

Department of Philosophy, History, Culture and Art Studies
University of Helsinki
Finland

The past hidden in our genes

Combining archaeological and genetic methodology:
Prehistoric population bottlenecks in Finland

Tarja Sundell

Academic dissertation

To be publicly discussed, by due permission of the Faculty of Arts
at the University of Helsinki in Auditorium XII
on the 10 of May, 2014 at 12 o'clock.

Helsinki 2014

Supervised by

Docent Päivi Onkamo
Department of Biosciences
University of Helsinki
Finland

and

Docent Petri Halinen
Department of Philosophy, History, Culture and Art Studies
University of Helsinki
Finland

Reviewed by

Dr Volker Heyd
Reader in Prehistoric Archaeology
Department of Archaeology & Anthropology
University of Bristol
UK

and

Docent Teppo Varilo
Department of Medical Genetics
University of Helsinki
Finland

Opponent

Dr Volker Heyd
Reader in Prehistoric Archaeology
Department of Archaeology & Anthropology
University of Bristol
UK

ISBN 978-952-10-9866-6 (paperback)

ISBN 978-952-10-9867-3 (PDF)

The artworks of the covers are drawn by Ella Sundell (front) and Emilia Sundell (back).

Unigrafia, Helsinki 2014.

To the women in my maternal line:

**Hilkka
Pirkko
Emilia and Ella
Pirjo and Vilhelmiina
Mirja and Anna**

*Ihmisen ainoa mahdollisuus
on pitää yllä katteetonta
optimismia.*

- Lauri Kerosuo

*The only possibility one has
is to maintain unjustified
optimism.*

- Lauri Kerosuo (translated by the author)

CONTENTS

CONTENTS	4
LIST OF ORIGINAL PUBLICATIONS	6
AUTHOR CONTRIBUTIONS	7
ABBREVIATIONS	8
PREFACE	9
ACKNOWLEDGEMENTS	10
ABSTRACT	12
1 INTRODUCTION	13
1.1 Research history	14
1.1.1 Previous archaeological research on prehistoric population events	14
1.1.2 Previous genetic research	16
1.1.3 Previous and contemporary multidisciplinary research	17
1.2 Genetic marker systems used in this study	18
1.3 Basics of human population genetics	18
1.4 Aims of the dissertation	20
2 MATERIALS AND METHODS	21
2.1 Population genetic simulations (I, III, IV)	21
2.1.1 Simulation tool	21
2.1.2 Simulation components	21
2.1.3 Simulation layout	23
2.1.4 Simulation scenarios	27
2.2 Spatial analyses (II, IV)	28
2.2.1 Bayesian spatial analysis	29
2.2.2 Computational approaches	29

2.3 Data sets for spatial analyses (II, III, IV)	30
2.4 Stone Artefact Database (IV)	30
3 RESULTS	31
3.1 Population genetic simulations (I, III)	31
3.1.1 Publication I: Effects of population bottlenecks on genetic diversity	31
3.1.2 Publication III: Effects of higher female-specific migration rate	35
3.2 Spatial distribution of archaeological finds and their relevance in the evaluation of population changes (II, IV)	37
3.3 Analyses of stone artefacts (IV)	39
3.3.1 Quantitative analysis of the Stone Artefact Database	39
3.3.2 The spatial distribution of stone artefacts	41
4 DISCUSSION	44
4.1 Summary of the results (I, II, III, IV)	44
4.1.1 Simulations	44
4.1.2 Spatial analyses and stone artefacts	45
4.2 Development of the approach used in this thesis	45
4.3 Future prospects	47
4.3.1 Incorporating geographical data	47
4.3.2 Stone artefact analyses	47
4.3.3 Simulation of autosomal markers	47
4.3.4 The coalescent approach	48
4.3.5 Population genetic measures	48
5 CONCLUSIONS	50
REFERENCES	52
APPENDIX 1: GLOSSARY	63

LIST OF ORIGINAL PUBLICATIONS

This thesis is based on the following original publications referred to in the text by their Roman numerals.

