

Genomic insights into the human population history of Australia and New Guinea



Anders Bergström

Wellcome Trust Sanger Institute

Magdalene College

University of Cambridge

This dissertation is submitted for the degree of Doctor of Philosophy

September 2017

Genomic insights into the human population history of Australia and New Guinea

Anders Bergström, Wellcome Trust Sanger Institute, Magdalene College, University of Cambridge

Abstract

The ancient continent of Sahul, encompassing Australia, New Guinea and Tasmania, contains some of the earliest archaeological evidence for humans outside of Africa, dating back to at least 50 thousand years ago (kya). New Guinea was also one of the sites where humans developed agriculture in the last 10 thousand years. Despite the importance of this part of the world to the history of humanity outside Africa, little is known about the population history of the people living here. In this thesis I present population-genetic studies using whole-genome sequencing and genotype array datasets from more than 500 indigenous individuals from Australia and New Guinea, as well as initial work on large-scale sequencing of other, worldwide, human populations in the Human Genome Diversity Project panel.

Other than recent admixture after European colonization of Australia, and Southeast Asian admixture in the lowlands of New Guinea in the last few millennia, the populations of Sahul appear to have been genetically independent from the rest of the world since their divergence ~ 50 kya. There is no evidence for South Asian gene flow to Australia, as previously suggested, and the highlands of Papua New Guinea (PNG) have remained unaffected by non-New Guinean gene flow until the present day. Despite Sahul being a single connected landmass until ~ 8 kya, different groups across Australia are nearly equally related to Papuans, and vice versa, and the two appear to have separated genetically already ~ 30 kya. In PNG, all highlanders strikingly appear to form a clade relative to lowlanders, and population structure seems to have been reshaped, with major population size increases, on the same timescale as the spread of agriculture. However, present-day genetic differentiation between groups is much stronger in PNG than in other parts of the world that have also transitioned to agriculture, demonstrating that such a lifestyle change does not necessarily lead to genetic homogenization.

The results presented here provide detailed insights into the population history of Sahul, and suggests that its history can serve as an independent source of evidence for understanding human evolutionary trajectories, including the relationships between genetics, lifestyle, languages and culture.

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the introduction of each chapter and/or specified in the text. None of the contents in this dissertation have been submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution. It does not exceed the prescribed word limit for the Biology Degree Committee.

Anders Bergström

September 2017

Acknowledgements

First of all, I would like to thank Chris Tyler-Smith and Yali Xue for their supervision and guidance, and for making it possible for me to work on such exciting projects. Thanks also to the members of the Tyler-Smith group for all the feedback on my work and the stimulating and far-reaching scientific discussions. Thanks to Manjinder Sandhu, Richard Durbin, Shane McCarthy, Deepti Gurdasani and others at the Sanger Institute for fruitful collaborations, and the institute itself, its graduate program and the Wellcome Trust for giving me this opportunity. I would also like to thank the numerous people who supported me and gave me opportunities to learn and do stimulating science before embarking on this PhD, in particular Jonas Warringer and colleagues in Gothenburg and Gianni Liti and his team in Nice.

Secondly, I am grateful to the numerous collaborators who in one way or another have enabled the work described in this thesis, including Stephen Oppenheimer, Alexander Mentzer and colleagues in Oxford and at the Papua New Guinea Institute of Medical Research for our highly fruitful collaboration on Papua New Guinea, John Mitchell and colleagues in Australia, Eske Willerslev, Anna-Sapfo Malaspinas, David Reich, Jaume Bertranpetit, Aylwyn Scally and many others I have interacted with and learnt from during the course of this work.

Lastly, thanks to my friends and family, near and far, for their support and patience over these years. The biggest thanks goes to Giulia, for her endless support and encouragement.

Contents

Chapter 1: Introduction	1
1.1 Outline of this thesis	1
1.2 Basic principles of population genetics	1
1.3 Technology for studying variation in genome sequences	3
1.4 Methods for processing short read sequencing data	6
1.5 Methods for analysing population histories using genetic data	7
1.6 A brief summary of current knowledge on human evolutionary history	9
1.6.1 Our closest hominin relatives	9
1.6.2 Africa	9
1.6.3 Out-of-Africa	10
1.6.4 Archaic admixture	11
1.6.5 Europe	12
1.6.6 East Asia	13
1.6.7 South Asia	13
1.6.8 The Americas	13
1.6.9 The Pacific	14
1.6.10 The effects of lifestyle on population history	14
1.6.11 Adaptation to local environments	15
Chapter 2: The genetic history of Australia	16
2.1 Introduction	16
2.2 Analysis of Aboriginal Australian Y chromosomes	18
2.3 Foreign admixture in Australia	21
2.4 Testing for South Asian gene flow to Australia	22
2.5 Archaic ancestry in Sahul	24
2.6 Out of Africa	25
2.7 Relationship to Eurasians	28
2.8 Relationships to Papua New Guineans	30
2.9 The time of separation between Aboriginal Australians and Papuans	32
2.10 Population structure and its time depth within Australia	34
2.11 Conclusions	34
2.12 Materials and methods	36
Chapter 3: The genetic history of Papua New Guinea	38
3.1 Introduction	38
3.2 Array genotyping of Papua New Guineans	39
3.3 Genetic identity of the HGDP-CEPH Papua New Guinean individuals	40

3.4	Southeast Asian admixture in PNG	40
3.5	The relationship to Aboriginal Australians	42
3.6	Local ancestry inference in PNG	45
3.7	Population structure in PNG	45
3.8	The time depth of population separation in PNG	48
3.9	Population size histories in PNG	50
3.10	Genetic differentiation in PNG	50
3.11	Diversity and isolation in PNG	51
3.12	A model for population history in Holocene PNG	53
3.13	Conclusions	55
3.14	Materials and Methods	57
Chapter 4: Large-scale sequencing of worldwide human populations		59
4.1	Introduction	59
4.2	Sequencing of the HGDP-CEPH panel	60
4.3	Characterization of cell line chromosomal abnormalities	61
4.4	Effects of chromosomal abnormalities on sequencing and genotyping	63
4.5	Variant call properties and filtering strategies	65
4.6	Rare variant sharing patterns	69
4.7	Conclusions	70
4.8	Materials and methods	71
Chapter 5: Future directions		72
5.1	Future directions for the population history of Sahul	72
5.2	Future directions for human population history generally	75
5.3	Concluding remarks	78
Bibliography		81