Detection of Single Nucleotide Polymorphisms in Cattle Genes Using Automated DNA Sequencing and Dideoxy Fingerprinting

Kittipat Ukoskit
Department of Biotechnology, Faculty of Science and Technology,
Thammasat University (Rangsit Campus), Thailand

Abstract

Single nucleotide polymorphisms (SNPs) are the most common form of DNA sequence polymorphisms and of mutation in vertebrates. Any detection approach must involve the determination of DNA sequence and allele frequencies which are important for linkage analysis studies and high-throughput genotyping methods. In this experiment two SNPs detection methods, direct automated DNA sequencing and dideoxy Fingerprinting (ddF) were used to detect SNPs in Insulin-like Growth Factor Binding Protein-3 gene (IGFBP-3) and Growth Hormone gene (GH) of cattle. Using direct automated DNA sequencing, 2 SNPs representing C/T and G/T and 2 SNPs representing A/C and C/T were identified in the IGFBP-3 gene and the GH gene, respectively. Dideoxy GTP used as the chain terminator in the ddF method generated fingerprints corresponding to the G/T substitution and the C/T substitution in the IGFBP-3 gene and in the GH gene, respectively. The frequencies of the minor allele were estimated by the ratio of peak heights from chromatograms of sequencing traces generated from the pooled DNA. The results indicated that the G/T substitution with 29% of allele frequency for the minor allele was present in the population and informative for high throughput genotyping and for genetic mapping.

1. Introduction

Single nucleotide polymorphisms (SNPs) are the most common form of DNA sequence variation in vertebrates. Two to three polymorphic sites may be found per kilobase in some regions of human genomic DNA (1). Although the density of SNP markers is uneven across the genome, it has been estimated that the average density of SNP loci in human genome map is a 2-3 cM. Taking advantage of their density, mapping of complex traits can be efficiently investigated (2). Because SNPs have only two alleles (biallelics), genotyping known SNPs requires only a plus/minus assay allowing easier automation. Since the informativeness of SNP markers is determined by the frequency of the minor allele, any SNP discovery method must involve not only the determination of DNA sequence but also the determination of allele frequency.

Many investigators reported that dideoxy fingerprinting (ddF) was more sensitive and reliable than single strand conformation polymorphism (SSCP) for detection of point mutation (3,4,5). The ddF technique involves a Sanger sequencing reaction with one dideoxy nucleotide followed by non-denaturing gel electrophoresis. The sequencing product mobility is determined by size and secondary structure of single-strand conformation. In this study two SNP detection techniques, direct automate DNA sequencing and ddF were compared to investigate SNPs in Insulin-like Growth Factor Binding Protein-3 gene (TGFBP-3) and Growth hormone gene (GH) in cattle. The allele frequencies were estimated with chromatograms of sequencing traces generated from DNA samples of individual and from a pooled DNA of a population of fourteen individuals.

2. Materials and Methods

2.1 PCR amplification

Fourteen cattle semen samples were collected randomly and used in this experiment. Genomic DNA of each individual was extracted using the modified method of Heyen et al. (6). The genomic DNA sample from each individual and the pooled DNA consisting of equal amounts of DNA from fourteen individuals were amplified by PCR. Two pairs of PCR primers were used to amplify a 300-bp segment of the GH gene and a 600-bp segment of the IGFBP3 gene. Approximately 20 ng of genomic DNA was amplified in 50-μl PCR reaction consisting of standard PCR buffer (50 mM KCl, 10 mM Tris-HCl pH 8.3, 1.5 mM $MgCL_2$, 0.001% gelatin), 0.2 μM of each primer, 0.2 mM dNTPs and 0.2 unit of *Taq* DNA polymerase (GIBCOBRL). Thermal cycling was performed on PTC-100™ (MJ Research, Inc.) with an initial denaturing step at 95 °C for 5 minutes followed by 40 cycles of denaturation at 95 °C for 30 seconds, primer annealing at 60 °C or 61 °C (for the IGFBP3 gene or the GH gene respectively), for 1 minute, and primer extension at 70 °C for 1 minute. The PCR product was separated on a 1% agarose gel in 0.5 TBE buffer and visualized with ethidium bromide straining. The PCR products were purified according to the manufacturer's instruction using purification kit of Genomed, Astral Scientific. The purified PCR products were used for a sequencing template.

