

# DISCOVERY, ANALYSIS, AND UTILITY OF BOVINE SINGLE NUCLEOTIDE POLYMORPHISMS

Gabrielle Simone Sellick
July, 2002

A thesis submitted in partial fulfilment of the award of Ph.D. in the Department of Animal Science at The University of Adelaide, Australia

## TABLE OF CONTENTS

Table of Contents	<u>Page</u> ii
Abstract	viii
List of Figures	ix
List of Tables	xi
Abbreviations	xiii
Declaration	xvi
Dedication	xvii
Acknowledgments	xviii
Acknowledgments	
CHAPTER 1. Introduction and literature review	1
1.1 Introduction	2
1.2 Objectives of this study	2
1.3 Literature review	3
1.3.1 Typing genetic variation	3
1.3.1.1 Protein polymorphisms	3
1.3.1.2 Polymorphisms within repetitive DNA elements	4
1.3.1.3 Single nucleotide polymorphisms	8
1.3.2 SNP discovery methods	10
1.3.2.1 Established SNP detection methods	11
1.3.2.1.1 DNA sequencing	11
1.3.2.1.2 Restriction fragment length polymorphisms (RFLP)	12
1.3.2.1.3 Single stranded conformation polymorphism (SSCP)	12
1.3.2.1.4 Dideoxy fingerprinting (ddF)	14
1.3.2.1.5 Denaturing gradient gel electrophoresis (DGGE)	15
1.3.2.1.6 Heteroduplex analysis (HA)	15
1.3.2.1.7 Chemical cleavage of mismatch (CCM)	16
1.3.2.1.8 RNase cleavage	16
1.3.2.2 New SNP detection methods	17
1.3.2.2.1 SNP discovery from overlapping genomic sequences	18
1.3.2.2.2 Variant detection arrays (VDA)	19
1.3.3 Diagnostic methods for genotyping known SNPs	21
1.3.3.1 Allele-specific amplification (ASA)	21
1.3.3.2 Allele specific hybridization (ASH)	24

	iii
1.3.3.2.1 Array hybridization assays	24
1.3.3.2.2 Homogenous hybridization assays	26
1.3.3.3 Oligonucleotide ligation assay (OLA)	27
1.3.3.4 Primer extension	28
1.3.4 SNP validation and allele frequency determination	30
1.3.5 Functional applications of SNPs	32
1.3.5.1 SNPs as markers in genetic studies	32
1.3.5.2 SNPs as markers for routine parentage and identity testing	34
1.4 Research plan	36
CHAPTER 2. General methods	37
2.1 Isolation of genomic DNA from whole blood	38
2.2 Optimization of PCR conditions	39
2.3 Analysis of PCR products by agarose gel electrophoresis	40
2.4 Preparation of PCR template for automated sequencing	40
2.5 Automated cycle sequencing of PCR products	40
2.6 Description of the Adelaide/AgResearch cattle QTL mapping design	41
CHAPTER 3. Discovery of Bovine SNPs	43
3.1 Introduction	44
3.2 Specific methods	46
3.2.1 PCR-RFLP analysis	46
3.2.2 Dideoxy fingerprinting	46
3.2.2.1 Primer selection and design	49
3.2.2.2 Template production	49
3.2.2.3 End-labelling primers	50
3.2.2.4 Cycle-sequencing with a single ddNTP	50
3.2.2.5 DdF gel electrophoresis	51
3.3 Results	52
3.3.1 Amplification success rate of bovine, ovine and CATS primers	52
3.3.2 PCR-RFLP analysis of 22 bovine SNPs	52
3.3.3 Analysis of ddF as a tool for novel SNP discovery	53
3.4 Discussion	56