- I **Sundell T.**, Heger M., Kammonen J. & Onkamo P. 2010. Modelling a Neolithic Population Bottleneck in Finland: A Genetic Simulation. *Fennoscandia Archaeologica* XXVII: 3-19.
- II Kammonen J., **Sundell T.**, Moltchanova E., Pesonen P., Oinonen M., Heger M., Haimila M. & Onkamo P. 2012. Bayesian spatial analysis of archaeological finds and radiocarbon datings: An example from Finland 4000-3500 cal BC. In Zhou, M., Romanowska, I. Wu, Z., Xu, P. & Verhagen, P. (eds.). *Revive the Past. Proceedings of the 39th Annual Conference of Computer Applications and Quantitative Methods in Archaeology (CAA)*. 318-325.
- III **Sundell T.**, Kammonen J., Heger M., Palo J. & Onkamo P. 2013. Retracing Prehistoric Population Events in Finland Using Simulation. In Earl, G., Sly, T., Chrysanthi, A., Murrieta-Flores, P., Papadopoulos, C., Romanowska, I. & Wheatley, D. (eds.). *Archaeology in the Digital Era. Papers from the 40th Annual Conference of Computer Applications and Quantitative Methods in Archaeology (CAA)*. 93-104.
- IV **Sundell T.**, Kammonen, J., Halinen, P., Pesonen, P. & Onkamo, P. In press. Archaeology, genetics and a population bottleneck in prehistoric Finland. *Antiquity*.

The publications are reproduced with the kind permission of their copyright holders.

AUTHOR CONTRIBUTIONS

	I	II	III	IV
Designed the study	TS,MHe,JK,PO	JK,TS,EM,PO	TS,JK	TS,JK
Provided the material	TS,MHe,JK,PO	JK,PP,MO,MHa	TS,JK,JP	TS,JK,PH,PP
Analyzed the data	TS,MHe,JK,PO	JK,EM	TS,JK,MHe	TS,JK,PH
Wrote the paper	TS,MHe,JK,PO	JK,TS,PP	TS,JK,MHe	TS,JK, PO
Supervized the paper	PO	PO	PO,JP	PO,PH

TS Tarja Sundell
 JK Juhana Kammonen
 MHe Martin Heger
 EM Elena Moltchanova
 PP Petro Pesonen
 MO Markku Oinonen
 MHa Miikka Haimila
 JP Jukka Palo
 PH Petri Halinen
 PO Päivi Onkamo

ABBREVIATIONS

ABC	Approximate Bayesian Computing
aDNA	ancient DNA
AMR	Ancient Monuments Register
BC	before Christ
BEDLAN	Biological Evolution and Diversification of Languages
bp	base pair
BP	before present
BYM	Besag-York-Mollie
CSC	Finnish IT Center for Science
CW	Corded Ware
DNA	deoxyribonucleic acid
FDH	Finnish Disease Heritage
F_{ST}	a measure of genetic distance among subpopulations
GTK	Geological Survey of Finland
HVS	hypervariable segment of mtDNA
INLA	Integrated Nested Laplace Approximations
LD	linkage disequilibrium
M1	Pioneering Stage
M2	Ancylus Mesolithic
M3	Litorina Mesolithic
MCMC	Markov chain Monte Carlo
MRCA	most recent common ancestor
mtDNA	mitochondrial DNA
N1	Early Neolithic
N2	Middle Neolithic
N3	Late Neolithic
NE	North-eastern
N_e	effective population size
PCR	polymerase chain reaction
SAD	Stone Artefact Database
SGU	Geological Survey of Sweden
SW	South-western
TCW	Typical Comb Ware
Y-STR	Y chromosomal short tandem repeat

PREFACE

My first touch with prehistoric people took place in today's Vantaa when my late father took me to see a Stone Age digging site when I was seven years old. I remember being intrigued by the red ochre graves and when walking back did not lift my gaze from my feet in hope of finding sherds of ceramics on the road. Some time after this I was taken through the National History Museum and its prehistoric artefact collections by my godmother, historian Sirkka Kauhanen. These moments were unforgettable and I became fascinated by ancient people and their histories.

Although archaeology has always interested me it was only in my late thirties that I became involved in actual archaeological studies at the University of Helsinki. At the time, I was warned not to quit my day job and turn archaeology, then solely a nice hobby, into real work. As I was starting my history studies at the Open University of Helsinki, still holding on to my permanent job, my history teachers Jari Aalto and Professor Tuomas Heikkilä had me convinced that it was cool to study history. Moreover, I was greatly influenced by certain people who made it easier for me to take the turn in my life; my friend Jaana Aho, who assured me that it was never too late to change the course no matter what. I will never forget the help and support I received from my friend and colleague Tiina Heikkinen when starting our archaeological studies together. Important were the memorable archaeology excursions tended in a motherly way by the late Elvi Linturi.