2.2 DNA sequencing

Li-Cor cycle sequencing reactions were carried out using SequiTherm EXCEL™ II DNA Sequencing Kits-LC (EPICENTRE TECHNOLOGIES, Inc.). The 17 μl volume of premix was made in a 0.5 ml microfuge tube containing 7.2 μl of 3.5X sequencing buffer, 2 ρ mole of IRD-labeled primer, 4.8μl (100 to 250 fmole) of purified PCR product, 5 units of DNA polymerase and ddH_2O to 17 μl. For each template, 2 μl of each termination mix, A, C, G, or T were added to the A, C, G, or T tubes respectively followed by the addition of 4 μl of the premix to each of the four tubes of the termination mix. The PCR reactions were overlaid with paraffin oil and performed in a PTC-100™ (MJ Reseàrch, Inc.). The temperature profiles included a 5 minutes denaturation step at 95 °C followed by 30 cycles each consisting of 30 seconds denaturation at 95 °C, 15 seconds annealing at 50 °C, and 1 minute extension at 70 °C. After the PCR, each sample was resuspended in 3 μl of stop/loading buffer (98% formamide, 0.1% bromophenol blue, 0.1% xylene cyanol, 10 mM EDTA), heated for 3 minutes at 94 °C, chilled on ice and loaded onto Li-Cor automated DNA sequencer according to the manufacturer's instructions. The Li-COR used infrared technology (wave length 700 nm and 800nm) to detect DNA fragment. Fluorescently labeled M13 Forward and Reverse were used as primers to sequence in both directions. The cycle sequencing products were resolved in four lanes. The bands were detected by the laser and visualized on the monitor computer.

2.3 Analysis of SNPs

The chromatograms of sequencing traces from the fourteen DNA samples and the pooled DNA sample were analyzed with Sequencher 4.0 (Gene Codes Corporation, Ann Arbor, MI). SNPs were identified by comparing base-by-base and searching for sequence variants by the Sequencher program. The analyzed DNA sequencing traces were visually inspected to identify changes in the peak height. The allele frequencies were estimated by comparing the ratio of the relative peak heights of the corresponding bases between the sequence of the pooled DNA sample and the sequence of a heterozygous individual. The estimated allele frequencies were determined by the formula described by Knok et al. (8).

2.4 Dideoxy fingerprinting

Dideoxy GTP was used as the chain terminator in the termination reaction. The Sanger termination reaction cycling parameters were the same as that of cycle sequencing reaction previously described. After the PCR each reaction was added with stop/loading buffer. The mix was heated for 3 minutes, and chilled on ice until loading. Two μl of each reaction were loaded on 8% non-denaturing polyacrylamide gel with glycerol. Electropheresis was performed in Li-Cor automated DNA sequencer at room temperature at 2 W constant for 20 hours. The bands were detected by the laser and visualized on the monitor computer.

3. Results and Discussion

3.1 Determination of SNPs

Sequencing traces of fourteen individuals and the pooled sample were compared base by base to scan for DNA variations in the PCR products of the IGFBP3 gene and the GH gene. Peak height variations were visually analyzed. The characteristic of *Taq* polymerase in cycle sequencing can create a problem for the detection of SNPs because the dissimilar rate of incorporation of the dideoxynucleotides can cause the peak intensity representing one allele to be dramatically greater than that of the other allele. The base used to determine heterozygote may be missed by exiting base-calling software (9). In this study three criterions namely, a greatly decreased signal intensity, the presence of a concomitant appearance of a new base peak underneath, and a significant change in peak height in the immediately 3′ base were utilized for identifying heterozygous base in a sequencing trace.