		iv
CH.	APTER 4. Characterization of SNPs	61
л 1	Introduction	62
	Specific methods	66
7.2	4.2.1 Sequence analysis	66
	4.2.1.1 Estimation of bovine sequence polymorphism	66
	4.2.1.2 CpG dinucleotide involvement	67
	4.2.2 Gene ontology assignment	67
4.3	Results	68
	4.3.1 Screening of genes and identification of polymorphisms	68
	4.3.2 Characteristics of bovine SNPs	68
	4.3.2.1 Types and distribution of polymorphisms in genes	68
	4.3.2.2 Distribution of polymorphisms among genes	73
	4.3.3 Analysis of CpG dinucleotide involvement	75
4.4	Discussion	77
CH	APTER 5. Analysis of SNP allele frequencies in cattle	86
5.1	Introduction	87
	Specific methods	90
	5.2.1 Chromosomal assignment of genes	90
	5.2.1.1 Ovine BAC library screening	91
	5.2.1.2 Physical mapping by fluorescence in situ hybridization	91
	5.2.1.3 Fluorescence in situ hybridization using tyramide signal	
	amplification	92
	5.2.1.4 Dual-FISH	92
	5.2.2 Determination of SNP allele frequency	92
	5.2.2.1 Genomic DNA isolation and pooling	93
	5.2.2.2 PCR amplification, DNA sequencing and sequence analysis	95
	5.2.3 Estimating exclusion probability $(P_{\rm E})$	95
5.3	Results	96
	5.3.1 Comparative physical mapping of type I markers in cattle and sheep	96
	5.3.2 Physical Distribution of SNPs	99
	5.3.3 Distribution of SNP allele frequency in different classes of nucleotide	
	sites 100	100
	5 3 4 Analysis of SNP allele frequency among groups of cattle breeds	100

*	
٠,	ı

	5.3.5	Preliminary SNP marker selection for cattle parentage and identity	
		testing	105
5.4	Discussi	on	107
СН	APTER	6. Analysis of SNP genotyping methods	113
6.1	Introduc	tion	114
6.2	Specific	methods	118
	6.2.1	Genomic DNA samples	118
	6.2.2	SNP genotyping by direct hybridization on microspheres	118
		6.2.2.1 PCR primer design and amplification	120
		6.2.2.2 Microsphere capture probe and synthetic control design	120
		6.2.2.3 Coupling of capture probes to microspheres	120
		6.2.2.4 Exonuclease treatment of PCR products	121
		6.2.2.5 Hybridization of templates to capture oligonucleotide-coupled	
		microspheres	- 121
		6.2.2.6 Flow cytometry and genotype analysis	122
	6.2.3	Structural analysis of single stranded PCR templates	123
	6.2.4	SNP genotyping by indirect colorimetric primer extension	123
		6.2.4.1 PCR primer design, amplification and extension primer design	125
		6.2.4.2 Coupling of extension probes to plates	126
		6.2.4.3 Exonuclease digestion of PCR templates and primer extension	
		in 96-well plates	126
		6.2.4.4 Data analysis and genotype assignment	128
6.3	Results		129
	6.3.1	SNP allele discrimination by direct hybridization on microspheres	129
		6.3.1.1 Variations in microsphere coupling	129
		6.3.1.2 Optimization of synthetic oligonucleotide allele discrimination	129
		6.3.1.3 Hybridization experiments with undigested PCR amplicons	130
		6.3.1.4 Undigested PCR amplicon titration	132
		6.3.1.5 Variation in hybridization time	134
		6.3.1.6 Varying the stage of SA-PE addition	134
		6.3.1.7 Varying hybridization scheme	135
		6.3.1.8 Exonuclease treatment of PCR amplicons	135
	6.3.2	2 Analysis of single stranded PCR amplicon secondary structure	135
	6.3.3	SNP genotyping using indirect colorimetric primer extension	138

