

Ancient DNA from pre-Columbian South America

Guido Marcelo Valverde Garnica

Australian Centre for Ancient DNA
Department of Genetics and Evolution
School of Biological Sciences
The University of Adelaide

A thesis submitted for the degree of Doctor of Philosophy at
The University of Adelaide

October 2015

TABLE OF CONTENTS

Thesis abstract	9
Research aims of the project.....	10
Thesis structure.....	11
Thesis declaration	13
Acknowledgements	14
 CHAPTER 1	
General Introduction.....	15
A) Population history of pre-Columbian Americas.....	17
<i>Archaeological evidence.....</i>	18
<i>Linguistic evidence</i>	22
<i>Evidence from cranial morphology</i>	23
<i>Genetic evidence</i>	25
a) Mitochondrial DNA.....	25
b) Human mitochondrial DNA haplogroups.....	27
c) Mitochondrial DNA diversity in the Americas	30
d) Complete mitochondrial data	31
e) Y-chromosome	31
f) Autosomal DNA	33
g) Ancient DNA in the Americas.....	34
h) Hypotheses on the peopling of the Americas based on modern genetic data.....	37
Conclusion from genetic data	43
Final remarks	43
B) Population history of indigenous Americans after the European arrival	44
The Prehispanic “Disease-free-Eden” / Virgin soil Paradigm.....	44
Black Legend Paradigm.....	46
Final remarks	48
C) Ancient DNA: Potentials and pitfalls of ancient DNA analyses	48
The DNA molecule.....	51
DNA degradation.....	52
Ancient DNA characteristics	53
a) DNA fragmentation	53

b) Blocking lesions	53
c) Nucleotide modification	54
d) PCR inhibitors	55
e) Jumping PCR	56
Survival of DNA molecules	56
Contamination of DNA	57
a) Sample collection and handling	57
b) Laboratory Procedures	58
Criteria of authenticity	59
Top 10 criteria of authenticity for DNA for ancient and degraded samples	60
Ancient DNA: Applications and perspectives.....	62
a) Evolutionary and phylogenetic relationships of extinct and extant species.....	62
b) Molecular paleopathology.....	63
c) Hominin evolution	65
d) Human evolutionary history.....	67
Next-Generation Sequencing (NGS).....	71
Applications of Next-Generation Sequencing technologies.....	71
a) de novo whole-genome sequencing	71
b) Re-sequencing or targeted sequencing.....	72
c) DNA Methylation.....	73
d) Metagenomics	73
Ancient DNA and Next-Generation Sequencing	74
Final remarks.....	78
REFERENCES.....	81

CHAPTER 2

Ancient DNA from pre-Columbian populations in the Central Andes of South America: a diachronic study of mtDNA haplogroup diversity based on Hypervariable region I	103
Statement of authorship.....	105
ABSTRACT	110
1. INTRODUCTION	111
2. MATERIALS AND METHODS.....	114
2.1. Sample description and DNA extraction.....	114

2.2. PCR amplification and sequencing of mtDNA Hypervariable	
Region I	115
2.3. Genotyping of mtDNA by Multiplex PCR (GenoCoRe22)	115
2.4. Genotyping of mtDNA by Multiplex PCR (AmericaPlex26)	116
2.5. Sequence analysis	116
2.6. Populations for comparative analyses	116
2.7. Population genetics and statistical analyses	117
2.8. Authentication of aDNA	118
3. RESULTS AND DISCUSSION	118
3.1. Performance of the three typing methods	118
3.2. Contamination detection	119
3.3. Overall success rates of analysed samples	120
3.4. Haplogroup composition and frequency based genetic distances across cultural periods	120
3.5. Genetic comparison of ancient and modern populations	123
3.6. Genetic comparison of ancient populations	123
3.7. Analysis of Molecular Variance (AMOVA)	124
4. CONCLUSION	126
Acknowledgments	126
REFERENCES	127
List of Figures and Tables	133

CHAPTER 3

High-resolution mitochondrial genome sequencing from ancient Pre-Columbian populations reveals new insights into South American genetic diversity

Statement of authorship	163
ABSTRACT	168
1. INTRODUCTION	169
2. MATERIAL AND METHODS	172
2.1. Selection of samples for genomic library preparation	172
2.2. Radiocarbon dating	173
2.3. DNA extraction library preparation and mitochondrial capture (ACAD)	173
2.4. Genomic Library preparation for degraded DNA samples (ACAD)	173
2.5. DNA Extraction and library preparation (UCSC Human Paleogenomic Lab)	175