Figure 1 illustrats the example of the chromatograms of the window CANGG sequencing trace of the two homozygotes carrying the allele CACGG and CATGG (underlined bases indicated a polymorphic site) and of the heterozygote with the allele CAC/TGG of the IGFBP-3 gene. The heterozygous signal at the polymorphic site (sequence 3 in figure 1)showed the decrease in peak height of approximately 50% of the C peak (blue line) with the concomitant appearance of the similar peak height of the T peak (red line) underneath when compared to its homozygous counterpart. Because the heterozygous peaks represented sequences from two simultaneously amplified DNA fragments from a duplicated regions of the genome. Furthermore, two alleles exerting different influences on the height of the G peak immediately 3′ to the polymorphic site (sequence 1, 2 and 3 in figure 1) helped to recognize heterozygotes with confidence. When the sequencing traces of the two allelic homozygotes were compared to that of the heterozygote, a large G peak of sequence 1 (larger than the allelic C peak) following the C in the CACGG homozygote, a small G peak of sequence 2 (smaller than the allelic T peak) following the T in the CATGG homozygote, and a medium size G peak following the composite C/T peaks in the hetrozygote were observed. Based on these criterions, 2 SNPs representing C/T and G/T and 2 SNPs representing A/C and C/T were identified in the IGFBP-3 gene and the GH gene, respectively.

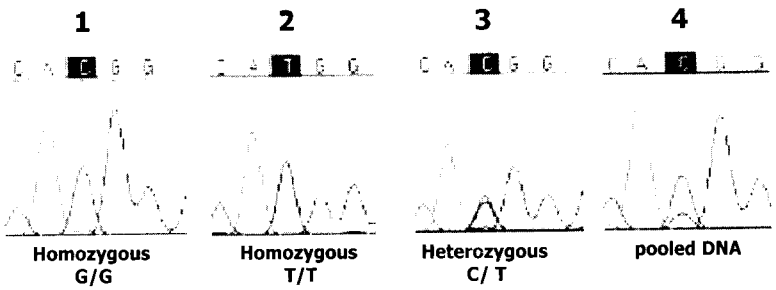

Figure 1 Examples of chromatograms of sequencing traces showing the polymorphic site (a base in the solid box). The C/T substitution was identified in the IGFBP-3 gene.

Table 1 Identification of SNPs in the PCR products of both the IGFBP-3 gene and the GH gene of fourteen bulls. The predicted allele frequencies computed by using the sequencing trace of the pooled DNA sample were compared to the observed allele frequencies determined by genotyping fourteen individuals using the sequencing trace of each individual.

PCR product of	Genotype	Number of individual	Observed frequency	Predicted frequency
IGFBP-3 gene	C/C	9	C = 0.71	C = 0.69
	C/T	2		
	T/T	3	T = 0.29	T = 0.31
	G/G	11	G = 0.89	Too low to
	G/T	3		
	T/T	0	T = 0.11	estimate
GH gene	C/C	13	C = 0.96	Too low to
	C/T	1		
	T/T	0	T = 0.04	estimate
	A/A	11	A = 0.89	Too low to
	A/C	3		
	C/C	0	C = 0.11	estimate

Table 1 showes the results of scanning SNPs in the IGFBP-3 gene and the GH gene of fourteen bulls. The allele frequencies were estimated using observed and predicted approaches. The observed allele frequencies were determined by the number of individual which was genotyped by comparing the sequencing trace of each individual. Whereas the predicted allele frequencies were computed by using the formula $p = 0.5(A_{pool}/A_{het})$, $q = 1 - p$ where A_{pool} and A_{het} were relative peak heights of the major allele (8). The relative peak heights of A_{pool} and A_{het} were measured from the sequencing traces of the pooled DNA and the heterozygote, respectively. A comparison of the observed allele frequencies with the predicted allele frequencies of the pooled DNA was presented in Table 1. In only the case of the C/T substitution of the IGFBP-3 gene possessing 29% of the observed minor allele frequency, the predicted allele frequencies could be estimated from the pooled DNA. Therefore, the pooled DNA approach can be utilized to reduce the number of sequencing reactions required to identify SNPs possessing high allele frequency in the population from anonymous STSs. The reagent and labor cost of SNPs development can be lowered significantly.

