

Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

Modeling the role of social structures in population genetics

A thesis presented in partial fulfilment of the requirements for the degree of

Doctor of Philosophy
in
Statistics

at Massey University, Manawatu
New Zealand.



MASSEY
UNIVERSITY

Elsa Gratianne Guillot

2011-2015

Par paresse, nous attendons de la science qu'elle réponde à nos questions; les scientifiques eux-mêmes se prêtent parfois au jeu et acceptent d'être présentés comme "ceux qui savent", ceux qui apportent les réponses. Cela est parfois vrai, mais la science est un territoire qui se définit surtout par ses frontières; et, aux frontières de la science, tout est en question.

Albert Jacquard, *Au Péril de la science ?*

Editions du Seuil (1984)

We lazily expect science to answer our questions; scientists themselves sometimes play along and accept to be presented as 'those who know', those who bring the answers. Sometimes it is true, but science is a territory primarily defined by its frontier; and, at the frontiers of science, all is question.

Acknowledgments

First, I would like to thank my supervisor Dr Murray Cox, for taking me on the PhD journey, guiding me along the way and supporting my work. I would also like to thank Prof. Martin Hazelton, my co-supervisor, advising me through my PhD, or, how I learned to stop worrying and love statistics.

I wish to thank the Institute of Fundamental Sciences of Massey University for granting me a scholarship to study a PhD.

Most of this work was done in collaboration with a group of researchers who provided data from Indonesian populations. I would particularly like to thank Prof. Steve Lansing, expert on Indonesian anthropology, and Prof. Herawati Sudoyo, deputy director of the Eijkam Institute both of who I have had the chance to meet. Their contributions to this work is acknowledged in authorship.

I would like to thank my colleagues at the Institute of Fundamental Sciences, including the staff who has been extremely helpful and welcoming. I have had the privilege to exchange with many post-grads, post-doc and researchers, but I would particularly like to thank those who helped create SMARTPOP, a great programming enterprise (Tim, Pablo, Chris). Thanks to members of the computational biology group, the Massey evolution community as well as the statistics group with who I have been working.

I would like to thank friends from France and New Zealand, flatmates, as well as handball, netball and squash teammates who made this journey an enjoyable one. A very special thank to Mark for supporting my endeavour and sharing my passion of science.

Finally, I would like to thank my family, who supported me and stayed close to me, despite the distance.

Merci à tous.

Abstract

Building on a theoretical framework, population genetics has been widely applied to diverse organisms, from bacteria to animals. On humans, this has led to the reconstruction of history, the timing of settlements, and migration between populations. Mostly based on the coalescent theory, modern population genetic studies are challenged by human social structures, which are difficult to incorporate into analytically models. The implications of social structure on population genetics are mostly unknown. This work presents new modeling and inference methods to model the role of social structure in population genetics. The applications of these new techniques permit to gain better understanding of the history and practices of a number of Indonesian island communities.

This thesis comprises three published, organized as sequential chapters. The Introduction describes population genetic models and the statistical tools that are used to make inferences. The second chapter presents the first paper, which measures the change of population size through time on four Indonesian islands structured by history and geography. The third chapter presents SMARTPOP, a new simulation tool to study social structure, including mating systems and genetic diversity. The fourth chapter focuses on Asymmetric Prescriptive Alliance, a famous kinship system linking the migration of women between communities with cousin alliance. The fifth chapter presents a conclusion and future directions. In combination, this body of work shows the importance of including social structure in population genetics and proposes new ways to reconstruct aspects of social history.

Contents

Contents	ix
List of Figures	xi
List of Tables	xiii
1 Introduction	1
1.1 Human population genetics	1
1.2 Population genetics data	4
1.3 Models in population genetics	7
1.3.1 Theoretical models	7
1.3.2 Simulations	9
1.4 Statistical inference	10
1.4.1 Summary statistic methods	11
1.4.2 Bayesian methods	11
1.4.2.1 Markov Chain Monte Carlo	12
1.4.2.2 Approximate Bayesian Computation	13
1.5 Structure of human populations	15
1.5.1 Social structure shapes genetic patterns	16
1.5.2 The genetic footprint of social systems	18
1.5.3 Human mating systems	21
1.6 Rationale of the study	23
2 Population Demography	25
2.1 Preamble	25
2.1 Paper	27

CONTENTS

3 SMARTPOP	47
3.1 Preamble	47
3.2 Upgrades to the software	48
3.2 Paper	50
4 Asymmetric Prescriptive Alliance	67
4.1 Preamble	67
4.1 Manuscript	68
5 Conclusions and Perspectives	97
References	103
6 Appendix	119
6.1 Contributions to publications	119

List of Figures

1.1	Scheme of inheritance of sex-specific genetic material in humans	6
1.2	Map of matrilocality	17
2.1	Map of Indonesian samples	30
2.1	Bayesian skyline plots per island	35
2.1	Distribution of modern effective population sizes and growth rates	37
2.1	Pooled Bayesian skyline plots	38
2.1	Distribution of population sizes peak times and surface area of the Sunda Shelf	39
3.2	Four models showing the range of capability of SMARTPOP.	54
3.2	Time to the most recent common ancestor as a function of female population size.	58
3.2	Pairwise diversity simulated with SMARTPOP and SIMCOAL	59
3.2	Allelic diversity simulated with SMARTPOP and SIMCOAL	60
3.2	Simulated times to equilibrium as function of θ	63
3.2	Comparison of simulated diversity between buffering and sampling schemes	65
4.1	Map of Asymmetric Prescriptive Alliance	68
4.1	Kinship under Asymmetric Prescriptive Alliance.	85
4.1	Genetic diversity (θ_π) under Asymmetric Prescriptive Alliance.	86
4.1	Rate of actual Mother's Brother's Daughter marriage observed.	87
4.1	Posterior distribution of Approximate Bayesian Computation in Rindi.	88
4.1	Folded site frequency spectrum showing ascertainment bias.	91
4.1	Homozygosity under Asymmetric Prescriptive Alliance.	94

LIST OF FIGURES

4.1	Approximate Bayesian Computation cross-validation applied to the Rindi analysis	95
4.1	Approximate Bayesian Computation cross-validation applied to the theoretical Analysis	95
4.1	Flowchart of the SMARTPOP algorithm	96

List of Tables

1.1	Patrilocality: expected genetic diversity patterns	17
1.2	Sex-linked genetic patterns associated with post-marital residence rules	18
1.3	Modified table from Lansing et al. [2011].	19
2.1	Modern effective population sizes and sample sizes	31
3.2	Comparison of summary statistics formulas	61
3.2	Speed gain from buffering	64
4.1	Generalized additive model regression.	83
4.1	Model paramaters.	91