

# **Investigation and Application of Methods for Ancient DNA Research**

Stephen Malone Richards

A dissertation submitted to the University of Adelaide  
in accordance with the requirements of  
the degree of PhD  
in the Faculty of Science,  
School of Earth & Environmental Sciences

September 2014

## Dedication

This thesis is dedicated to Ignatius J. Reilly.

## Table of Contents

Dedication	ii
Table of Contents	iii
List of Figures	vii
List of Tables	ix
List of Appendixes	x
Abstract	xi
Thesis Declaration	xiii

### **Chapter I: Introduction**

1.0 History of Ancient DNA	1
1.1 DNA Sequencing and Polymerase Chain Reaction (PCR)	3
1.2 High Throughput Sequencing (HTS)	4
1.3 Microarray Technologies	12
2.0 Ancient Bison: An Ideal Model for Megafaunal Evolution	13
3.0 Outline of Thesis Chapters	16
References	20

### **Chapter II: Optimizing Ancient DNA Sequencing Libraries**

Statement of Authorship	25
Abstract	26
Introduction	26
Methods	29
Sample	30
aDNA Extraction	30
Library Construction and Amplification	31
Generation of TLR8 Probe	32
Hybridization Capture of TLR8	32
Ion Torrent Sequencing	33
Data Analysis	33
Results	34
Library Characteristics	34
Shotgun Reads Mapped to Genome and Mitochondrial References	36
Shotgun and Hybridization Capture Reads Mapped to the TLR8 Reference	37
Uracil-Induced Misincorporation	38
Discussion	38
Library Characteristics	38
Shotgun Reads Mapped to the Cattle Reference Genome	40
Enriched Reads Mapped to the Cattle TLR8 Reference	42
Uracil-Induced Misincorporation	45
Conclusion	45
References	47
Supplemental Methods	66

### **Chapter III: Isothermal Amplification in Hybridization Capture of Ancient DNA**

Statement of Authorship	88
Abstract	89

Introduction	89
Methods	93
Amplification Protocols	94
Sample	95
Extraction of aDNA	95
Library Construction	95
Generation of TLR8 Probe	96
Hybridization Capture of TLR8	97
Ion Torrent Sequencing	97
Data Analysis	98
Results	99
Whole Extract Amplification	99
Characteristics of Filtered Reads from TLR8 Hybridization	
Capture	99
Mapping of Filtered TLR8 Enriched Reads	101
Nucleotide Misincorporation	102
Discussion	102
Library Yields for Whole Amplification	102
Characteristics of Filtered Reads from TLR8 Hybridization	
Capture	104
Mapped TLR8 Unique Reads	106
Nucleotide Misincorporation	107
Future Directions	108
Conclusion	109
References	111
Supplemental Methods	130

#### **Chapter IV: Detection of Altered Bases in Ancient DNA Using SMRT Sequencing**

Statement of Authorship	149
Abstract	150
Introduction	150
Methods	155
Sample	155
Reagents and Materials	156
aDNA Extraction	156
SMRTbell Preparation and Sequencing	157
Data Analysis	157
Results	158
Discussion	160
References	166

#### **Chapter V: Paleoclimatic Impacts on European Bovid Megafauna in the Late Pleistocene**

Statement of Authorship	175
Abstract	177
Introduction	177
Methods	181
Samples	181
Ancient DNA Extraction and Amplification	182

Amplicon Sequencing.....	184
Amplification of Template for <i>In Vitro</i> Transcription ( <i>IVT</i> ) of Probe for Whole Mitochondrial Genome Enrichment.....	184
Transcription of <i>B. taurus</i> Mitochondrial <i>IVT</i> Templates.....	185
Fragmentation of Mitochondrial <i>IVT</i> RNA.....	186
Biotinylation of Fragmented RNA.....	186
Repetitive Sequence Blocking RNA.....	187
<i>Bison X</i> Sequencing Library Construction and Amplification.....	188
Primary Mitochondria Hybridization Capture.....	188
Primary Hybridization Capture Amplification.....	190
Secondary Mitochondria Hybridization Capture.....	190
Secondary Hybridization Capture Amplification.....	190
Sequencing of Enriched Mitochondrial Libraries.....	191
NGS Data Analysis.....	191
Radiocarbon Dating.....	192
Phylogenetic Analysis.....	192
Genetic Identification of the New Specimens.....	193
Estimation of Evolutionary Timescale.....	193
Survey of Temperature and Paleovegetation Record.....	195
Morphological Comparisons.....	195
Results and Discussion.....	196
Ancient DNA Mitochondrial Typing.....	196
Position of New Samples in the Bovid Mitochondrial Phylogeny.....	196
Late Pleistocene Movements of Bison in Europe.....	198
References.....	202
Supplementary Methods.....	216

