Genome-wide copy number variation in sheep: detection and utility as a genetic marker for quantitative traits, with reference to gastrointestinal nematodiasis

Thesis presented in partial fulfilment of
the requirements for the degree of

Doctor of Philosophy
in
Animal science

At Massey University, Palmerston North,
New Zealand

Juncong Yan

2018

Abstract

Gastrointestinal nematodes are perhaps the most important parasites of domestic sheep world-wide. Genetic selection for nematode resistance in domestic sheep is being promoted in many countries including New Zealand. There are several strategies to identify genetic markers associated with quantitative traits. Single nucleotide polymorphism (SNP)-based strategies have been widely used in animal breeding. However, SNP cannot explain all the genetic variation for a particular trait. A new kind of variation, copy number variation (CNV) has been identified as contributing to genetic variation in production and disease traits. Compared with other domestic animals, CNV in sheep is poorly investigated. The primary objective of this thesis was to explore the utility of genome-wide CNV as a genetic marker for the analysis of quantitative traits in sheep. Five different studies were undertaken to fulfill the objective. The first two studies used 50 K SNP BeadChip genotype data and next generation sequencing (NGS) data to detect CNV. Extensive CNV differences were evident between breeds as well as detection algorithms. NGS-based detection resulted in better CNV resolution than that by SNP. Subsequently, a genome-wide association study (with a small sample size) using CNV detected from a high density (HD) SNP genotype data identified four CNV regions to be significantly associated with a couple of traits pertaining to gastrointestinal nematodiasis in Romney sheep, while no significant SNP associations were found. Somatic mosaicism of CNV, influenced by age (high in foetuses, compared to adults), individuals, detection algorithm and type of tissue analysed, was also evident in separate study. The final study detected CNV differences and SNP based selection signatures in two Romney lines selected for gastrointestinal nematode resistance or resilience. Several significant SNPs and line-specific CNV regions were identified. However, only one SNP overlapped to a CNV region, indicating that SNP-based selection signatures and CNV could represent different aspects of sheep immunogenetics. Overall, CNV could be a potential

genetic marker, albeit with methods for detection and validation needing to be refined. The conclusions from this thesis expand our understanding of CNV in sheep and its potential application prospects for genetic breeding of sheep in the future.

Acknowledgements

First and foremost, I would like to acknowledge my supervisors, Dr Rao Dukkipati, Prof Hugh Blair and Associate Prof Patrick Biggs. Rao, you like a father, endlessly supported and encouraged me during my PhD study. You had no complaints about my poor English and tolerated that for four years. Your passion, patience and meticulousness for science let me know what a scientist should be. Hugh, you like a grandfather, did your best to provide funds for my study. I cannot count how many times you have helped me. Your academic brilliance illuminated my academic path like a beacon and you have been my model in heart. Patrick, you opened the gate of bioinformatics for me and show me a total new interesting world.

Furthermore, I would like to acknowledge Prof Dorian Garrick, Dr Keren Dittmer, Dr Sarah Pain, Dr Andrew Greer, Mr Joseph Hamie for original data support and Dr Kristene Gedye, Rosemary Heathcott for technical support. I would also like to acknowledge the overseas laboratory group, Key Laboratory of Genetics Breeding and Reproduction of Grass Feeding Livestock; Key Laboratory of Animal Biotechnology of Xinjiang, Xinjiang Academy of Animal Science, in China. The data you supported is very important for my research. I also thank my all colleagues in room 3.06. Your support and harmonious environment in office has been crucial for my successful completion. Your help let me quickly adapt to New Zealand's life.

I would like to acknowledge research funding from different sources within Massey University as well as the Massey-Lincoln and Agricultural Industry Trust, for this project. Besides, I very much appreciate the financial support from the Massey University Doctoral scholarship, which really helped me a lot and let me focus on study. I also appreciate the financial support from the IVABS postgraduate fund which helped me to attend my first international academic conference in Australia.

I would like to acknowledge NeSi's (New Zealand eScience Infrastructure) support for high performance computing for the NGS data analysis.

Finally, I would like to thank my family. Dad, Mum and my brother, your encouragement is the power for me to climb the mountain of academic success. Your never-ending support has helped me succeed. Thank you for always believing in me.

Preface

I have undertaken this thesis in the form of publishable experimental chapters using a format of thesis by publication. The current status and publication outlet are described in the following list.

Chapter 1: Literature review

Chapter 2: Genome-wide detection of autosomal copy number variants in several sheep breeds using Illumina OvineSNP50 BeadChips.

Juncong Yan, Hugh T. Blair, Mingjun Liu, Wenrong Li, Sangang He, Lei Chen, Keren E. Dittmer, Dorian J. Garrick, Patrick J. Biggs, Venkata S.R. Dukkipati*
Published in Small Ruminant Research, 2017, 155: 24-32.
(doi:10.1016/j.smallrumres.2017.08.022)

All molecular work, data analysis, interpretation of results and manuscript write-up were completed by Juncong Yan. The original SNP data was provided by Mingjun Liu, Wenrong Li, Sangang He, Lei Chen, Keren E. Dittmer and Dorian J. Garrick. The manuscript was checked by supervisors, Venkata S.R. Dukkipati, Hugh T. Blair and Patrick J. Biggs.

Chapter 3: Detection of copy number variation in sheep by whole genome sequencing

Juncong Yan, Hugh T. Blair, Keren E. Dittmer, Patrick J. Biggs, Venkata S.R. Dukkipati
To be submitted to BMC Genomics

All molecular work, data analysis, interpretation of results and manuscript write-up were completed by Juncong Yan. The original NGS data was provided by Keren E. Dittmer. The manuscript was checked by supervisors, Venkata S.R. Dukkipati, Hugh T. Blair and Patrick J. Biggs.