Besides archaeology I have always been interested in genes. Why go further than our own inheritance to look for traces of our past? When attending an archaeogenetics lecture in 2004 held by Professor Antti Sajantila it suddenly became clear to me how it would be possible to combine the two passions, archaeology and genetics. Inspired by ancient DNA I wrote my first seminar on DNA research, concentrating on Neanderthals. Following the suggestion of Professor Mika Lavento I wrote my second seminar on investigating plausible prehistoric genetic bottlenecks in Finland. This subject inspired me to continue research towards my Master's thesis. While working on my thesis I soon found out that I needed help with genetics. I was advised to turn towards Docent Jukka Palo who kindly initiated my genetics studies by explaining genetic definitions and theoretical background to me. I also got in contact with Docent Päivi Onkamo who patiently answered my numerous questions on genetics and Finns, helping me greatly with my Master's thesis.

In 2006 the three of us, Päivi Onkamo, Jukka Palo and I formed a brainstorming group which was to be the beginning of the research group Argeopop. The multidisciplinary research project Argeopop (<http://www.helsinki.fi/bioscience/argeopop/>) was officially founded in January 2008, when the Helsinki University three-year research grant started. This was followed by a four-year-grant by the Academy of Finland. Due to these grants I was able to start working on my Doctoral Thesis in the project. This was ideal for me since at that time I had small children and knew that my path as an archaeologist would not be that of a field archaeologist's but instead, something more sedentary.

ACKNOWLEDGEMENTS

This thesis was carried out in the Department of Philosophy, History, Culture and Art Studies and the Department of Biosciences at the University of Helsinki, during the years 2008-2013. The funding for this study has been provided by the University of Helsinki Funds, the Academy of Finland, and the Finnish Concordia Fund, and I am very grateful for this. Over the years I have received several travel grants from the University of Helsinki, University of Helsinki Funds and the Finnish Concordia Fund. These grants have enabled me to attend international conferences and get valuable ideas and feedback for my papers.

A number of people have contributed to this work by helping and encouraging me. Without their support, this thesis would not have been possible. First I wish to thank my supervisors Docents Päivi Onkamo and Petri Halinen for all their help and guidance throughout my PhD project. They have provided valuable comments on previous articles and various versions of the manuscript. Thank you for enabling me to work in a challenging multidisciplinary group. The reviewers of this thesis, Doctor Volker Heyd and Docent Teppo Varilo, are gratefully acknowledged for their constructive comments on the manuscript.

Of all my many co-authors and collaborators, I owe most to Juhana Kammonen. One could not wish for better co-operation and friendship. My co-author and friend Martin Heger is thanked for being a true friend and the analytical mind when most needed. My colleague and friend Sanni Översti is thanked for invaluable help and unforgettable paja-days. I have been so lucky to share the same office with all of you.

I am also grateful to my other co-authors and colleagues in the Argeopop-group and at Hjelt Institute for their help, support and company. Special thanks go to Antti Sajantila, Jukka Palo, Petro Pesonen, Anu Neuvonen, Mikko Putkonen, Markku Oinonen and Elena Moltchanova.

I want to express thanks to my colleagues, past and present, at the Department of Philosophy, History, Culture and Art Studies at the University of Helsinki; thanks go especially to Tuija Kirkinen, Anna Wessman, Kristiina Mannermaa, Teija Alenius, Sanna Kivimäki, Mervi Suhonen and Georg Haggrén for support and encouragement. My particular thanks go to Tuovi Laire for all her help and friendly guidance during my years at the Department of Archaeology. I am also grateful for all the aid offered by Mika Lavento. He has always supported my work and is gratefully thanked for being willing to write me various recommendations for scholarship applications.

I owe special thanks to the Soiniemi Project for inspiring me in ways unseen. The Hawaii group, Liisa Ukkola-Vuoti, Sauli Vuoti, Katri Kantojärvi, Martin Heger and Päivi Onkamo are thanked for our unforgettable conference trip to Hawaii and palju-nights afterwards. My friends at Viikki on the 7th floor: Teija Ojala, Matti Kankainen, Jaana Oikkonen, Kerttu Mäkilä, Patrik Koskinen and Petri Törönen are thanked for their refreshing company and conversations over lunches, coffee breaks and picnics.

I wish to thank my dear friends Mervi and Janna for being there. Inka, Kai, Tilli, Kimmo, Sarkka, Jan, Leena, Mats, Minna, Hena, Saija, Mikko, Tara, Mika, Afa and their children are thanked for all the memorable summer cottage days and midsummers together.

