Cloning and sequencing of the complicated rDNA gene family of *Bos taurus*

D. Tang[1,2], H. Jiang[2], Y. Zhang[1], Y. Li[2], X. Zhang[2], T. Zhou[2]

[1]College of Life Science, Foshan University, Foshan, Guangdong, P.R. China
[2]College of Life Science and Technology, Jinan University, Guangzhou, Guangdong, P.R. China

ABSTRACT: The rDNA genes coding for ribosomal RNA (rRNA) in animals are repeat sequences with high GC content and complicated structure. Based on the sequences of human ribosomal DNA repeat unit and transcription unit and the long and accurate PCR method with *LA Taq* DNA polymerase and GC buffer, we were able to amplify the complicated repeat sequences of rDNA genes in *Bos taurus*. Three rDNA genes and 2 internal transcribed spacer (ITS) fragments were cloned and confirmed by sequencing. The conditions for the cloning of complicated DNA sequences such as special rules of primer design, improvement of the reaction system, selection of DNA polymerase and adjustment of cycle parameters were discussed.

Keywords: GC-rich; repetitive sequences; PCR; rDNA; rRNA; *Bos taurus*

The 18S, 5.8S, 28S rDNA genes and internal transcribed spacers (ITS) between rDNA genes in animals are linked as a transcription unit (Figure 1). In order to develop the technique of multiple locus gene targeting at repeat sequences of *Bos taurus in vivo* (Tang et al., 2002), i.e. to insert foreign genes into the ITS between the 3 rDNA genes, the rDNA gene family of *Bos taurus* needs to be cloned and the targeted locus needs to be found by sequencing.

The rDNA gene families in animals are repeat sequences with high GC content and complicated structure. The repeats are often a stretch of G or C. Therefore the cloning of the rDNA gene family is very difficult. Currently few rDNA gene families of animals were cloned. The objective of the present study is to clone and sequence the complicated repeat sequences of the rDNA gene family of *Bos taurus* as insertion sites for multiple locus gene targeting.

MATERIAL AND METHODS

Designing primers

Based on the sequences of the human ribosomal DNA repeat unit and transcription unit (GenBank accession No. U13369) (Gonzalez et al., 1995), 16 pairs of primers were designed for PCR amplification of the rDNA gene family of *Bos taurus*. Only 5 pairs of primers could effectively amplify the genes and ITS fragments, and these are presented in Table 1.

Isolation of genomic DNA and PCR amplification

Genomic DNA was extracted and purified from the sperm cells of *Bos taurus* following the SDS method. Because the rDNA gene family contains

Supported by the National Natural Science Foundation of China (Project No. 30470978 30671194) and Nature Science Foundation of Guangdong Province (Project No. 04011645).

Table 1. The primers used to amplify the rDNA genes in *Bos taurus*

PCR product	Primer name	Primer sequence 5' → 3'	Primer orientation	Positions in human sequence	Annealing temperature (°C)
BT498 bp	P1	TACCTGGTTGATCCTGCCAGTAGCATATGC	F	18S rDNA 1–30	60
	P2	GGGAGTGGGTAATTTGCGCGCCTGCTGCCT	R	18S rDNA 469–498	
BT2538 bp	P3	GCTACCACATCCAAGGAAGGCAGCAGGC	F	18S rDNA 452–479	62
	P4	TTCTTCATCGACGCACGAGCCGAGTGAT	R	5.8S rDNA 17–44	
BT1479 bp	P5	CGGTGGATCACTCGGCTCGTGCGTCGAT	F	5.8S rDNA 11–38	61
	P6	GGCTGCATTCCCAAGCAACCCGACTCCG	R	28S rDNA 267–294	
BT2319 bp	P7	CCTGGCTTGCTTTGTGGCGTGTATG	F	Based on sequence of BT1479	60
	P8	GTTGATTCGGCAGGTGAGTTGTTAC	R	28S rDNA 1994–2018	
BT3326 bp	P9	AAGCCGCTGTGGAGCAATGAAGGTGAAG	F	Based on sequence of BT2319	61
	P10	GGACAAACCCTTGTGTCGAGGGCTGACT	R	28S rDNA 5009–5035	

high GC content and repeat sequences, high-fidelity DNA polymerases such as *Pfu* (Promega, USA) and *Ex Taq* (TAKARA, Japan) were used to amplify the rDNA gene family of *Bos taurus*. However, specific PCR products could not be obtained. Finally the Long and Accurate PCR *Taq* polymerase (*LA PCR Taq*) (TAKARA, Japan) with the GC buffer (TAKARA, Japan) was chosen to amplify the rDNA gene family. PCR amplifications were performed in a Biometra thermocycler. Reactions were carried out in 50 µl volumes containing 10 pmol of each primer, 200 µM of each dNTP, 25 µl 2 × GC buffer I or buffer II, 0.5 u of *LA PCR Taq* polymerase, and ~100 ng of genomic DNA. The primers P1/P2, P3/P4, P5/P6, P7/P8 and P9/P10 were annealed at 60°C, 62°C, 61°C, 60°C and 61°C, respectively, and 30 cycles were used.