	V1
6.3.3.1 Effect of DNA quality on assay robustness	142
6.4 Discussion	146
6.4.1 Assessment of the allele-specific hybridization experiments	146
6.4.2 Assessment of the primer extension experiments	150
6.4.3 Comparison of allele-specific hybridization and primer extension assays	153
CHAPTER 7. Nucleotide diversity and haplotype structure of five bovine	
candidate genes in two diverse cattle breeds	156
7.1 Introduction	157
7.2 Specific methods	162
7.2.1 Candidate gene SNP discovery	162
7.2.2 Animals	165
7.2.3 SNP genotyping	165
7.2.4 SNP haplotype analysis	166
7.3 Results	167
7.3.1 SNP discovery screening of five bovine candidate genes	167
7.3.2 Characterization of candidate gene haplotype frequencies and structure in	
Limousin and Jersey cattle	170
7.4 Discussion	176
CHAPTER 8. Genetic linkage and linkage disequilibrium analysis of complex	
traits in cattle	180
8.1 Introduction	181
8.2 Specific methods	189
8.2.1 Animals	189
8.2.2 Slaughter and phenotypic data collection	189
8.2.3 Marker selection and genotyping	191
8.2.4 Linkage mapping	192
8.2.4.1 Map construction	192
8.2.4.2 Statistical linkage analysis	192
8.2.4.3 Thresholds	193
8.2.5 Linkage disequilibrium statistical analysis	193
8.3 Results	195
8.3.1 Haplotype phase assignment	195

	V11
8.3.2 Candidate gene linkage analysis	195
8.3.2.1 Genetic mapping and chromosomal location	195
8.3.2.2 Quantitative trait loci (QTL) mapping	197
8.3.3 Candidate gene linkage disequilibrium analysis	201
8.4 Discussion	206
CHAPTER 9. General Discussion	212
9.1 Isolation of SNPs in livestock species	213
9.2 Validation of SNPs in livestock species	215
9.3 Genotyping of SNPs	217
9.4 SNP applications in livestock species	219
9.4.1 Example: MSTN in cattle	221
9.5 Summary	226
APPENDICES	229
Appendix I: Solutions	230
Appendix II: Primers, amplification conditions and dideoxy fingerprinting results for	
100 bovine loci	237
Appendix III: SNP sequence context of 244 bovine SNPs	253
Appendix IV: Primers and reaction conditions used to screen the ovine BAC library	263
Appendix V: Allele frequency and gene map location of 187 bovine SNPs in four	
distinct cattle samples	264
Appendix VI: Oligonucleotides used for SNP genotyping by direct hybridization on	
microspheres	269
Appendix VII: Oligonucleotides used for primer extension based SNP genotyping	
(Experiment 1)	272
Appendix VIII: Oligonucleotides used for primer extension based SNP genotyping	
(Experiment 2)	274
Appendix IX: SNP sequence context, position and gene location of 54 bovine SNPs	
isolated from sequencing 5 candidate genes	277
Appendix X: Predictive equations used to calculate MEAT%, FAT% and BONE%	
in the Australian Limousin and Jersey Backcross progeny	280
REFERENCES	283
	220
ADDENDUM	338

#### **ABSTRACT**

DNA sequence variation provides the fundamental material for improving livestock through selection. To date, few studies in cattle have concentrated on the isolation of single nucleotide polymorphisms (SNPs) to describe rates of nucleotide diversity across the bovine genome. In an attempt to determine the future value of SNP markers in livestock genomics, the screening of 97 bovine gene fragments amplified from a diverse panel of individuals was undertaken. Over 200 bovine SNPs were isolated from 70 known genes (1 SNP every 231bp), allowing the first large-scale analysis of SNP occurrence and comparison of the rates of bovine nucleotide polymorphism with other mammalian species. To further investigate the properties of bovine SNPs, pooled DNA sequencing analysis of 48 bovine gene fragments containing 187 SNPs were analyzed for minor allele frequency in four samples constructed from 10 common *Bos taurus* and *Bos indicus* cattle breeds. Analysis showed 28% of SNPs were common (≥ 20%) in all four pools providing vital evidence for the existence of informative bovine SNPs shared across many diverse cattle populations.

