

Technical Note: Whole-Genome Amplification of DNA Extracted from Cattle Semen Samples

R. J. Hawken,*¹ J. A. L. Cavanagh,†‡# J. R. S. Meadows,* M. S. Khatkar,†# Y. Husaini,†# K. R. Zenger,†# S. McClintock,§# A. E. McClintock,†‡# and H. W. Raadsma†#

*CSIRO Livestock Industries, Queensland Biosciences Precinct, St Lucia, 4067, Queensland, Australia

†The Centre for Advanced Technologies in Animal Genetics and Reproduction (Reprogen) The University of Sydney, Camden, 2570, New South Wales, Australia

‡The Centre for Advanced Technologies in Animal Genetics and Reproduction (Reprogen), Armidale, 2350, New South Wales, Australia

§Animal Genetics and Breeding Unit, University of New England, Armidale, 2351, New South Wales, Australia

#Cooperative Research Centre for Innovative Dairy Products, Melbourne 3000, Victoria, Australia

ABSTRACT

The bovine genome sequence project and the discovery of many thousands of bovine single nucleotide polymorphisms has opened the door for large-scale genotyping studies to identify genes that contribute to economically important traits with relevance to the beef and dairy industries. Large amounts of DNA will be required for these research projects. This study reports the use of the whole-genome amplification (WGA) method to create an unlimited supply of DNA for use in genotyping studies and long-term storage for future gene discovery projects. Two commercial WGA kits (GenomiPhi, Amersham Biosciences, Sydney, Australia, and REPLI-g, Qiagen, Doncaster, Australia) were used to amplify DNA from straws of bull semen, resulting in an average of 7.2 and 67 μ g of DNA per reaction, respectively. The comparison of 3.5 kb of sequences from the amplified and unamplified DNA indicated no detectable DNA differences. Similarly, gene marker analysis conducted on genomic DNA and DNA after WGA indicated no difference in marker amplification or clarity and accuracy of scoring for approximately 10,000 single nucleotide polymorphism markers when compared with WGA samples genotyped in duplicate. These results illustrate that WGA is a suitable method for the amplification and recovery of DNA from bull semen samples for routine genomic investigations.

Key words: whole-genome amplification, GenomiPhi, REPLI-g, DNA extraction

A focus of livestock breeding programs has been the improvement of key industry traits through the selection of superior breeding stock. To this end, research groups have investigated the presence of QTL explaining phenotypic variance for these key traits, with

the ultimate aim of identifying a suite of genetic markers that can be used to increase the efficiency and accuracy of selection of breeding stock. In the past, QTL have been routinely identified in cattle using many hundreds of microsatellite markers strategically selected for genome coverage, or specific genome regions for high-resolution mapping (Georges et al., 1995; Spelman et al., 1996; Ashwell et al., 1998; Schrooten et al., 2000; Holmberg and Andersson-Eklund, 2004; Olsen et al., 2005). With the bovine genome sequence project in the first phase of completion, many thousands of single nucleotide polymorphisms (SNP) are in the public domain (International Bovine Genome Sequence Consortium, <http://www.hgsc.bcm.tmc.edu/projects/bovine/>). These SNP markers in combination with high-throughput low-cost SNP genotyping technologies make it possible to screen more than 10,000 SNP across large cattle populations to accelerate QTL and genetic marker discovery. For these high-throughput genotyping studies, a large amount of starting DNA is generally required (6 μ g for 11,000 to 15,000 SNP). In dairy QTL studies using EBV from progeny-tested bulls, the most convenient tissue to use for DNA extraction is semen obtained from straws prepared for AI. However, the amount of DNA produced from a single straw of semen is often inadequate for such high-throughput genotyping technologies and long-term storage for future gene discovery research. Furthermore, some semen samples are very scarce and difficult to obtain for research purposes because they represent historic bulls or non-commercial bulls for which a limited number of semen straws remain in storage at AI facilities.

