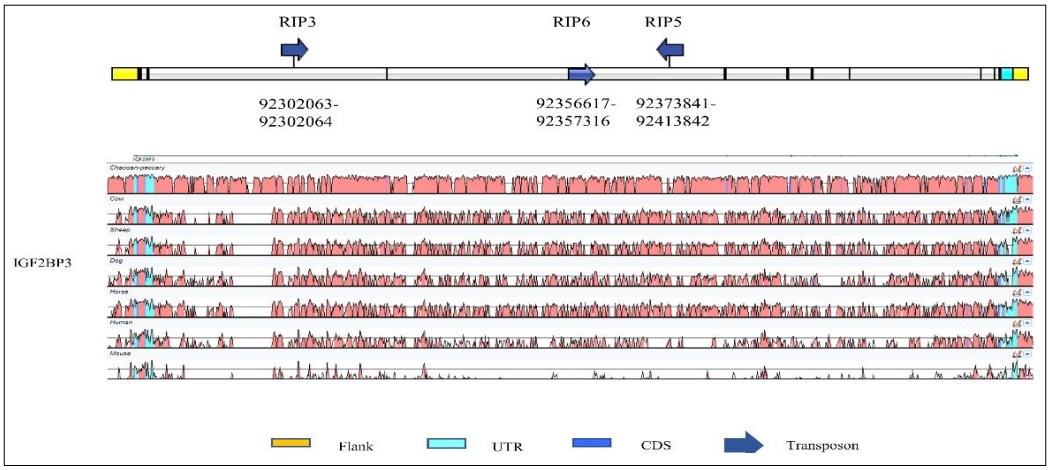

Fig 3: Schematic diagram of gene structure, RIP distribution and conservative analysis results of IGF2BP3. Yellow flank region, Light blue UTR, Dark blue CDS, Arrow RIPs.

CONCLUSION

In this study, sixteen predicted RIPs in IGF2BP genes were identified and three RIPs were verified by PCR. These three RIPs were all caused by the youngest SINEA1 retrotransposons and located in IGF2BP3 introns. The polymorphism of these RIPs in Chinese native pig breeds is higher than that in commercial pigs which is consistent with the intensive selection of commercial pigs.

ACKNOWLEDGEMENT

This research was supported by the Jiangsu Postdoctoral Science Foundation[2021K221B] to Cai Chen, the National Natural Science Foundation of China [grant numbers 31872977 and 32002146], the Independent Innovation Fund Project of Agricultural Science and Technology in Jiangsu Province [CX (19)2016], the Priority Academic Program Development of Jiangsu Higher Education Institutions and the High-end Talent Support Program of Yangzhou University to Chengyi Song.

Conflict of interest: None.

REFERENCES

Barrett, J.C., Fry, B., Maller, J., Daly, M.J. (2005) Haploview: Analysis and visualization of LD and haplotype maps. Bioinformatics. 21: 263-265. doi:10.1093/bioinformatics/bth457.

Caballero, J., Smit, A.F.A., Hood, L., Glusman, G. (2014) Realistic artificial DNA sequences as negative controls for computational genomics. Nucleic Acids Research. 42: 1-12. doi: 10.1093/nar/gku356.

Chen, C., Wang, W., Wang, X., Shen, D., Wang, S., Wang, Y., Gao, B., Wimmers, K., Mao, J., Li, K., Song, C. (2019). Retrotransposons evolution and impact on lncRNA and protein coding genes in pigs. Mobile DNA. 10: 1-24. doi:10.1186/s13100-019-0161-8.

Cohick, W.S., Clemmons, D.R. (1993). The insulin-like growth factors. Annu. Rev. Physiol. 55: 131-53.

Conson, A.R.O., Taniguti, C.H., Amadeu, R.R. andreotti, I.A.A., De Souza, L.M., Dos Santos, L.H.B., Rosa, J.R.B.F. (2018). High-resolution genetic map and QTL analysis of growth-related traits of hevea brasiliensis cultivated under suboptimal temperature and humidity conditions. Frontiers in Plant Science. 9: 1255. doi:10.3389/fpls.2018.01255.

Cordaux, R., Batzer, M.A. (2009) The impact of retrotransposons on human genome evolution. Nature Reviews Genetics. 10: 691-703. doi:10.1038/nrg2640.

DeChiara, T.M., Efstratiadis, A., Robertsen, E.J. (1990). A growth-deficiency phenotype in heterozygous mice carrying an insulin-like growth factor II gene disrupted by targeting. Nature. 345: 78-80. doi:10.1038/345078a0.

Du, Z., D'Alessandro, E., Zheng, Y., Wang, M., Chen, C., Wang, X., Song, C. (2022). Retrotransposon insertion polymorphisms (RIPs) in pig coat color candidate genes. Animals: An Open Access Journal from MDPI. 12: doi:10.3390/ani12080969.

Hofig, A. Simmen, F.A., Bazer, F.W., Simmen, R.C. (1991). Effects of insulin-like growth factor-I on aromatase cytochrome P450 activity and oestradiol biosynthesis in preimplantation porcine conceptuses *in vitro*. Journal of Endocrinology. 130: 245-250. doi:10.1677/joe.0.1300245.

Jadhav, P.V., Das, D.N., Tarate, S.B. (2020). Genetic polymorphism in TLR2 genes of HF cross bred cattle through PCR SSCP. Indian Journal of Animal Research. 54: 430-433. doi: 10.18805/ijar.B-3796.