To clarify the level of assay flexibility, ease of implementation, and overall rates of genotyping success, two distinct SNP genotyping technologies based on either primer extension or primer hybridization were assessed. Analysis of different SNP base variants and in/dels coupled with the genotyping of over 1000 individuals per SNP clearly showed that primer extension-based SNP genotyping was a robust technology suitable for generating confident SNP genotypes. Furthermore, resultant SNP genotypes were used to highlight the successful application of SNPs in the genetic linkage mapping of type I markers for enhanced comparative map resolution and the development of candidate gene SNP haplotypes for linkage (QTL) and linkage disequilibrium mapping of 11 complex traits in cattle. Compelling evidence for a QTL on BTA2 affecting economically important carcass traits was found in the region harbouring the MSTN gene. Linkage disequilibrium analysis confirmed the association of MSTN haplotypes with total meat and fat percentage traits in Limousin cattle. In total, almost 290 bovine SNPs located within 70 known genes were described, resulting in a resource which will continue to provide key DNA markers for future comparative mapping strategies and association studies of candidate genes with complex traits in cattle.

## **LIST OF FIGURES**

<u>]</u>	Page
Figure 1.1: SSCP and ddF strategies for the discovery of single nucleotide	
polymorphisms	13
Figure 1.2: Variant detection arrays	20
Figure 1.3: Principles of four techniques for distinguishing between single nucleotide	
polymorphisms	23
Figure 1.4: Modular design of assays available for SNP genotyping	31
Figure 3.1: Dideoxy fingerprinting gels showing fragments amplified from a panel of	
24 unrelated cattle	54
Figure 4.1: Distribution of SNPs by type of nucleotide substitution (a) and cSNPs by	
codon position (b)	71
Figure 4.2: Functional class distribution of the 97 genes screened for polymorphisms (a	ı)
and resultant cSNPs found (b)	74
Figure 5.1: Breed pooling strategy used to determine SNP allele frequency in cattle	94
Figure 5.2: Localization of a BAC clone containing the KRTAP6.1 gene to 1q1.2 by	
TSA-FISH to R-banded male bovine metaphase chromosomes	98
Figure 5.3: Determination of SNP allele frequency differences for CAS279 in 4 diverse	<b>;</b>
groups of cattle	101
Figure 5.4: SNP minor allele frequency distribution within four distinct cattle pools	
and one superpool	104
Figure 6.1: Schematic representation of the individual steps of SNP genotyping by	
direct hybridization on the surface of microspheres	119
Figure 6.2: Schematic representation of the individual steps of SNP genotyping by	
indirect colorimetric primer extension in 96 well plate format	124
Figure 6.3: SNP genotyping by multiplexed direct hybridization on microspheres	
for the loci LEP154	133
Figure 6.4: Predicted amplicon secondary structure for four loci tested by direct	
hybridization at two different incubation temperatures	137
Figure 6.5: Effect of DNA quality on primer extension assay robustness	144

Figure 6.6: Possible effects of rare and unknown SNP alleles on primer extension	
genotyping results generated for LIP158	145
Figure 7.1: Organization of five bovine candidate genes and location of 54 identified	
SNPs	168
Figure 7.2: Comparison of haplotype number and frequency in five candidate genes	
in Limousin and Jersey cattle	174
Figure 8.1: Principles of linkage and linkage disequilibrium analysis	183
Figure 8.2: Bovine genetic linkage maps of BTA1, BTA2, BTA4, BTA18 and BTA19	196