## **Chapter VI: Elucidating Bovid Evolution with Genotyping Technologies**

Statement of Authorship.....	222
Abstract.....	225
Introduction.....	225
Genotyping Microarrays in Evolutionary Biology.....	225
Illumina BovineSNP50 BeadChip.....	226
Ascertainment bias.....	227
Bison Evolution.....	228
Evolutionary History of American Bison.....	229
Genetic Markers.....	229
Sub-species Distinction of Plains and Woods American Bison.....	230
Modern Genotyping Data.....	231
Genotyping aDNA.....	231
Hybridization Capture.....	234
Methods.....	235
Modern Genotyping Data.....	235
SNP Character Analysis.....	235
Phylogenetic Analysis.....	236
Extraction of aDNA.....	237
BovineSNP50 BeadChip Genotyping: steppe bison aDNA.....	238
Hybridization Capture of SNPs from Steppe Bison aDNA Libraries.....	238
Genomic Mapping of Steppe Bison SNP Enriched Libraries.....	241

Results/Discussion	242
Analysis of Modern Genotyping	242
Modern Bison Genotyping Character Composition	243
Modern Bison Phylogeny	244
Molecular Clock Analysis	246
Genotyping Steppe Bison aDNA with the BovineSNP50 BeadChip	247
Hybridization Capture of SNPs	248
Conclusion	251
References	252

## **Chapter VII: Conclusion**

1.0 Overview	268
2.0 Population-Level Studies	269
2.1 Methodologies for Generating Genome-Wide Data	269
2.2 Advantages and Disadvantages of Genome-wide Methodologies	271
2.3 Examples of Population-Level Studies	271
2.4 Sample Size in Modern DNA and aDNA Population-Level Studies	274
3.0 aDNA Molecular Adaptation Studies	274
3.1 Molecular Adaptation – Gene Expression	275
3.2 Molecular Adaptation – Gene Copy Number Variation	275
3.3 Molecular Adaptation – Mutations in <i>Cis</i> -Acting Elements	277
3.4 Molecular Adaptation – Epigenetic Modifications	279
3.5 Molecular Adaptation – Gene Product Activity	282
4.0 Relevance of Thesis to Population-Level and Molecular Adaptation Studies	284
4.1 Hybridization Capture	284
4.2 Library Fidelity	284
4.3 Identification of Nucleotide Damage	285
4.4 Identification of Epigenetic Modifications	286
4.5 Analytical Tools	286
5.0 Conclusion	287
References	289
Appendix A	295

## List of Figures

### **Chapter I: Introduction**

Figure 1. Illustration of cluster generation for Illumina’s single end sequencing .....	8
Figure 2. Illumina’s sequencing by synthesis using reversible terminator nucleotides .....	9
Figure 3. Cost of DNA Sequencing and Moore’s Law .....	11

### **Chapter II: Optimizing Ancient DNA Sequencing Libraries**

Figure 1. Library construction and amplification .....	50
Figure 2. Boxplots for shotgun filtered read GC content and length .....	51
Figure 3. Boxplot for TLR8 enriched filtered read GC content and length .....	52
Figure 4. Chromosomal distribution of shotgun unique reads .....	53
Figure 5. Length distribution for the unique TLR8 enriched reads .....	54-55
Figure 6. mapDamage plots for shotgun unique reads .....	56
Figure 7. mapDamage plots for captured TLR8 unique reads .....	57
Figure S1. TLR8 RNA probe synthesis .....	58
Figure S2. Restriction enzyme recognition sites .....	59
Figure S3. Examples of qPCR amplification curves .....	60
Figure S4. TLR8 mapped read coverage .....	61

### **Chapter III: Isothermal Amplification in Hybridization Capture of Ancient DNA**

Figure 1. Flow diagrams for the hybridization capture protocols .....	115
Figure 2. Boxplots of filtered read GC content and read length distribution .....	116
Figure 3. TLR8 unique read coverage .....	117
Figure 4. TLR8 unique read length distribution .....	118-119
Figure 5. mapDamage misincorporation plots .....	120
Figure S1. Rolling circle amplification .....	121
Figure S2. Recombinase polymerase amplification .....	122
Figure S3. Agarose gel of RCA product from steppe bison aDNA .....	123
Figure S4. Illustration of the fragmentation of RCA product from aDNA .....	124
Figure S5. Example of TLR8 qPCR amplification curves .....	125

### **Chapter IV: Detection of Altered Bases in Ancient DNA Using SMRT Sequencing**

Figure 1. Images generated with SMRT View showing modified bases detected in the subreads clusters of single SMRTbells .....	169
Figure 2. An image generated in SMRT View showing modified bases detected in the subreads of multiple SMRTbells .....	170

### **Chapter V: Paleoclimatic Impacts on European Bovid Megafauna in the Late Pleistocene**

Figure 1. Phylogenetic tree of control region sequences from 350 bovid samples .....	207
Figure 2. (a) Bovid phylogeny estimated from whole mitochondrial genome sequences .....	208
Figure 2. (b) Allometric scaling of metacarpal measurements between three bison groups .....	208