Recently, a whole-genome amplification (WGA) technique has evolved that enables the amplification of minute amounts of DNA into large quantities of DNA suitable for large-scale genomic research projects (Dean et al., 2001, 2002; Hosono et al., 2003). This technique has been used for many human studies to facilitate the recovery of patient DNA in cases where the availability of patient tissue has been limited (Anchordoquy et al.,

Received October 3, 2005.

Accepted December 23, 2005.

¹Corresponding author: Rachel.Hawken@csiro.au

Table 1. Yield and amplification efficiencies of DNA obtained for 8 animals, and average, minimum, and maximum yields are presented for all 67 samples using 2 commercial kits¹

Lab no.	Date of birth	Yield of DNA following extraction from semen, μg	GenomiPhi		REPLI-g		Amplification efficiency	
			Amount of DNA starting, ng	Yield of DNA following amplification, μg	Amount of DNA starting, ng	Yield of DNA following amplification, μg	GenomiPhi	REPLI-g
65	07/07/1979	17.5	35	5.5	87.5	45	$\times 157$	$\times 514$
81	08/30/1981	25	50	5.5	125	42	$\times 110$	$\times 336$
47	07/17/1984	5.5	11	4	27.5	44	$\times 364$	$\times 1,600$
72	04/22/1986	13	26	7	65	40	$\times 269$	$\times 615$
402	06/19/1998	23	46	11	115	55	$\times 239$	$\times 478$
395	07/01/1998	6	12	0.2	30	57	$\times 16$	$\times 1,900$
407	07/10/1998	2.5	5	0.5	12.5	48	$\times 100$	$\times 3,840$
390	07/13/1998	43	86	13	215	49	$\times 151$	$\times 228$
DNA quantities for all 67 samples								
Average		11	22.5	7.2	57	67	$\times 1,483$	$\times 3,100$
Minimum		0.25	0.5	0.05	1.25	21	$\times 4$	$\times 228$
Maximum		43	85.5	29	213	128	$\times 59,000$	$\times 32,000$

¹GenomiPhi, Amersham Biosciences, Sydney, Australia; REPLI-g, Qiagen, Doncaster, Australia.

2003; Barker et al., 2004; Bergen et al., 2005a,b; Holbrook et al., 2005; Jiang et al., 2005; Park et al., 2005; Sun et al., 2005; Tzvetkov et al., 2005). The utility of WGA DNA derived from semen in large-scale genomic investigations has not been demonstrated to date. Frozen semen stored in extender diluents for commercial AI purposes represents an additional complexity in which the additives may affect the quality and purity of DNA, potentially interfering with the WGA procedure.

This paper reports on the utility of the WGA procedure on DNA extracted from cattle semen stored in commercial diluents in deep-frozen straws, in an effort to produce an infinite source of DNA for gene discovery purposes. The hypothesis to be tested is that WGA of DNA from bull semen provides sufficient DNA for genotyping purposes and does not introduce significant DNA errors that could alter DNA genotyping results and bias QTL discovery analyses.

Extraction of DNA was achieved using a method described by Heyen et al. (1997) from one-half the contents of a single straw of semen from each of 67 bulls (straws of semen were kindly supplied by Genetics Australia, Bacchus Marsh, Australia); each sample represented 0.07 mL of diluted sperm cells in diluent solution. The yield and concentration of DNA extracted was estimated using PicoGreen DNA quantitation (PicoGreen dsDNA quantitation reagent; Molecular Probes, Inc., Invitrogen, Melbourne, Australia). A large amount of variation was observed in the yield of DNA from different samples (8 examples and averages for all 67 samples are presented in Table 1). This large amount of variation in DNA concentration and yield may reflect the semen count within semen straws or may be the result of a variable DNA extraction procedure.