2.6. Mitochondrial DNA enrichment using biotinylated RNA probes (ACAD)	175
2.7. Hybridization capture for mtDNA (ACAD)	176
2.8. Bioinformatics / DNA sequencing and sequence assembly.....	179
2.9. Sequencing analysis and SNP calling	179
2.10. Population genetics and statistical analyses.....	180
3. RESULTS	181
3.1. Mitochondrial lineages.....	181
3.2. Data sequence quality	181
3.3. Haplogroup composition and haplotype diversity	182
3.4. Genetic distances and population affinities.....	182
3.5. Phylogenetic analysis of mtDNA genomes.....	183
3.6. Haplotype sharing of complete mtDNA genomes	183
4. DISCUSSION	184
Acknowledgments.....	186
REFERENCES.....	187
List of Figures and Tables.....	193
SUPPLEMENTARY INFORMATION.....	209

CHAPTER 4

Ancient DNA analysis from the ‘Huaca Pucllana’ archaeological site in Central coastal Peru: mitochondrial DNA variation in the context of the Wari Empire expansion during the Middle Horizon..... 243

Statement of authorship..... 245

ABSTRACT..... 248

1. INTRODUCTION..... 249

The Wari of the Middle Horizon..... 250

2. MATERIAL AND METHODS

2.1. Sample description and archaeological context

2.2. Radiocarbon dates from Huaca Pucllana samples

2.3. Sample preparation and DNA extraction

2.4. Library preparation and hybridization capture of mtDNA

2.5. DNA sequencing and sequence assembly.....

2.6. Sequence analysis.....

2.7. Population genetic network and statistical analyses

3. RESULTS	259
3.1. Summary statistics and haplotype sharing.....	259
3.2. Mitochondrial DNA haplogroups and sequence coverage	259
3.3. Huaca Pucllana mtDNA diversity in the South American context.....	260
4. DISCUSSION	260
5. CONCLUSIONS	263
Acknowledgments	264
REFERENCES	265
List of Figures and Tables	271
SUPPLEMENTARY INFORMATION	279

CHAPTER 5

AmericaPlex26 – A SNaPshot multiplex system for genotyping the main human mitochondrial founder lineages of the Americas

Coutinho A, Valverde G, Fehren-Schmitz L, Cooper A, Barreto Romero MI, Espinoza, Llamas B, and Haak W. 2014. AmericaPlex26: a SNaPshot multiplex system for genotyping the main human mitochondrial founder lineages of the Americas. PLoS ONE 9(3):e93292.

Statement of authorship	291
-------------------------------	-----

CHAPTER 6

General Discussion.....

Thesis summary	307
Main findings of the thesis	308
Significance and contribution to current knowledge	310
Authenticating aDNA data and contamination control	313
Library preparation, sequencing and ancient DNA damage.....	315
South America: unique and attractive region of the world for aDNA research	320
DNA preservation of samples from South America.....	322
From archaeological excavation to genomic data generation.....	325
Limitations and future directions.....	329
Concluding remarks.....	331
REFERENCES	333

Thesis abstract

Ancient DNA (aDNA) research in the Americas represents a genetic strategy to investigate demographic and historical events of populations in the continent with the added bonus of having a direct and temporal perspective. This thesis aims to explore human mitochondrial DNA (mtDNA) diversity from a large number of pre-Columbian samples in a diachronic transect through time in order to refine our understanding of the genetic structure and diversity of ancient civilizations in the Central Andes of South America prior to the arrival of Europeans.

I used a combination of traditional PCR-based methods and the latest technological advances for DNA analysis (i.e. Next-Generation Sequencing – NGS) to generate high-resolution mtDNA data to explore the past genetic diversity of South American populations.

This work shows the perspective of aDNA research to identify temporal transitions in the genetic composition in the Central Andes of South America in real-time, since I aimed to incorporate samples from all cultural archaeological periods, improving the spatial and geographic coverage.

By comparing the results with genetic data from modern-day native populations, this thesis will also address the potential impact of the European colonization on indigenous populations to understand the evolutionary history of Native Americans. To that end, the acquisition of high-resolution genomic data from ancient specimens showcase the true potential of aDNA research to uncover (re-discover) lost genetic diversity or lost mtDNA lineages from pre-Columbian populations, which cannot be inferred from modern-day populations.

I aim to provide an accurate description of patterns of genetic diversity through time, reconciling and contrasting the genetic data with valuable archaeological information, and to test for demographic and population continuity or discontinuity in pre-contact South America. Finally, this thesis adds important perspective to the existing knowledge about mtDNA diversity and population prehistory in the Central Andes.

Research Aims of the project

- a) To explore novel ancient mitochondrial DNA data from Native Americans from the Central Andes of South America across several archaeological periods in order to contrast this information with available HVR-I data from present-day and ancient populations to better understand the overall pre-Columbian mtDNA genetic diversity.
- b) To generate complete mitochondrial DNA genome data at highest level of resolution from selected ancient samples from South America to characterize and explore the potential of mitochondrial genomes to unveil mtDNA genetic diversity in South America before the European arrival.
- c) To reconcile genetic and archaeological information in the light of temporal sampling to reconstruct the population history of pre-Columbian ancient groups in restricted geographic locations in South America.
- d) To combine the advantages of traditional and novel methods for aDNA analysis in order to develop and establish a new technique for genotyping ancient specimens exclusively from pre-contact Americas.