Figure 3. Geographical origin and chronology of study bison samples .....	209-210
Figure 4. Maximum-clade-credibility tree of <i>Bison X</i> .....	211
Figure S1. Date-randomization test.....	212
Figure S2. Comparison of Nitrogen 15 and Carbon 13 values from the surveyed samples through time .....	215

**Chapter VI: Elucidating Bovid Evolution with Genotyping Technologies**

Figure 1. Reanalyzed phylogenetic tree of 47 cattle breeds .....	257
Figure 2. SNP character composition plots.....	258
Figure 3. Reanalysis of European and American bison genotyping data .....	259
Figure 4. The maximum clade credibility tree estimated using Bayesian analysis of 55 bison and Yak BovineSNP50 genotyping data .....	260
Figure 5. Random phylogenetic placement of steppe bison Bovine SNP50 replicates.....	261
Figure S1. Schematic of probe tiling.....	262
Figure S2. ML phylogenetic tree showing the position of the low quality modern bison samples.....	263
Figure S3. ML phylogenetic tree of European and American bison calculated without heterozygote characters.....	264



## List of Tables

### **Chapter II: Optimizing Ancient DNA Sequencing Libraries**

Table 1. Library characteristics .....	62
Table 2. Unique shotgun reads mapping to cattle and <i>Bison bison</i> references .....	63
Table 3. Mapping of unique reads to the cattle TLR8 gene .....	64
Table S1 Primers and oligonucleotides .....	65

### **Chapter III: Isothermal Amplification in Hybridization Capture of Ancient DNA**

Table 1. Whole extract amplification 1 (WEA1) yields .....	126
Table 2A. Enriched library characteristics .....	127
Table 2B. Characteristics of TLR8 mapped reads .....	127
Table 3. Single nucleotide polymorphism profiles .....	128
Table S1. Oligonucleotides .....	129

### **Chapter IV: Detection of Altered Bases in Ancient DNA Using SMRT Sequencing**

Table 1. Distribution of subread clusters and modified bases .....	171
Table 2. Examples of local sequence context that produced multiple altered bases in steppe bison aDNA .....	172
Table 3. Characteristics of modified bases called in subreads from single SMRTbells .....	173-174

### **Chapter V: Paleoclimatic Impacts on European Bovid Megafauna in the Late Pleistocene**

Table 1. List of all samples from Urals, North Sea, Caucasus and Austria analyzed in this study .....	206
Table S1. List of published mitochondrial control region sequences used for phylogenetic analysis .....	213
Table S2. List of published whole mitochondrial genome sequences used for phylogenetic analysis .....	214
Table S3. Mitochondria control region primers .....	214
Table S4. Oligonucleotides for whole mitochondrial genome hybridization capture .....	221

### **Chapter VI: Elucidating Bovid Evolution with Genotyping Technologies**

Table 1. Steppe bison samples .....	265
Table 2. Read depth coverage of SNPs targeted for enrichment by Hybridization capture .....	266
Table S1. Primers and Oligonucleotides .....	267

List of Appendixes

Appendix A: Additional manuscript produced during candidature.....295

## Abstract

The introduction of high throughput sequencing (HTS) in 2005 caused a revolution in the field of ancient DNA (aDNA). Using the large sequencing capacity of HTS, researchers have overcome the abundant environmental contamination present in most aDNA extractions to reconstruct the genomes of long extinct organisms, such as an archaic horse that perished >500,000 years ago. The proliferation of genomes engendered by HTS has also led to the development of potential ancillary technologies for aDNA research such as genotyping microarrays. In this thesis, HTS and genotyping techniques were developed or refined to improve the application of aDNA to larger biological questions in evolution. This thesis successfully: *a) describes an in-house hybridization capture system that uses RNA probes generated from long-range PCR amplicons, b) demonstrates that recombinase polymerase amplification is a less biased alternative to PCR in hybridization capture of aDNA, c) develops an analytical approach that improves phylogenies generated with data from the Illumina BovineSNP50 BeadChip (a commercially available genotyping microarray).* In contrast, an attempt to determine the identity of modified nucleotides in aDNA with Pacific Bioscience's Single Molecule Real-Time (SMRT) sequencing prove to be unsuccessful and genotyping of ancient bison aDNA with the BovineSNP50 BeadChip generated inconsistent results. Furthermore, a hybridization capture probe design was tested and found to be unsuitable for aDNA enrichment. For the larger biological aspect of this thesis, several of the methods developed were used to study bison, because these animals are ideal models of megafauna evolution. Using the in-house hybridization capture system, whole mitochondrial genomes were enriched from aDNA and used to help identify a new extinct species of bison. Furthermore, the new analytical approach for BovineSNP50 BeadChip data was used to demonstrate a

significant genetic split between American woods and plains bison, which supports separating these animals at least at the subspecies level. This genetic split suggests that woods and plains bison should be conserved as separate species, which has considerable economic and political implications.

## Thesis Declaration

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, 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. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I give consent to this copy of my thesis when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

The author acknowledges that copyright of published works contained within this thesis resides with the copyright holder(s) of those works. I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

Stephen M. Richards

September 13, 2014