Whole genome amplification on the DNA samples extracted from bull semen was completed using 2 commercial kits; GenomiPhi (Amersham Biosciences, Sydney, Australia), and REPLI-g (Qiagen, Doncaster, Australia). Whole-genome amplification using the GenomiPhi kit was accomplished using 1 μL of the stock DNA according to the manufacturer's protocol (see Table 1 for DNA concentrations). Whole-genome amplification using the REPLI-g kit was accomplished using 2.5 μL of stock DNA. Amplified DNA from both kits was quantified using the PicoGreen quantification method. An average of 7.2 μg of genomic DNA was produced from the GenomiPhi amplification procedure, whereas an average of 67 μg of DNA was produced using the REPLI-g kit. The starting concentration, WGA amplification yield, and amplification efficiencies of 8 DNA samples and an average for the 67 samples are presented in Table 1.

As expected, a large difference in yield and amplification efficiency was observed between the 2 kits. This difference is attributed to the WGA DNA yields targeted by each kit; the GenomiPhi kit is targeted for the production of smaller amounts of amplified DNA (4 to 7 μg) than the REPLI-g kit (45 μg). There were 2 protocol differences between the kits. First, a lesser amount of starting DNA was used for the amplification procedure using the GenomiPhi kit. Second, the amplified DNA from the GenomiPhi reaction was cleaned using an ethanol precipitation step according to the manufacturer's instructions. Given these differences in procedure, both WGA kits produced an average DNA yield above or at the specified target yield advised by the manufacturer, as illustrated in a comprehensive comparison of WGA kits by Bergen et al. (2005a).

Table 2. Gene fragments sequenced in each of 8 animals using unamplified DNA and GenomiPhi and REPLI-g whole-genome amplification DNA samples¹

Gene name ²	Length sequenced, bp	No. of SNP ³ within the sequenced region
PCSK1	324	4
OCLN	210	2
NRAMP	350	0
NFE2	405	1
HSPC148	256	1
CRSP6	422	1
HP	284	3
GNAS	396	1
IL16	437	0
DES	421	1
Total	3,505	14

¹GenomiPhi, Amersham Biosciences, Sydney, Australia; REPLI-g, Qiagen, Doncaster, Australia.

²PCSK1 = bovine sequence similar to *Homo sapiens* proprotein convertase subtilisin/kexin type 1; OCLN = bovine sequence similar to *H. sapiens* occludin; NRAMP = bovine sequence similar to *H. sapiens* natural resistance-associated macrophage protein 1; NFE2 = bovine sequence similar to *H. sapiens* nuclear factor (erythroid-derived 2); HSPC148 = bovine sequence similar to *H. sapiens* hypothetical protein hspc148; CRSP6 = bovine sequence similar to *H. sapiens* cofactor required for Sp1 transcriptional activation, subunit 6; HP = bovine sequence similar to *H. sapiens* haptoglobin; GNAS = bovine sequence similar to *H. sapiens* GNAS complex locus; IL16 = bovine sequence similar to *H. sapiens* interleukin 16; DES = bovine sequence similar to *H. sapiens* desmin.

³SNP = Single nucleotide polymorphism.

To determine if the WGA procedure introduced errors in the DNA sequence, which could be mistaken as DNA polymorphisms in future SNP discovery and genotyping research, 8 DNA samples (listed in Table 1) were sequenced across 10 gene regions using both the unamplified and the WGA DNA (produced from both WGA kits). These 10 gene regions have been sequenced previously for SNP discovery and cover 3.5 kb of gene sequence (Table 2; Hawken et al., 2004). Twenty nanograms of DNA from each of the unamplified DNA samples, and the GenomiPhi and REPLI-g reactions were used in 20- μ L PCR reactions using HotStar Taq (Qiagen) and standard PCR conditions (Hawken et al., 2004). Amplified products were treated with exosapIT (Amersham Biosciences) to remove unincorporated nucleotides and amplification oligonucleotides. Each PCR amplicon was sequenced in both the forward and reverse direction using the PCR amplification oligonucleotides. Sequencing was accomplished using BigDye version 3.1 (Applied Biosystems, Scoresby, Australia) on an ABI 377 sequencer according to the manufacturer's instructions. Sequences were aligned to each other using Sequencher (Gene Codes, Ann Arbor, MI).