## LIST OF TABLES

<u>P</u>	Page
Table 3.1: Primers, reaction conditions and allele frequencies of 22 bovine PCR-RFLPs	47
Table 4.1: DNA sequence diversity for different classes of nucleotide sites	69
Table 4.2: CpG mutations at NCGN sequences	76
Table 5.1: Physical map locations of ovine BAC clones localized to R-banded	
metaphase chromosomes by fluorescence in situ hybridization	97
Table 5.2: Population allele frequencies of 187 SNPs from 48 bovine genes	103
Table 5.3: Polymorphisms with a 20% minor allele frequency or greater in all four	
cattle populations	106
Table 7.1: Summary of previous studies reporting effects of GH1, LEP, LIPE, MSTN	
and POU1F1 on body composition in mammals	161
Table 7.2: PCR primers, reaction conditions and SNPs identified from bovine	
candidate gene SNP screening	163
Table 7.3: Candidate gene sequence polymorphisms genotyped in the Limousin and	
Jersey backcross progeny and associated SNP population frequencies	171
Table 7.4: Australian dam haplotype allele and animal genotype frequencies	
for fivebovine candidate genes	173
Table 7.5: Haplotype sharing and pairwise linkage disequilibrium among Limousin and	
Jersey cattle	175
Table 8.1: Summary of $F$ st values and QTL location for traits showing suggestive	
linkage (P < 0.05) using genome-wide threshold levels	198
Table 8.2: Adjusted phenotypic correlations (> +/- 0.30) between carcass muscularity	
and fatness traits in the Jersey and Limousin backcross progeny	199
Table 8.3: Effects of suggestive QTL located on BTA2 on carcass traits in units of	
phenotypic standard deviation	200
Table 8.4: ANOVA variance ratios (Fst) and tests of significance for main effects and	
their interactions	203

Table 8.5: Estimated genetic effects of the MSTN haplotypes on eleven carcass traits inLimousin and Jersey cattle205

#### **ABBREVIATIONS**

A adenine

Abs absorbance

**AP** alkaline phosphatase

**ASA** allele specific amplification

**ASH** allele specific hybridization

BAC bacterial artificial chromosome

**bp** base pair

**BSA** bovine serum albumin

BTA Bos taurus

C cytosine

°C degrees Celcius

cm centimetre

cM centiMorgan

**CpG** C and G dinucleotide pair (running 5' to 3')

**cSNP** coding single nucleotide polymorphism

**dATP** 2'-deoxyadenosine 5'-triphophate

dCTP 2'-deoxycytosine 5'-triphophate

ddF dideoxy fingerprinting

**dGTP** 2'-deoxyguanosine 5'-triphophate

dH<sub>2</sub>O distilled water

**DNA** deoxyribonucleic acid

**dNTP** deoxyribonucleotide triphosphate

dsDNA double stranded deoxyribonucleic acid

dTTP 2'-deoxythymidine 5'-triphophate

EDC 1-ethyl-3-(3-dimethyl-aminopropyl) carbodiimide hydrochloride

EDTA ethylenediamine tetra acetic acid

**EST** expressed sequence tag

**EtBr** ethidium bromide

**FISH** fluorescence in situ hybridisation

**FITC** fluorescein isothiocyanate

fmol fentamole

**g** gram

**G** guanine

**hr** hour

HRP

horse radish peroxidase

**HSA** 

Homo sapiens

in/del

insertion/deletion

kb

kilobase pair

L

litre

LB

Luria broth

LD

linkage disequilibrium

LOD

Logarithm of odds score

 $\mathbf{M}$ 

molar

MAS

marker assisted selection

mA

milliampere

Mb

million base pairs

**MFI** 

mean fluorescence intensity

min

minute

mM

millimolar

mRNA

messenger ribonucleic acid

ng

nanograms

nm

nanometre

OAR

Ovis aries

OD

optical density

**ODA** 

*N*,*N*-dimethyloctylamine

**PBS** 

phosphate buffer saline

PCR

polymerase chain reaction

PE

phycoerythrin

 $P_{\mathrm{E}}$ 

probability of exclusion

pers. comm.