Finally, I would like to thank my husband Jokke and our daughters Emilia and Ella for loving and supporting me. Thank you, Jokke, for taking such good care of our children when I was not there. Emilia and Ella, I am so proud of you. Mariliis, thank you for always being willing to help with our children when needed, you are part of the family. I warmly thank my mother Pirkko and my late father Veikko, my sister Mirja with her family Pasi, Anna and Eero, and my late sister Pirjo with her family Hannu and Vilhelmiina for loving and believing in me. Both my father and my little sister Pirjo passed away last year before they could see the end of this project, you are dearly missed. I warmly acknowledge my other relatives, Raija, Tapio, Take, Mika, Tuomas, Marianne and their children for special moments together. I am grateful to my American family for everything they have done for me – my Mom and Dad, Mary Jane and V. Carl Gacono, my sisters Becky and Mary Ann, my brothers Jeff, Kory, Chris and Carl, as well as other family and friends in the USA including my Dutch brother Leon. I have been very fortunate to have a family like you across the sea.

Kauniainen, April 2014

Tarja

ABSTRACT

In this doctoral dissertation both archaeological and genetic methodology was used to investigate the potential existence of prehistoric population bottleneck(s) in Finland. This was performed by applying forward-time population genetic simulations based on archaeological knowledge. Different scenarios were run to evaluate the genetic effects of past demographic events, and the results were compared with the known population genetic features of the current gene pool. The focus in the simulations was on uniparental markers, the mitochondrion and the Y chromosome. The simulations were carried out in the forward time population genetic simulation environment simuPOP.

In addition to genetic simulations, we performed analyses of the Stone and Ceramic Artefact Database as well as the radiocarbon date database to find out temporal differences in the intensity of the archaeological signal. We found similar evidence in these analyses for a marked increase in the archaeological signal 4000-3500 cal BC followed by a distinct weakening. Previously, the Stone Artefact Database has been used to study singular artefacts and a few artefact types exclusively. This is the first time the Database including thousands of objects belonging to several Mesolithic and Neolithic time periods has been data mined to a large extent.

Furthermore, we carried out a Bayesian spatial analysis of radiocarbon datings and archaeological finds from Finland and ceded Karelia, to better understand the overall geographical distribution of human activity through time. The analyses were carried out within the framework of Bayesian spatial modelling, using the Besag-York-Mollie (BYM) model to build spatial distributions of various archaeological datasets in Finland. This model is based on image analysis and assumes similarity of neighbouring areas in geospatial applications. The methodology presented here is one of the first efforts of applying Bayesian spatial analysis with different types of archaeological data in Finland.

The main results of the study can be crystallized as follows: Archaeological and genetic evidence, together with the stone artefact analyses, indicate that there has been at least one Neolithic bottleneck in Finland. Secondly, we show that immigration from neighbouring populations, even if very limited but constant over prolonged time periods, can have drastic effects on a population's genetic composition. Finally, female-specific higher migration rate, compared to a gender-neutral migration rate, brings the simulated genetic diversity closer to the observed contemporary genetic diversity in Finland. Our simulations also showed that a tight prehistoric bottleneck can still have a noticeable effect on genetic diversity even today, after thousands of years.

Keywords: prehistoric population, population simulation, spatial analysis, simuPOP, Bayesian, Stone Artefact Database.

1 INTRODUCTION

Prehistoric demographic changes have been studied by for example investigating summed probability distribution of radiocarbon dates through time (*e.g.* Forsberg 1996; Shennan & Edinborough 2007) as well as evaluating nutritional resources available for hunter-gatherers in the area (*e.g.* Tallavaara & Seppä 2012). Genetic data on modern populations can, to some extent, reveal population histories and provide important insights for investigating the process that shaped their evolution. For example, in a recent paper, 50,000 years of population movements were reconstructed from mitochondrial lineages reflecting the earliest settlers to Neolithic population dispersals (Tumonggor *et al.* 2013). Nevertheless, in this case only maternal lineages were followed. The analysis of modern data can also be biased to mainly express the extreme events, leaving other demographic events undetected.

Ideally, ancient DNA analyses allow one to directly examine genetic variation of ancient individuals and as such to make conclusions concerning the prehistoric populations to which these individuals belonged. A recent paper, Skoglund *et al.* 2012, showed that through analysis of DNA extracted from 5000-year-old ancient Scandinavian human remains, the hunter-gatherers were most similar to that of extant northern Europeans contrasted to a farmer sample from the same time, which was genetically most similar to that of extant southern Europeans.