Sequence alignments and trace files from each of 8 DNA samples for the 10 amplicons can be found at

<http://www.livestockgenomics.csiro.au/WGAsequence/>. In total, 3,505 bp were sequenced across each of 8 animals, resulting in approximately 28 kb of sequence to assess DNA differences introduced as a result of the WGA procedure (Table 2). No differences in DNA sequence were observed following the comparison of DNA sequences between unamplified and WGA DNA. Furthermore, SNP detected within the gene fragments by analysis of sequencing traces showed identical alleles in the WGA DNA compared with the unamplified DNA.

To further assess the accuracy of genotyping calls and identify any allelic imbalance resulting from the WGA procedure (illustrated by Bergen et al., 2005a), 2 unamplified and corresponding WGA individuals (REPLI-g) were genotyped using highly saturated genome-wide SNP loci (MegAllele Genotyping Bovine 10,000 SNP Panel, ParAllele, San Francisco, CA). In addition, 17 samples that had been amplified by WGA were genotyped in duplicate to compare the genotyping accuracy and consistency rate to the WGA vs. non-WGA genotyping results. The no-call rate, inconsistent genotype rate, and individual multilocus heterozygosity values (Slate and Pemberton, 2002) were calculated between genotypes of duplicates animals and between WGA and non-WGA samples. The results indicate that the difference in multilocus heterozygosity and no-call rates between the WGA and non-WGA genotypes were similar to the WGA samples genotyped in duplicate (Tables 3 and 4). The inconsistent results appeared to be higher in the comparison of WGA vs. non-WGA DNA (Table 3). However, they were also similar to the WGA vs. WGA inconsistent calls for the same animals, indicating that the inconsistent calls were a result of the genotyping procedure, rather than the WGA procedure. Finally, these data suggest that the WGA procedure (from the REPLI-g kit) did not introduce detectable DNA errors and did not preferentially bias particular genomic regions when using bovine DNA extracted from semen.

The listed price of each of these kits is AUS\$774 (US\$420) for the 100-reaction (4 to 7 μ g) GenomiPhi kit and AUS \$2,855 (US\$1,500) for the 100-reaction (45 μ g) REPLI-g kit. This equates to AUS\$7.74 (US\$4.20) per sample, producing an average yield of 7.2 μ g using the GenomiPhi kit, and AUS\$28.50 (US\$15.00) per sample, producing an average yield of 67 μ g of DNA using REPLI-g. For those semen samples that are in short supply, WGA appears to be a robust technique to produce a sufficient amount of DNA for future large-scale genotyping and other gene discovery studies. Whole-genome amplification is a valuable tool for the recovery of DNA from semen samples of limited availability.

Table 3. Multilocus heterozygosity (MLH), no-call rates, and inconsistent genotyping calls from 9,710 markers across 2 DNA samples (1432 and 1545) comparing genotypes between whole-genome amplification (WGA, in duplicate: WGA1 and WGA2) and non-WGA DNA

DNA ID	Sample comparison	DNA concentration used for WGA (ng/ μ L)	No call in both ¹	No call in either ¹	Inconsistent calls	MLH	Difference in MLH between samples compared
1432	Non-WGA ² vs. WGA1	157 ²	37	392	68	0.3195	-0.0008
		78				0.3203	
1432	Non-WGA vs. WGA2	157 ²	37	345	53	0.3195	0.0033
		78				0.3162	
1432	WGA1 vs. WGA2	78	40	275	71	0.3203	0.0041
						0.3162	
1545	Non-WGA vs. WGA1	116 ²	35	608	102	0.3055	0.0226
		82				0.2829	
1545	Non-WGA vs. WGA2	116 ²	29	531	101	0.3055	0.0194
		82				0.2862	
1545	WGA1 vs. WGA2	82	77	730	106	0.2829	-0.0033
						0.2862	

¹No call = failure to obtain an accurate genotype; no call in both = the number of identical single nucleotide polymorphisms for which an accurate genotype could not be determined in both DNA samples; no call in either = total number of single nucleotide polymorphisms for both samples for which an accurate genotype could not be determined.