Cloning and sequencing

The PCR products were separated on 0.8% agarose/TAE gel and purified from the gel with the E.Z.N.A.Gel Extraction kit (OMEGABIO-TEK, USA) according to the manufacturer's instructions. The PCR products were cloned into pMD18-T vector (TAKARA, Japan) and transformed into *E. coli* JM109. The plasmids were extracted from the white colonies with the E.Z.N.A.Plasmid Miniprep kit (OMEGABIO-TEK, USA) following the recommended protocol. To screen positive colonies, the plasmids containing the fragment were identified by PCR with the original primers. Positive plasmids with the inserted fragment were sequenced for obtaining the sequence of the rDNA gene family of *Bos taurus*. Sequencing was carried out under contract at Bioasia Company. An ABI 3730 genetic analyser utilising a capillary sequencing technology was used.

RESULTS AND DISCUSSION

PCR amplification

The rDNA gene family of *Bos taurus* was amplified by 5 pairs of primers, P1/P2, P3/P4, P5/P6, P7/P8 and P9/P10 designed according to the sequences of human ribosomal DNA repeat unit and transcription unit. The specific PCR products were

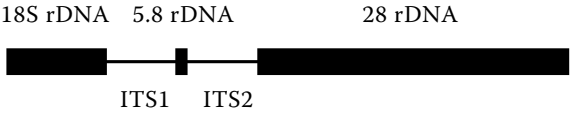

Figure 1. The structure of the rDNA transcription unit in animals

obtained as seen by agarose/TAE gel electrophoresis (Figure 2). The sizes of the fragments were as expected, i.e. about 500 bp, 2 500 bp, 1 500 bp, 2 500 bp and 3 300 bp, respectively.

Cloning and sequencing

Colonies in which the rDNA genes could be amplified using the original primers were considered positive. The plasmid DNA was isolated and sequenced.

The sizes of the fragments of 3 rDNA genes and 2 ITS, BT498, BT2538, BT1479, BT2319 and BT3326, were 498 bp, 2 538 bp, 1 479 bp, 2 319 bp and 3 326 bp, respectively. From these the complete *Bos taurus* sequence of the rDNA gene family of 8 675 bp was assembled by using DNAMAN software (GenBank accession No. DQ222453).

The sizes of *Bos taurus* 18S rDNA gene, ITS1, 5.8S rDNA gene, ITS2 and 28S rDNA gene are 1 873 bp, 1 072 bp, 150 bp, 1 038 bp and 4 542 bp, respectively (GenBank accession Nos. AY779625, AY779626, AY779627, AY779628 and AY779629). By comparison with the sequence of the human rDNA family, these genes and spacers have 98%, 56%, 91%, 11% and 66% of identity, respectively. The average GC content of the complete sequence

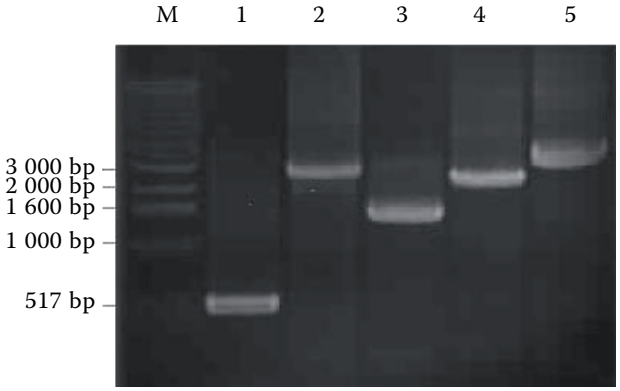

Figure 2. Ethidium bromide stained agarose gel showing PCR fragments BT498 bp, BT2538 bp, BT1479 bp, BT2319 bp and BT3326 bp (lanes 1–5) amplified with the primers presented in Table 1. M – 1 kb ladder

is as high as 61.6%, and the GC contents of the genes and spacers are 56.1%, 80%, 50.7%, 67.5% and 58.6%, respectively.

Strategies for the amplifying and cloning of complicated sequences with high GC content

The rDNA gene family of *Bos taurus* consists of DNA sequences with high GC content. The average GC content is 61.6%, and the GC content of ITS1 is as high as 80%. The rDNA gene family of *Bos taurus* is composed of repeat sequences containing mostly stretches of G or C, with complicated structure. For instance, there is sequence with repeats of stretches of 8 Gs. The following strategies were used to amplify the complicated repeat sequences.