However, ancient DNA studies face power issues in detecting demographic changes. A realistic estimation of *e.g.* population allele frequencies or size would require at least dozens of individuals to be studied, which is not realistic; the skeletal record of prehistoric populations before sedentary dwelling is sparse. Especially in Finland, due to the destructive acidic soil, adequately preserved organic remains from the Mesolithic and Neolithic yielding DNA are practically non-existent. Thus, we need to assess the prehistoric population events differently. In this thesis we used population genetic simulations based on detailed archaeological knowledge to reach a synthesis. We combine knowledge from two different fields of science to reconstruct prehistoric demographic events: population size estimates based on archaeological data with the most probable scenarios of *e.g.* migration rates acquired from population genetic simulations.

When assessing population size the principal assumption is that the archaeological signal, evidenced by for example summed radiocarbon date distributions, correlates with the population size: the stronger the detected archaeological signal is, the larger the population that left the signal has been. This does not yield absolute values of the population sizes at given times but, instead, estimates the relative population sizes between consecutive time periods.

1.1 Research history

1.1.1 Previous archaeological research on prehistoric population events

The prehistoric population events and settlement history of Finland have been studied by a number of archaeologists *e.g.* Siiriäinen 1981; Meinander 1984; Nuñez 1987, 1997a, 1997b; Huurre 1990, 2001; Lavento 1997, 1998, 2001; Carpelan 1999a-b; Edgren 1999; Halinen 1999a-b, 2005, 2011; Nuñez & Okkonen 1999; Carpelan & Parpola 2001; Mökkönen 2002, 2011; Pesonen 2002, 2005; Takala 2004; Rankama & Kankaanpää 2008; Tallavaara *et al.* 2010. According to these studies cultural influences have reached the area from different source regions in the south, east and north contributing to the composition and structure of people and the culture.

Finland was settled during the Mesolithic by prehistoric hunter-gatherers and is assumed to have been continuously inhabited ever since. The first postglacial pioneers arrived *c.* 11,000 BP (*c.* 8850 cal BC). The earliest human culture in the area is called Suomusjärvi culture, which prevailed until *c.* 5100 cal BC. This distinctive culture was formed when small immigrating population groups, arriving to Finland from various directions, together formed a new culture. The subsistence was based on fishing, hunting and gathering; the manufacture of ceramics was still unknown.

The spread of Typical Comb Ware (TCW) culture into Finland *c.* 4000-3900 cal BC is apparent and brings significant changes both into the quality and quantity of archaeological finds (*e.g.* Vuorinen 1982; Meinander 1984; Carpelan 1999a; Halinen 1999b; Pesonen 2002; Edgren 2007) (Figures 1-2). Whether this implies significant migration or simply a result of cultural continuity has been an enduring question in Finnish archaeology. Nevertheless, the changes in the material culture of the local hunter-gatherers are so remarkable that the possibility of prominent migration cannot be rejected, which we also take into consideration in the empirical part of this work. The climate was at its thermal maximum, which probably contributed positively to the resources available to inhabitation (Tallavaara & Seppä 2012). These conditions would have favoured living conditions and increased the number of people in the area whether the explanation was migration or diffusion.



Figure 1 *Typical Comb Wear sherds from settlement sites in Vantaa, South Finland. Photo by István Bolgár, National Board of Antiquities, 2008.*



Figure 2 *Leaf-shaped flint and slate arrowheads associated with Typical Comb Wear, from settlement sites in Vantaa, South Finland. Photo by István Bolgár, National Board of Antiquities, 2008.*

Additionally, the spread of Corded Ware (CW) culture (*c.* 2800 cal BC) has influenced the repertory of archaeological finds in the country (*e.g.* Carpelan 1999; Edgren 1999). CW is commonly seen as a result of a notable migration in which the southern and western parts of the country receive new immigrants from the south introducing sedentary farming (Halinen 1999a; Cramp *et al.* submitted; Halinen *et al.* submitted). The new material culture also influences the indigenous inhabitants through mutual contacts. In the eastern and northern parts of the country the archaeological picture remains the same, however, the inhabitants of central and northern parts of Ostrobothnia relate to the newcomers by elevating the level of social organization by building large stone structures, *e.g.* cairns and so called “giant’s churches” (Halinen 1999a; Okkonen 2003; Costopoulos *et al.* 2012; Halinen *et al.* submitted).