²These samples were not whole-genome amplified.

Table 4. Multilocus heterozygosity (MLH), no-call rates, and inconsistent genotyping calls from 9,710 markers across 17 samples produced by whole-genome amplification (WGA) and genotyped in duplicate

DNA ID	DNA concentration used for WGA (ng/ μ L)	No call in both ¹	No call in either ¹	Inconsistent calls	MLH (duplicate samples)	Difference in MLH between duplicates
15	138	59	194	21	0.2978	0.0002
					0.2976	
21	71	97	601	69	0.2808	0.0025
					0.2784	
133	20	38	206	12	0.3266	0.0008
					0.3274	
135	101	35	142	16	0.3116	0.0009
					0.3107	
413	109	47	132	16	0.2842	0.0004
					0.2838	
433	54	38	118	15	0.3154	0.0012
					0.3142	
447	11	45	138	6	0.2575	0.0014
					0.2589	
453	46	42	107	4	0.3009	0.0002
					0.3007	
1542	168	45	132	13	0.2916	0.0000
					0.2916	
1594	155	27	209	7	0.3011	0.0004
					0.3014	
1595	67	34	108	4	0.3027	0.0004
					0.3023	
1601	57	99	202	9	0.3086	0.0001
					0.3085	
1617	103	43	120	6	0.3227	0.0000
					0.3227	
1619	68	73	335	33	0.2960	0.0015
					0.2975	
1768	23	50	129	2	0.3066	0.0003
					0.3063	

¹No call = failure to obtain an accurate genotype; no call in both = the number of identical single nucleotide polymorphisms for which an accurate genotype could not be determined in both DNA samples; no call in either = total number of single nucleotide polymorphisms for both samples for which an accurate genotype could not be determined.

ACKNOWLEDGMENTS

The authors wish to thank Joel Peacock, Gina Attard, Lee Webley, and Marilyn Jones for their laboratory support. This work was supported by the Cooperative Research Centre for Innovative Dairy Products, Australia.