However, figure 2 illustrates that the A/C SNP with < 20% allele frequency for the minor allele (the blue peak representing the C allele of sequence 3 in figure 2) were unable to be identified quantitatively with the sequencing traces of the pooled DNA sample. Because the presence of signal intensity of the minor allele in the sequencing trace of the pooled DNA sample was too low (see Table 1), the predicted allele frequencies can not be estimated accurately. It has been previously reported that the pooled DNA sequencing approach was a useful approach to identify SNPs with the minor of > 15% frequency (11). It was unlikely that a SNP with the minor allele frequency of < 15% could be identified with confidence using the

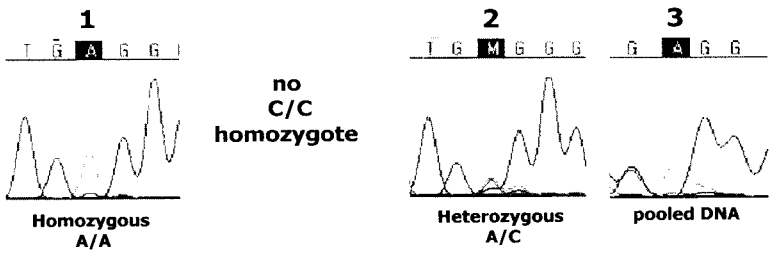

Figure 2 The chromatograms of sequencing traces representing the A/C substitution identified in the GH gene. The presence of signal intensity of the C peak (the blue peak) in the sequencing trace of the pooled DNA sample (sequence 3) was too low (see Table 1 for the observed allele frequency). The predicted allele frequencies can not be estimated using the sequencing trace of the pooled DNA sample.

sequencing trace of the pooled DNA sample. Consequently, SNPs discovered by use of the population pool method are informative markers not rare polymorphisms which are useful for individual genotyping and genetic mapping.

3.2 Analysis of ddF

ddF is a hybrid technique that combines aspects of single-strand conformation polymorphism (SSCP) and didexy sequencing and can detect the presence of single base and sequence variant in PCR amplification products. In this study, dideoxy GTP was chosen for the termination reaction. The informative dideoxy component is abnormal when an extra band is produced by a sequence change that creates an extra G (substitutions of A→G, C→G and T→G) or when a segment is eliminated by a change of G to another. The SSCP component is informative if all the bands subsequent to the sequence change show altered migration when compared to the wild type individuals.

Figure 3 illustrated the ddF patterns of the 600-bp PCR product of the IGFBP-3 gene (figure 3A) and the 300-bp PCR product of the GH gene (figure 3B). In figure 3A, the extra bands of the individual 4, 10 and 12 indicated by arrows were associated with the G/T heterozygotes determined by using direct DNA sequencing. With the ddGTP as the chain terminator, the G/T substitution was expected to be associated with the loss of termination segment. However the absence of the band representing the dideoxy component did not appear in this ddF pattern. Although the mutant allele (the minor allele) containing T nucleotide was unable to produce the G terminated segment corresponding to the absence of the band, the wild type allele (the major allele) containing G nucleotide at the polymorphic site was still able to produce the G terminated segment. Therefore the mobility shift of the extra bands occurring in the three individuals was associated with the informative SSCP component generated from the mutant allele.