According to the archaeological signal, there seems to be a clear decline between the numbers of the settlement sites in inland culture, especially in eastern Finland and the Ancient Saimaa area, when turning from the late Neolithic to the Early Metal Period (Lavento 1997, 2001; Saipio 2008) suggesting a population bottleneck (Lavento 2001). The Early Metal Period dwelling sites are usually smaller and spatially differently distributed than in the previous period. The small number of structures found and the smaller sizes of sites could suggest temporary sites, possibly associated with hunter-gathering population and their seasonal settlement model. Despite the uncertainties in defining what constitutes a dwelling site, the difference between the numbers of Stone Age and Early Metal Period sites appears substantial. The above-mentioned facts may thus reflect a genuine decrease in population size, which, if profound enough, is genetically defined as a bottleneck. The archaeological evidence suggests that this may have been especially notable in eastern Finland between the Late Neolithic and the Early Metal Period (Lavento 2001). However, Lavento does not perform further conclusions or detailed genetic interpretations in his dissertation.

In this work we wanted to combine the information acquired from both archaeological and genetic research, on top of the archaeological work that has been done before. Especially, the thesis concentrates on genetic and archaeological information of population bottlenecks, taking into account the known archaeological details.

1.1.2 Previous genetic research

Geneticists have carried out numerous studies concerning the population genetics and the origins of Finns *e.g.* Nevanlinna 1972, Cavalli-Sforza *et al.* 1994; Sajantila *et al.* 1995, 1996; Lahermo *et al.* 1996,1998; Lahermo 1998; De la Chapelle & Wright 1998; De la Chapelle 1999; Kittles *et al.* 1998, 1999; Savontaus & Lahermo 1999; Norio 2000, 2003a-c, 2004; Peltonen *et al.* 2000; Varilo *et al.* 2000, 2003; Kere 2001; Uimari *et al.* 2005, Lappalainen *et al.* 2006, 2008; Salmela *et al.* 2006, 2008; Service *et al.* 2006, Hedman *et al.* 2007; Hannelius *et al.* 2008; Jakkula *et al.* 2008; McEvoy *et al.* 2009; Palo *et al.* 2009. With the development of new genetic methods and tools (PCR in 1980s, high-throughput

genotyping in 2000's *etc.*) it has become possible to study genetic differences between populations in high resolution. Several of the genetic studies mentioned above show that there are significant differences between the western and eastern parts of Finland. Especially, this concerns the Y chromosome distribution which shows substantial differences between the East and West. Moreover, there is a somewhat reduced genetic diversity in the Y chromosome in Eastern Finland.

When compared to west European populations, Finns are considered a genetic outlier (Cavalli-Sforza *et al.* 1994; Lao *et al.* 2008). Our overall genomic variation slightly but significantly differs from other European populations, for example showing a modest shift towards Asians (Salmela *et al.* 2008). Moreover, the overall genomic variation is slightly decreased, and linkage disequilibrium (LD) elevated compared to neighbouring populations (See chapter 4.3.5) (Varilo 1999). Uniquely, the population genetics of Finns is distinguished by the special 'Finnish Disease Heritage', FDH (Norio *et al.* 1973; Norio 2003a-c), the enrichment of rare endemic genetic diseases found in Finland. The FDH consists of 36 monogenic genetic diseases or disorders which are more common among ethnic Finns, especially in East and North, than other populations. In contrast, we almost totally lack some of the genetic diseases widespread elsewhere. There is wide agreement among geneticists, that the existence of FDH as well as above mentioned genomic features are most plausibly explained by population bottlenecks (Sajantila *et al.* 1996; Norio 2003a-c) or enhanced drift, followed by population bottlenecks, during the last 500 years, especially in eastern and northern Finland (Varilo *et al.* 2000, 2003). The direct genetic consequences of population bottlenecks are well known in theoretical population genetics: decrease in genetic diversity, higher inbreeding level as well as elevated LD see *e.g.* Jobling *et al.* 2013). A schematic diagram showing the effects of a bottleneck on genetic diversity is given in Figure 3. Taken all this together, from the genetics point of view it is quite indisputable that a bottleneck or bottlenecks have taken place.

1.1.3 Previous and contemporary multidisciplinary research

Investigating past population histories is a complex task which benefits from the synthesis of different disciplines. In addition to our own project, Argeopop (<http://www.helsinki.fi/bioscience/arageopop>), a few multidisciplinary projects have been launched. The project EUROEVOL 2010-2014 (<http://www.ucl.ac.uk/euroevol>) studies the cultural evolution of Neolithic Europe and the role of farming in transforming early European societies, *c.* 6000-2000 cal BC. The project aims to bring the different sub-fields of cultural evolutionary theory and methods together and create standardised, spatially-referenced datasets (*e.g.*, botanical, faunal, C₁₄ dates, material culture) focusing on broader scale patterns of exchange and influence within western Europe, from the late Mesolithic until the early Bronze Age. Their first major publication deals with regional population collapses in mid-Holocene Europe (Shennan *et al.* 2013).