Chapter 4: Genome-wide association study in sheep selectively bred for resistance or resilience to gastrointestinal nematodes

Juncong Yan, Hugh T. Blair, Andrew Greer, Joseph Hamie, Patrick Biggs, Venkata S.R. Dukkipati

To be submitted to Journal of Animal Breeding and Genetics

All molecular work, data analysis, interpretation of results and manuscript write-up were completed by Juncong Yan. The original phenotype data was provided by Andrew Greer and Joseph Hamie. Patrick J. Biggs provided bioinformatics support for gene annotation. The manuscript was checked by supervisors, Venkata S.R. Dukkipati, Hugh T. Blair and Patrick J. Biggs.

Chapter 5: Somatic mosaicism of copy number variation in sheep using Ovine Infinium® HD SNP BeadChip

Juncong Yan, Hugh T. Blair, Patrick J. Biggs, Sarah J. Pain, Venkata S.R. Dukkipati

To be submitted to PLOS ONE

All molecular work, data analysis, interpretation of results and manuscript write-up were completed by Juncong Yan. The original tissue samples were provided by Sarah J. Pain. Patrick J. Biggs provided bioinformatics support for gene annotation. The manuscript was checked by supervisors, Venkata S.R. Dukkipati, Hugh T. Blair and Patrick J. Biggs.

Chapter 6: Detection of copy number variation and genome-wide positive selection signatures using Ovine Infinium® HD SNP BeadChip in two Romney lines, selected for resistance or resilience to gastrointestinal nematodes

Juncong Yan, Hugh T. Blair, Andrew Greer, Joseph Hamie, Patrick J. Biggs, Venkata S.R. Dukkipati

To be submitted to BMC Genomics

Part of this chapter was presented as a poster at the 22nd Association for the Advancement of Animal Breeding and Genetics conference held in Townsville, QLD, Australia, 2-5 Jul 2017.

Juncong Yan, Venkata S.R. Dukkipati, Hugh T. Blair, Patrick Biggs, Joseph Hamie and Andrew Greer (2017). A genome-wide scan of positive selection signature using Ovine Infinium® HD SNP BeadChip in two Romney lines, selected for resistance or resilience to nematodes. In Proceedings of the Association for the Advancement of Animal Breeding and Genetics Vol. 22 (pp.1).

All molecular work, data analysis, interpretation of results and manuscript write-up were completed by Juncong Yan. The original phenotype data was provided by Andrew Greer and Joseph Hamie. The manuscript was checked by supervisors, Venkata S.R. Dukkipati, Hugh T. Blair and Patrick J. Biggs.

Chapter 7: General discussion

Table of Contents

List of Figures

List of Tables

Common abbreviations

aCGH	array comparative genomic hybridization
AFLP	amplified fragment length polymorphism
AMD	age-related macular degeneration
AS	de novo assembly of a genome
BAF	B allele frequency
BLUP	best linear unbiased prediction
BP	biological process
CC	cellular component
CFH	complement factor H gene
CIITA	class II Major Histocompatibility Complex transactivator
CN-LOH	mosaic copy neutral loss of heterozygosity
CNV	copy number variation
CNVR	copy number variation region
CPU	central processing unit
dH_2O	deionised distilled water
DLRS	derivative log ratio spread
DNA	deoxyribonucleic acid
dNTP	deoxy-ribonucleoside triphosphate
dsDNA	double-stranded DNA
EHH	extended Haplotype Homozygosity
ELISA	enzyme-Linked ImmunoSorbent Assay
EMP2	empirical p-value, corrected for all tests
EPG	eggs per gram
FDR	false discovery rate
FEC	faecal egg count
FLK	an extension of Lewontin and Krakauer (LK) test, based on population's kinship (F) matrix
Fst	fixation index
GO	gene ontology
GWAS	genome-wide association study
hapFLK	haplotype structure accounted FLK
HGP	human Genome Project
HIV	human immunodeficiency virus
HMMs	hidden Markov models
IBD	identity by descent
iHH	integrated allele-specific EHH
iHS	integrated haplotype Score
IQRs	inter-quartile range
ISGC	International Sheep Genomics Consortium
KEGG	Kyoto Encyclopedia of Genes and Genomes
LD	linkage disequilibrium

LRR	log R ratio
LW	live weights
MF	molecular function
MHC II	major histocompatibility complex II
MZ	monozygotic twins
NAHR	non-allelic homologous recombinations
NeSi	New Zealand eScience infrastructure
NGS	next generation sequencing
PCA	principal components analysis
PCR	polymerase chain reaction
PEM	paired-end mapping
qPCR	quantitative polymerase chain reaction
Q-Q	quantile-quantile
QTL	quantitative trait loci
RAM	random-access memory
RAPD	random Amplified Polymorphic DNA
RD	read depth
REHH	relative EHH
RFLP	restriction fragment length polymorphism
RNA	ribonucleic acid
Rsb	across Population EHH
SLE	systemic lupus erythematosus
SM	somatic mosaicism
SNP	single nucleotide polymorphism
SR	split read
SRFA	selective restriction fragment amplification
SSRs	simple sequence repeats
SVS	Golden Helix SNP & Variation Suite
TMB	tetramethyl benzidine
VNTR	variable number of tandem repeats
WF	Wave factor
XP-EHH	across Population EHH
ZHp	Z-transformed Heterozygosity Value