In figure 3B, a mutation in the GH gene of the individual 10 (indicated by arrow) was identified in the ddF pattern. The polymorphic bands were associated with the C/T substitution of individual 10 examined by direct DNA sequencing. All the bands subsequent to the sequence change showed altered migration when compared to the wild type individuals. Since ddGTP was utilized as the chain terminator, the dideoxy component was not expected to be present in the C/T substitution of the IGFBP-3 gene and in the A/C substitution of the GH gene. Only SSCP component would be informative. Unfortunately the polymorphism accounting for the SSCP component was not observed in the ddF patterns. Liu and Sommer (7) reported that the SSCP component efficiencies varied substantially as a complex function of sequence and size of the fragment, gel metrics and electrophoretic temperature. For a given segment containing a mutation it was likely that mobility shift may be seen under some conditions but not under other conditions. As a consequence, SSCP did not detect all sequence changes in these two PCR fragments in this study. One should realize that if a ddNTP chosen in the termination reaction is not associated with any base variant in the DNA fragment, the dideoxy component will not be generated. The efficiency of ddF to detect mutation will be limited due to the weakness of the SSCP component.

4. Acknowledgement

I thank Dr. Bill Blackhall and Marilyn Jones for their technical assistance and Dr. Herman Raadsma for cattle semen samples. This work was a part of the fellowship program supported by Thailand-Australia Science & Engineering Assistance Project (TASEAP).