REFERENCES

- Anchordoquy, H. C., C. McGeary, L. Liu, K. S. Krauter, and A. Smolen. 2003. Genotyping of three candidate genes after whole-genome preamplification of DNA collected from buccal cells. *Behav. Genet.* 33:73–78.
- Ashwell, M. S., Y. Da, C. P. Van Tassell, P. M. Vanraden, R. H. Miller, and C. E. Rexroad, Jr. 1998. Detection of putative loci affecting milk production and composition, health, and type traits in a United States Holstein population. *J. Dairy Sci.* 81:3309–3314.
- Barker, D. L., M. S. Hansen, A. F. Faruqi, D. Giannola, O. R. Irsula, R. S. Lasken, M. Latterich, V. Makarov, A. Oliphant, J. H. Pinter, R. Shen, I. Sleptsova, W. Ziehler, and E. Lai. 2004. Two methods of whole-genome amplification enable accurate genotyping across a 2320-SNP linkage panel. *Genome Res.* 14:901–907.
- Bergen, A. W., K. A. Haque, Y. Qi, M. B. Beerman, M. Garcia-Closas, N. Rothman, and S. J. Chanock. 2005a. Comparison of yield and genotyping performance of multiple displacement amplification and OmniPlex whole genome amplified DNA generated from multiple DNA sources. *Hum. Mutat.* 26:262–270.
- Bergen, A. W., Y. Qi, K. A. Haque, R. A. Welsh, and S. J. Chanock. 2005b. Effects of DNA mass on multiple displacement whole genome amplification and genotyping performance. *BMC Biotechnol.* 5:24–34.
- Dean, F. B., S. Hosono, L. Fang, X. Wu, A. F. Faruqi, P. Bray-Ward, Z. Sun, Q. Zong, Y. Du, J. Du, M. Driscoll, W. Song, S. F. Kingsmore, M. Egholm, and R. S. Lasken. 2002. Comprehensive human genome amplification using multiple displacement amplification. *Proc. Natl. Acad. Sci. USA*; 99:5261–5266.
- Dean, F. B., J. R. Nelson, T. L. Giesler, and R. S. Lasken. 2001. Rapid amplification of plasmid and phage DNA using Phi 29 DNA polymerase and multiply primed rolling circle amplification. *Genome Res.* 11:1095–1099.
- Georges, M., D. Nielsen, M. Mackinnon, A. Mishra, R. Okimoto, A. T. Pasquino, L. S. Sargeant, A. Sorensen, M. R. Steele, X. Zhao, J. E. Womack, and I. Hoeschele. 1995. Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. *Genetics* 139:907–920.
- Hawken, R. J., W. C. Barris, S. M. McWilliam, and B. P. Dalrymple. 2004. An interactive bovine in silico SNP database (IBISS). *Mamm. Genome* 15:819–827.
- Heyen, D. W., J. E. Beever, Y. Da, R. E. Evert, C. Green, S. R. Bates, J. S. Ziegler, and H. A. Lewin. 1997. Exclusion probabilities of 22 bovine microsatellite markers in fluorescent multiplexes for semiautomated parentage testing. *Anim. Genet.* 28:21–27.
- Holbrook, J. F., D. Stabley, and K. Sol-Church. 2005. Exploring whole genome amplification as a DNA recovery tool for molecular genetic studies. *J. Biomol. Technol.* 16:125–133.
- Holmberg, M., and L. Andersson-Eklund. 2004. Quantitative trait loci affecting health traits in Swedish dairy cattle. *J. Dairy Sci.* 87:2653–2659.
- Hosono, S., A. F. Faruqi, F. B. Dean, Y. Du, Z. Sun, X. Wu, J. Du, S. F. Kingsmore, M. Egholm, and R. S. Lasken. 2003. Unbiased whole-genome amplification directly from clinical samples. *Genome Res.* 13:954–964.
- Jiang, Z., H. Wang, Y. Ma, and D. Wei. 2005. Characterization of two novel lipase genes isolated directly from environmental sample. *Appl. Microbiol. Biotechnol.* 8:1–6.
- Olsen, H. G., S. Lien, M. Gautier, H. Nilsen, A. Roseth, P. R. Berg, K. K. Sundaasen, M. Svendsen, and T. H. Meuwissen. 2005. Mapping of a milk production quantitative trait locus to a 420-kb region on bovine chromosome 6. *Genetics* 169:275–283.
- Park, J. W., T. H. Beaty, P. Boyce, A. F. Scott, and I. McIntosh. 2005. Comparing whole-genome amplification methods and sources of biological samples for single-nucleotide polymorphism genotyping. *Clin. Chem.* 51:1520–1523.
- Schrooten, C., H. Bovenhuis, W. Coppieters, and J. A. M. Van Arendonk. 2000. Whole genome scan to detect quantitative trait loci for conformation and functional traits in dairy cattle. *J. Dairy Sci.* 83:795–806.
- Slate, J., and J. M. Pemberton. 2002. Comparing molecular measures for detecting inbreeding depression. *J. Evol. Biol.* 15:20–31.
- Spelman, R. J., W. Coppieters, L. Karim, J. A. M. van Arendonk, and H. Bovenhuis. 1996. Quantitative trait loci analysis for five milk production traits on chromosome six in the Dutch Holstein-Friesian population. *Genetics* 144:1799–1808.
- Sun, G., R. Kaushal, P. Pal, M. Wolujewicz, D. Smelser, H. Cheng, M. Lu, R. Chakraborty, L. Jin, and R. Deka. 2005. Whole-genome amplification: Relative efficiencies of the current methods. *Leg. Med. (Tokyo).* 7:279–286.
- Tzvetkov, M. V., C. Becker, B. Kulle, P. Nurnberg, J. Brockmoller, and L. Wojnowski. 2005. Genome-wide single-nucleotide polymorphism arrays demonstrate high fidelity of multiple displacement-based whole-genome amplification. *Electrophoresis* 26:710–715.