Another interesting multidisciplinary endeavour is the LeCHE Project (Lactase Persistence in the early Cultural history of Europe), which studies the trait of Lactase Persistence and how this genetic trait appeared and spread in Europe. The project connects archaeologists, zoo-archaeologists, and researchers of genetics and proteomics as well as stable isotopes (<https://sites.google.com/a/palaeome.org/leche/>). The project has, for example, published an article on the evolution of lactase persistence in Europe based on a synthesis of archaeological and genetic evidence (Leonardi *et al.* 2012).

In Finland linguistic and archaeological research together has been used to describe the earliest contacts and formation of the Uralic and Indo-European language families (Carpelan & Parpola 2001; Parpola 2012). Moreover, a new project, BEDLAN (Biological Evolution and Diversification of Languages), combines biological theories and methods for studying language data and uses Uralic languages and Finnish dialects as study objects, contrasting that to *e.g.* paleoclimatic changes. However, projects combining archaeology with genetics have not been undertaken before.

1.2 Genetic marker systems used in this study

The human genome, containing all the genetic information of an individual, is stored within the 23 chromosome pairs. These contain 22 pairs of autosomes and one pair of sex chromosomes (XX in females and XY in males). In addition, mitochondria, energy generating organelles in cells, have their own DNA, mitochondrial DNA (mtDNA). Mitochondria and their genomes are maternally transmitted, meaning that the offspring inherit their mitochondria only from the mother. The Y chromosome is present as a single copy only in males (XY) and is passed on from fathers to sons.

Mitochondrial DNA and the Y chromosome both lack recombination and are thus inherited as whole entities; that is to say they are uniparentally inherited and their composition alters only via mutations. Thus, they enable the study of historic and prehistoric maternal and paternal lineages.

1.3 Basics of human population genetics

Population genetics studies the genetic variation and the evolutionary forces in populations. Human population genetics examines these processes in our own genus *Homo*, most often focusing to extant human populations.

Population genetic analyses are based on allele frequencies (glossary in Appendix 1). The main factors causing changes in allele frequencies are genetic drift, mutation, migration and selection, explained below. This work concentrates on genetic drift, mutation and migration.

Genetic drift

Genetic drift is defined as random fluctuation of allele frequencies in population due to the different contribution of individuals to the next generation of a population. Drift is the main process behind population differentiation and generally spoken leads to loss of genetic diversity. Population bottlenecks and founder effects are special occurrences of genetic drift.

A population bottleneck is an event in which a considerable part of the population is prevented from reproduction. The population undergoes a considerable decrease in size (or the number of reproducing individuals), which may happen as a sudden incident or over a longer time period. Only the survivors will pass their genes on to the next generations, which leads to a reduction of genetic diversity compared to the situation which prevailed before the bottleneck. The longer and more severe the bottleneck is, the more genetic diversity is lost. A founder effect occurs when a small number of individuals from a larger base population colonize a new site and no significant gene flow occurs thereafter between these two. The genetic consequences are very similar to a bottleneck.

Mutation

Mutation is a process which changes DNA sequence. Mutations create genetic variation and result in permanent differences between the ancestral and descendant copies of DNA. Even deleterious recessive mutations may remain in populations, especially if there is strong genetic drift; a text book example of such a process is The Finnish Disease Heritage.

Migration

Migration is a process in which populations and individuals move from one occupied area to another, whereas in colonization the movement occurs to previously unoccupied land. The overall outcome is homogenization of allele frequencies in the populations, but if the populations are small, drift overrides these effects. Migration can change allele frequencies in both the departure and the arrival areas.

Selection

Natural selection refers to the different contribution of different alleles to the next generation. The contribution depends on the allele effects on the survival and reproductive potential of the individuals that carry them. Selection can occur at any life stage between individual's fertilization and generating one's own fertile offspring. As a consequence, alleles favoured by selection are increased in population, whereas harmful alleles are in principle eliminated. Again, drift may run over the effects of selection in small populations.

1.4 Aims of the dissertation

The aim of this dissertation was to find proof and to evaluate the narrowness of prehistoric population bottleneck(s) in Finland. These include:

1. Find out most probable population histories in Finland by performing population genetic simulations with archaeologically justified development of population size (including bottlenecks *etc.*) and compare the results with the present day genetic diversity to evaluate between alternative models (I, III)
2. Carry out Bayesian spatial analyses to distinguish population size fluctuations in space and time (II, IV)
3. Perform analyses of the new Stone Artefact Database to evaluate whether it produces supporting evidence for a Neolithic population bottleneck (IV).