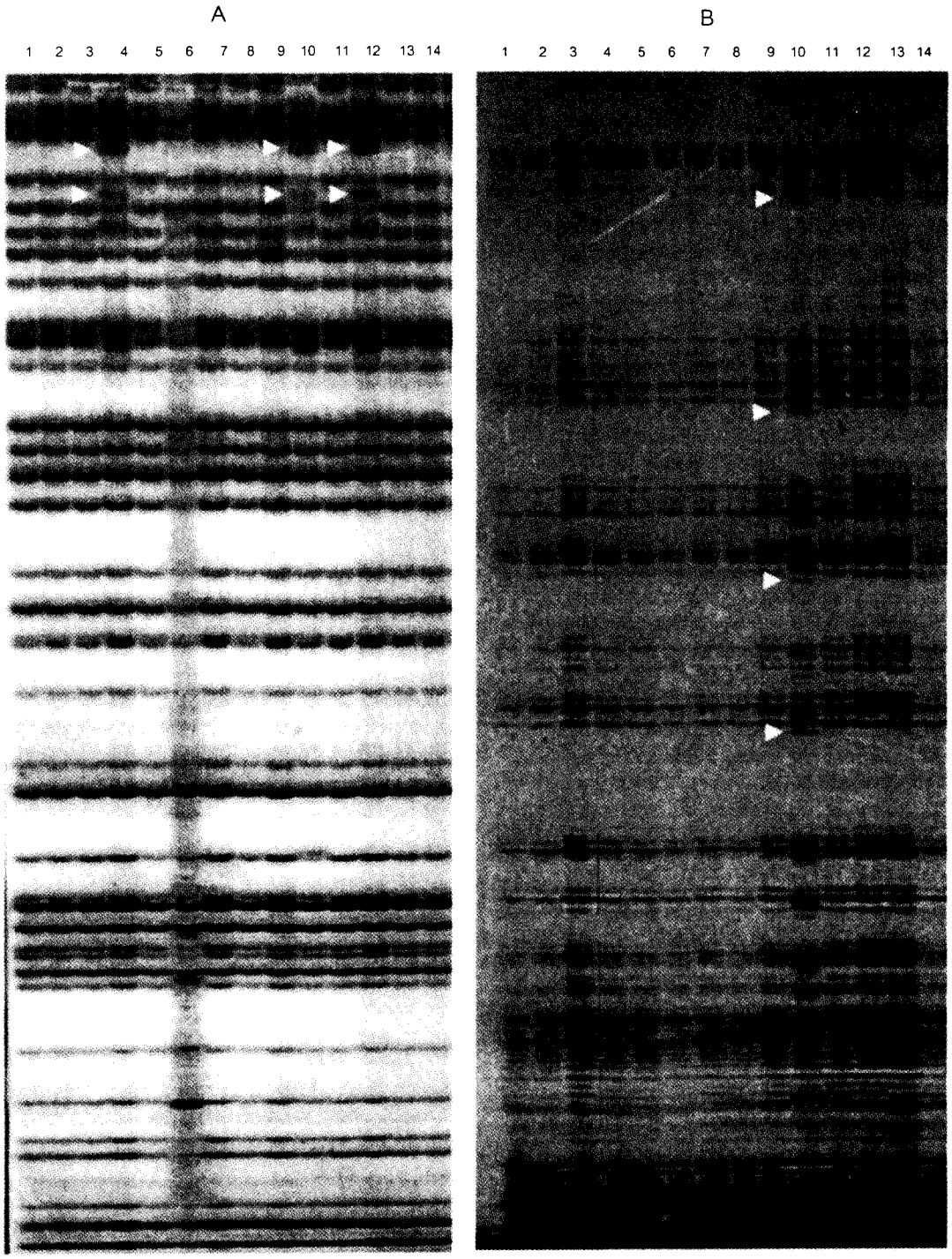

Figure 3 ddF patterns of the IGFBP3 gene (A) and the GH gene (B). Fourteen bulls were screened using ddGTP termination reaction. A) The individual 4, 10, 12 showed an alteration of the band pattern (indicated by arrows) corresponding to the G→T substitution in the IGFBP3 gene. B) The individual 10 showed the extra bands (indicated by arrows) corresponding to the C→T substitution in the GH gene.

5. References

[1] Cooper, D.N., Smith, B.A., Cooke, M.J., Niemann, S. and Schmidtke, J., An Estimate of Unique DNA Sequence Heterozygosity in the Human Genome. Hum. Genet. 69, pp. 201-2051, 985.

[2] Zho, L.P., Aragaki, C., Hsu, L. and Quiaoit, F., Mapping of Complex Traits by Single-Nucleotide Polymorphism. Am. J. Hum. Gent. 63, pp.225-240, 1998.

[3] Sarkar, G., Yoon, H.–S. and Sommer, S.S., Dideoxy fingerprinting (ddF): A rapid and efficient screen for the presence of mutation. Genomics 13, pp.441-443, 1992.

[4] Eillison, J., Squires, G., Crutchfield, C. and Goldman, D., Detection of mutations and polymorphisms using fluorescence-base dideoxy fingerprinting (F-ddF). BioTechniques 17(4), pp.742, 743, 746, 748, 749-753, 1994.

[5] Liu, Q., Feng, J. and Sommer, S.S., Bi-directional didexoy fingerprinting (Bi-ddF): a rapid method for quantitative detection of the mutations in genomic regions of 300-600 bp. Hum. Mol. Genet. 5(1), pp.107-114, 1996.

[6] Heyen, D.W., Beever, J.E., Da, Y., Evert, R.E., Green, C., Bates, S.E.R., Ziegle, J.S. and Lewin, H.A., Exclusion probabilities of 22 bovine microsatellite markers in fluoresenct multiplexes for semiautomated parentage testing. Animal Genetics, 28(1), pp.21-27, 1997.

[7] Liu, Q., and Sommer, S.S., Parameters affecting the sensitivities of dideoxy fingerprinting and PCR. PCR Method Appl. 4(2), pp.97-108, 1994.

[8] Kwoy, P.-Y., Carlson, C., Yager, T., Ankener, W. and Nickerson, D.A., Comparative analysis of human DNA variation by fluoresence-base sequencing of PCR products. Genomics 23, pp.138-144, 1994.

[9] Parker, L.T., Deng, Q., Zakeri, H., Carson, C., Nickerson, D.A. and Kwoy, P. -Y., Peak height variations in automated sequencing of PCR products using *Taq* dye-terminator chemistry. BioTechniques 19(1), pp.116-121, 1995.

[10] Parker, L.T., Zakeri, H., Deng, Q., Spurgeon, S., Kwoy, P.-Y., and Nickerson, D.A., AmpliTaq® DNA polymerase, FS dye terminator sequencing: Analysis of peak height pattern. BioTechniques 21(4), pp.694-699, 1995.

[11] Taillon-Miller, P., Piernot, E.E. and Kwoy, P.-Y., Efficient approach to unique single-nucleotide polymorphism discovery. Genome Res., pp.499-505, 1999.