Kaushik, P., Naskar, S., Handique, P.J., Rahman, H., Sarma, D.K. (2017). Genetic polymorphism of growth hormone releasing hormone gene in exotic and crossbred pigs. Indian Journal of Animal Research. 51: 231-234. doi: 10.18805/ijar.11162.

Kirby, C.J., Thatcher, W.W., Collier, R.J., Simmen, F.A., Lucy, M.C. (1996). Effects of growth hormone and pregnancy on expression of growth hormone receptor, insulin-like growth factor-I and insulin-like growth factor binding protein-2 and -3 genes in bovine uterus, ovary and oviduct. Biology of reproduction. 55: 996-1002. doi: 10.1095/biolreprod. 55.5.996.

Ko, Y., Lee, C.Y., Ott, T.L., Davis, M.A., Simmen, R.C.M., Bazer, F.W., Simmen, F.A. (1991). Insulin-like growth factors in sheep uterine fluids: Concentrations and relationship to ovine trophoblast protein-1 production during early pregnancy. Biology of Reproduction. 45: 135-142. doi: 10.1095/biolreprod. 45.1.135.

Lennard, S.N., Stewart, F., Allen, W.R. (1995). Insulin-like growth factor II gene expression in the fetus and placenta of the horse during the first half of gestation. Journal of Reproduction and Fertility. doi: 10.1530/jrf.0.1030169.

Li, J., Davis, B.W., Jern, P., Dorshors, B.J., Siegel, P.B. andersson, L. (2019). Characterization of the endogenous retrovirus insertion in CYP19A1 associated with henny feathering in chicken. Mobile DNA. doi: 10.1186/s13100-019-0181-4.

Liu, H., Xu, H., Lan, X., Cao, X., Pan, C. (2021). The InDel variants of sheep IGF2BP1 gene are associated with growth traits. Animal Biotechnology. 0: 1-9. doi:10.1080/10495398.2021.1942029.

Murphy, S.C., Evans, J.M., Tsai, K.L., Clark, L.A. (2018). Length variations within the Merle retrotransposon of canine PMEL: Correlating genotype with phenotype. Mobile DNA. 9: 1-11. doi:10.1186/s13100-018-0131-6.

Nonoshita, L.D., Wathen, N.C., Dsupin, B.A., Chard, T., Giudice, L.C. (1994). Insulin-like growth factors (IGFs), IGF-binding proteins (IGFBPs) and proteolyzed IGFBP-3 in embryonic cavities in early human pregnancy: their potential relevance to maternal-embryonic and fetal interactions. The Journal of Clinical Endocrinology and Metabolism. 79: 1249-1255. doi:10.1210/jcem. 79.5.7525630.

Reynolds, T.S., Stevenson, K.R., Wathes, D.C. (1997). Pregnancy-specific alterations in the expression of the insulin-like growth factor system during early placental development in the Ewe 1. endocrinology. 138: 886-897. doi:10.1210/endo.138.3.4983.

Wang, Z., Zhang, X., Jiang, E., Yan, H., Zhu, H., Chen, H., Liu, J., Qu, L., Pan, C., Lan, X. (2020). InDels within caprine IGF2BP1 intron 2 and the 3′-untranslated regions are associated with goat growth traits. Animal Genetics. 51: 117-121. doi:10.1111/age.12871.

Xing, J., Zhang, Y., Han, K., Salem, A.H., Sen, S.K., Huff, C.D., Zhou, Q., Kirkness, E.F., Levy, S., Batzer, M.A., Jorde, L.B. (2009). Mobile elements create structural variation: Analysis of a complete human genome. Genome Research. 19: 1516-1526. doi:10.1101/gr.091827.109.

Yang, G.L., Shi, C.X., Fu, D.L., Li, Z.Q. (2019). The mutations within MC1R, TYRP1, ASIP genes and their effects on phenotypes of coat color in wild pigs (Sus scrofa ussuricus). Indian Journal of Animal Research. 53: 288-293. doi:10.18805/ijar.B-804.

Yeh, F.C. Yang, R.C. Boyle, T. (1999). POPGENE version 1.32: Microsoft Windows-based freeware for population genetic analysis, quick user guide. Cent Int For Res Univ Alberta, Edmonton, Alberta, Canada. 1-29.

Zhao, P., Li, J., Kang, H., Wang, H., Fan, Z., Yin, Z., Wang, J., Zhang, Q., Wang, Z., Liu, J.F. (2016). Structural variant detection by large-scale sequencing reveals new evolutionary evidence on breed divergence between chinese and european pigs. Scientific Reports. 6: 18501. doi: 10.1038/srep18501.

Zheng, Y., Chen, C., Chen, W., Wang, X. Y., Wang, W., Gao, B., Wimmers, K., Mao. J D., Song, C. y. (2020). Two new SINE insertion polymorphisms in pig Vertnin (VRTN) gene revealed by comparative genomic alignment. Journal of Integrative Agriculture. 19: 2514-2522. doi: 10.1016/S2095-3119(20)63255-5.

Zhou, J., Dsupin, B.A., Giudice, L.C., Bondy, C.A. (1994). Insulin-like growth factor system gene expression in human endometrium during the menstrual cycle. Journal of Clinical Endocrinology and Metabolism. doi: 10.1210/jcem.79.6.7527408.

Zhou, Z., Li, M., Cheng, H., Fan, W., Yuan, Z., Gao, Q., Xu, Y., *et al.* (2018). Erratum to: An intercross population study reveals genes associated with body size and plumage color in ducks (Nature Communications. 9, 1, (2648), 10.1038/s41467-018-04868-4). Nature Communications. 9:712100. doi:10.1038/s41467-018-06521-6.