Based on the sequences of human ribosomal DNA repeat unit and transcription unit, 16 pairs of primers were designed for amplifying the rDNA gene family of *Bos taurus*. However, only 5 pairs of primers could amplify the specific products from the genomic DNA of *Bos taurus*. The studies show that for the design of primers for complicated repeat sequences special rules should be observed besides abiding by the common principles. (1) The binding sites of the primers should be far from the sequences with continuous G or C repeats. (2) The annealing temperature of the primers should be higher than 60°C. (3) The GC content of primers should be about 50–70%. The GC content in the middle of the primers can be relatively higher. The stretches of G, C, A or T should be avoided. (4) The longer the primers, the better the efficiency of amplification under the conditions of assuring the specificity and validity of PCR amplification.

Genomic DNA as long as 42 kb could be amplified successfully in the middle of the 90's. Based on this technique, a Long and Accurate PCR (LA PCR) was developed to amplify the complicated repeat sequences of rDNA genes. Compared to the regular PCR, the LA PCR has some special requirements in the reaction system, DNA polymerase and cycle parameters, etc. (Chen and Wang, 1996).

Some strategies for improving regular PCR were also useful in LA PCR. (1) DNA sequences containing GC-rich regions have a high denaturing temperature and sometimes cannot be amplified successfully. It is helpful to add 1–10% dimethyl sulphoxide (DMSO) to increase the efficiency of amplification of GC-rich regions. Its mechanism

maybe that DMSO can lead to instability of DNA structure by dehydration of DNA, lowering the melting point and denaturing temperature, and thus it increases the efficiency of amplification (Li et al., 2002). (2) 8% glycerol can markedly increase the production of the PCR product of large DNA fragments. Using both glycerol and DMSO is better than using only glycerol (He and Li, 2001). (3) The LA PCR needs a higher Mg^{2+} concentration. The Mg^{2+} concentration in regular PCR is 2.5 mM, but in LA PCR it is 5 mM. (4) A low concentration of KCl is useful for the annealing of primers and can promote the extension of DNA strand. (5) Another marked characteristic of the LA PCR reaction system is a requirement for high pH. In this system pH 8.8–9.2 was used. The high pH is useful to maintain the stability of DNA templates. Increased pH leads to higher amplification efficiency maybe due to decreased de-purination of the DNA template. GC buffer I and GC buffer II from some companies, such as TAKARA, were confirmed to be suitable for the amplification of complicated repeat sequences with high GC content.

The high fidelity DNA polymerase should be used for the amplification of complicated repeat sequences with high GC content. The *LA PCR Taq* polymerase used in the studies has the activity of 3' → 5' exonuclease according to the principle of LA PCR. When there are complicated secondary structures such as GC-rich region and repeat sequences in the DNA template, the *LA PCR Taq* polymerase with GC buffer I or GC buffer II is very effective in PCR amplification. The efficiency and fidelity of amplification of long DNA fragments, especially fragments longer than 10 kb, using *LA PCR Taq* polymerase is much higher than in other polymerases.

Studies of different denaturing time and temperature showed that the size and output of the amplified fragments were negatively correlated with denaturation time. If denaturation time and temperature are kept in the lowest range, the size and output of the amplified fragments were positively correlated with the extension time. Based on the above study, the cycle parameters were set as follows: denaturing at 94°C for 30 s, annealing at 60–68°C for 30 s and extension at 72°C for 30 s – 3 min.

When sequencing complicated repeat sequences with high GC content, sometimes only a short sequence can be obtained, or even they cannot be sequenced. So, the regular sequencing methods cannot be used for these sequences. The sequencing methods should be improved as follows: (1) Sequencing primers should be far from the sequences with stretches of G or C repeats. (2) Cycle parameters, such as the annealing temperature should be increased. (3) The reaction system should be changed such as using the GC buffer, adding reagents to promote DNA denaturation, etc.

Acknowledgements

We thank Yan Xia and Li Fang for their helpful discussion.

REFERENCES

Chen S.W., Wang X.Z. (1996): The PCR methods of amplifying the large fragments of target DNA. Biotechnology, *6*, 1–2.

Gonzalez I.L., Sylvester J.E. (1995): Complete sequence of the 43-kb human ribosomal DNA repeat: analysis of the intergenic spacer. Genomics, *27*, 320–328.

He Z.Y., Li Y.Q. (2001): The role of dimethyl sulfoxide on polymerase chain reaction for human apoE gene. J. Jinan Univ. (Nat. Sci.), *22*, 104–107.

Li A.L., Jiang T., Ma Z.Y., Jia J.Z. (2002): Several methodologies on improving the PCR products. Biotechnol. Bull., *6*, 33–35.

Tang D.S., Shi J.Q., Tian Y.B., Gu W.J. (2002): Multiple site gene targeting in somatic cells and desired site integration of genes. J. Foshan Univ. (Nat. Sci. Ed.), *20*, 64–68.

Received: 2005–04–13
Accepted after corrections: 2006–05–31

Corresponding Author

Dongsheng Tang, College of Life Science and Technology, Jinan University, Guangzhou, 510632 Guangdong, P.R. China
E-mail: tangdsh@163.com