Thesis structure

This PhD thesis is written as a combination of chapters in publication format and a published paper. The content of the chapters and relevant supplemental materials are as follows:

Chapter 1: General Introduction

- a. General description of the state-of-the art and current knowledge about population history in pre-Columbian Americas.
- b. General introduction to ancient DNA field, applications and limitations.
- c. General introduction to Next-Generation Sequencing technologies and applications in aDNA research.

Chapter 2: Ancient DNA from pre-Columbian populations in the Central Andes of South America: a diachronic study of mtDNA haplogroup diversity based on Hypervariable Region I

Exploration of the mtDNA genetic diversity with a large sample size in a broad geographical and chronological range, contrasting ancient and modern diversity in populations from South America based on analysis of the hypervariable region-I.

Chapter 3: High-resolution mitochondrial genome sequencing reveals a substantial loss of Pre-Columbian diversity

Application of Next-Generation Sequencing techniques to ancient samples from South America in order to retrieve mtDNA complete genomes by applying a temporal sampling through archaeological periods in the Central Andes of South America.

Exploration of mtDNA diversity in pre-Columbian Americas under a great level of genetic resolution.

Chapter 4: Ancient DNA analysis from the ‘Huaca Pucllana’ archaeological site in Central coastal Peru: Chronological study of mitochondrial DNA variation in the context of the Wari Empire expansion during the Middle Horizon

Exploration of changes in genetic diversity and micro-evolutionary processes in a transect through time in a restricted geographic area, i.e. coastal Peruvian populations, driven by the impact of pre-Columbian colonization led by the Wari Empire.

Chapter 5: AmericaPlex26 – A SNaPshot multiplex system for genotyping the main human mitochondrial founder lineages of the Americas (co-author)

Development of a laboratory analytical technique to perform an exhaustive screening of archaeological samples in the Americas for genotyping purposes and sample selection for Next-Generation Sequencing.

Published paper: Coutinho A, **Valverde G**, Fehren-Schmitz L, Cooper A, Barreto Romero MI, Espinoza IF, Llamas B, and Haak W. 2014. AmericaPlex26: a SNaPshot multiplex system for genotyping the main human mitochondrial founder lineages of the Americas. PLoS ONE 9(3):e93292.

Chapter 6: General Discussion

Summary and conclusion of the overall study.

Limitations of research, future perspectives and assessment for potential aDNA follow-up studies in South America.

Thesis declaration

This work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution to Guido Valverde and, to the best of my knowledge and belief, contains no material previously published or written by any other 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 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 (as listed below) resides with the copyright holders(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 catalogue, the Australian Digital Theses Program (ADTP) and also through web research engines, unless permission has been granted by the University to restrict access for a period of time.

The published works of my thesis include:

Coutinho A, Valverde G, Fehren-Schmitz L, Cooper A, Barreto Romero MI, Espinoza IF, Llamas B, and Haak W. 2014. AmericaPlex26: a SNaPshot multiplex system for genotyping the main human mitochondrial founder lineages of the Americas. PLoS ONE 9(3):e93292.

Guido Valverde

October, 2015

Acknowledgements

The present work has been made possible with the financial support of the Australian Research Council funding (ARC), Discovery Project “DP1095782”.

I would like to thank the School of Biological Sciences at the University of Adelaide for supporting my Candidature and the Australian Centre for Ancient DNA – ACAD, which provided the logistics to perform the project.

I would like to show my gratitude and admiration to my thesis supervisors Wolfgang Haak and Bastien Llamas for their understanding and patience guidance and encouragements during this time.

I am grateful to Alan Cooper for giving me the opportunity to work in this fascinating research field. I feel privileged to have been granted access to ancient DNA laboratories at ACAD and perform the laboratory analysis under the most impressive working conditions for ancient DNA research.

I would like to thank Lars Fehren-Schmitz for valuable advice and collaborative work.

I would like to thank Michael Herrera for valuable assistance and collaboration.

I would like to recognize all people and lab members at the Australian Centre for Ancient DNA for their intellectual, technical support and friendship.

I would like to thank the archaeologists at the Huaca Puellana archaeological site in Peru; the Department of Humanities and Archaeology, Pontificia Universidad Católica del Perú; Centro de Investigaciones Arqueológicas del Museo de Ancón, Perú; Unidad de Arqueología y Museos (UDAM), Ministerio de Culturas de Bolivia; Instituto de Alta Investigación (IAI), Laboratorio de Arqueología y Paleoambiente, Universidad de Tarapacá, Arica, Chile, for providing the valuable samples for analysis and for relevant advise and archaeological information.

I would like to address a very special thanks to Ana and Thomas Ryan for making me feel like home and support me during my stay in Adelaide as my Australian family.

I owe my deepest gratitude to my family and my parents Guido and Mery and my brother Alvaro for their constant support and encouragements despite the distance and during the last stage of completion of the thesis.