personal communication

pg

picograms

pmol

picomole

pNPPD

p-Nitrophenyl phosphate solution

PPD11

ρ-phenylenediamine dihydrochloride, pH 11

QTL

quantitative trait loci

RFLP

restriction fragment length polymorphism

RNA

ribonucleic acid

rpm

revolutions per minute

**SA-HRP** 

streptavidin-conjugated horse radish peroxidase

SA-PE

streptavidin-R-phycoerythrin

SDS

sodium dodecyl sulphate

sec

second

**SNP** 

single nucleotide polymorphism

SSC

sodium chloride sodium citrate

**SSCP** 

single stranded conformational polymorphism

**ssDNA** 

single stranded deoxyribonucleic acid

**STR** 

short tandem repeat

**STS** 

sequence tagged site

T

thymine

TAE

tris acetate ethylenediaminetetra-acetic acid

Taq

Thermus aquaticus

TBE

tris borate ethylenediaminetetra-acetic acid

TE

tris ethylenediaminetetra-acetic acid

**TEMED** 

N,N,N',N'-tetramethylethylenediamine

 $T_{m}$ 

melting temperature

**TMAC** 

tetramethylammonium-chloride

**TSA** 

tyramide signal amplification

μg

microgram

 $\mu$ l

microlitre

 $\mu$ **M** 

micromolar

U

restriction endonuclease unit

UTR

untranslated region

UV

ultraviolet light

 $\mathbf{V}$ 

volts

**VDA** 

variation detection array

 $\mathbf{v}/\mathbf{v}$ 

volume for volume

w/v

weight for volume

#### **DECLARATION**

I certify that this thesis contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. I give consent to this copy of my thesis, when deposited in the University Library, being available for loan and photocopying.

G. S. Sellick 23/7/02

## **DEDICATION**

For my father, who always wanted to see me wear a "black cap".

Dad, I know you will always be there close, somewhere.

For my mother, whose enduring love, strength and support has been forever inspirational.

### **ACKNOWLEDGMENTS**

I would sincerely like to thank my supervisor Dr. Cynthia Bottema for her resolute support and encouragement throughout my candidature and for the vast opportunities presented to me to work, learn and experience the many different facets of science both nationally and internationally. I would also like to thank my co-supervisors, Professor Herman Raadsma and Dr. Wayne Pitchford for their helpful editorial comments on my thesis.

I am indebted to Orchid BioSciences Inc. for their contributions of reagents and materials for the primer-extension based SNP genotyping implemented in this project. Great thanks must also go to the J.S. Davies Bequest (University of Adelaide) and the Foundation for Research, Science and Technology (New Zealand) for access to the many thousands of microsatellite genotypes and phenotypes generated as part of the highly successful Adelaide/AgResearch beef cattle QTL mapping trial.

Thanks also to John Rader (GeneScreen, Dallas, TX), Jim Jacobson and Santosh Arcot (Luminex, Austin, TX), Mark Kunkel, Mark Pohl and Minh Xu (Orchid BioSciences, Princeton, NJ) for their hospitality and helpful discussions on the many intricacies of SNP genotyping assay development and analysis; Marilyn Jones and Dr. William Blackhall (University of Sydney) for extraction of the cattle DNA pools and verification of the pooled SNP allele frequency sequencing strategy; Dr. Christopher Smyle, Bozena Kruk and Marilyn Jones for their assistance with the pooled SNP sequencing; Dr. Graham Webb for his help with FISH gene assignments; Adam Kister for generating the combined genotype databases and guidance in using the programs implemented in the linkage/linkage disequilibrium studies; Dave Rutley for developing the saleable meat yield equations and all the staff and students of the animal molecular genetics group at the University of Adelaide and at AgResearch for their tireless participation in phenotypic data collection from the QTL mapping animals.

Importantly, this study would not have been the same without the past and present staff and students of the Department of Animal Science and the Department of Molecular

Biosciences, University of Adelaide especially Jan Cook and Helena Kojevnikoff who have allowed me to experience and learn so much more than they will ever know.

Finally, my most heart-felt thanks must go to my dearest friends (Adam, Andrew, Karen, Kathryn, Kirsten and Trudy) and family who have stuck by me through it all! Words can not describe how much your encouragement, love, support and guidance have meant to me. The endless phone calls and e-mail messages, you have made me whom I am today.

Thankyou with all my heart.