2 MATERIALS AND METHODS

2.1 Population genetic simulations (I, III, IV)

Current genetic diversity provides only indirect evidence about the demographic events of past populations. Instead, population genetic simulation tools can be used to study the effects of population demography on genetic diversity over thousands of years. Simulations are a useful tool in studying population processes unreachable by other means.

2.1.1 Simulation tool

Population genetic simulations can be divided into two categories. Coalescent simulations simulate the population backward in time into the coalescent, that is, the most recent common ancestor from present genetic variation. This is usually implemented in the form of Approximate Bayesian Computing (ABC) (Beaumont 2002; Ray & Excoffier 2009).

With forward simulations one creates virtual populations which are then simulated through generations. The simulated populations are sampled and analyzed and the results compared with real data from present day genetic diversity. SimuPOP (Peng & Kimmel 2005), a forward time population genetics simulation environment, was used in this study. We chose simuPOP as it is the most versatile forward time population simulator at the moment and it allows continuous development of the simulation model. SimuPOP consists of a number of components from which users assemble a suitable simulator. These components are operated through Python script files. The simulations were run in the Murska supercomputer of the Finnish IT Center for Science (CSC).

2.1.2 Simulation components

We have performed two population genetic simulations moving forward in time (Publications I and III) to evaluate possible past demographic events in prehistoric Finland. The simulations started with small pioneer populations, migration from neighbouring populations as well as population bottlenecks were added according to the archaeological signal and finally the population grew exponentially, which we know from historical sources (detailed explanations in chapter 2.1.3). Our first simulation model (I) is further developed in the refined simulation model (III).

Both simulations have similar underlying components:

- The simulation moves forward in ten-year-steps
- Both mtDNA and Y chromosomes are simulated including mutations

- The simulated population is age-structured, that is to say generations overlap
- Reproductive ages are set at 20-60 years for males and 20-40 years for females
- The maximum lifespan is 60 years
- The natural mortality rate is 15% per ten years
- The mating of individuals as well as the selection of those individuals who die is based on random sampling
- Each mating produces 1 to N offspring according to Poisson distribution. Offspring is produced for each pair until it, together with the surviving part of the population, reaches the population size at the next simulation step
- The initial population size is set to grow exponentially up until approx. 5750 BP, after which the population begins to gradually decline over 1600 years towards the bottleneck
- The models include two archaeologically justified bottlenecks, first at 4100-3800 BP and a second, less severe at 1500-1300 BP.
- Each simulation scenario is run 1000 times to obtain enough replicates for evaluation of random variation to the result

The overall model is depicted in Figures 3-4. Differences between the components of the simulations in publications I and III:

Publication I

- Timespan simulated 9000 years
- 15 separate simulation scenarios (Table 1 in chapter 2.1.4 and Publication I)
- Immigration involved in the appearance of TCW and CW, extensive (20% / 9%) and smaller (5% / 2%)
- Constant gene flow from neighbouring metapopulations: archaic European, archaic Scandinavian and Saami
- Two bottlenecks: the first one with three severities, population census sizes 5000, 1000 and 200, the second with 10,000 individuals.

Publication III

- Timespan simulated 11,000 years
- 24 separate simulation scenarios (Table 2 in chapter 2.1.4 and Publication III)
- Immigration involved in the appearance of TCW and CW: moderate (8% / 3%) and small (2% / 0.8%)
- Constant gene flow from neighbouring metapopulations: archaic European, archaic Scandinavian and Saami, reduced to 1/10 compared with the previous simulation
- Two bottlenecks: the first one with two severities, population census sizes 1000 and 200, the second with 10,000 individuals
- Fluctuating start (serial founder effect) vs. stable start
- Finland divided into geographic sub-populations, migration specific division

- Internal gender-specific migration: the migration rate of females is ten times more frequent than that of males.

2.1.3 Simulation layout

Chapter 2.1.3 is based on the refined model (Publication III). The simulations begin at 11,000 BP when the first postglacial pioneers settled the country. The population growth is simulated according to the strength of the archaeological signal (Tallavaara *et al.* 2010). We employ two archaeologically justified bottlenecks: one at 4100-3800 BP and another at 1500-1300 BP. The second, Iron Age bottleneck, has been less severe. The simuPOP package accepts absolute BP years only. Therefore we use BP as the unit of time in all the simulations, contrary to the other datings in this research.

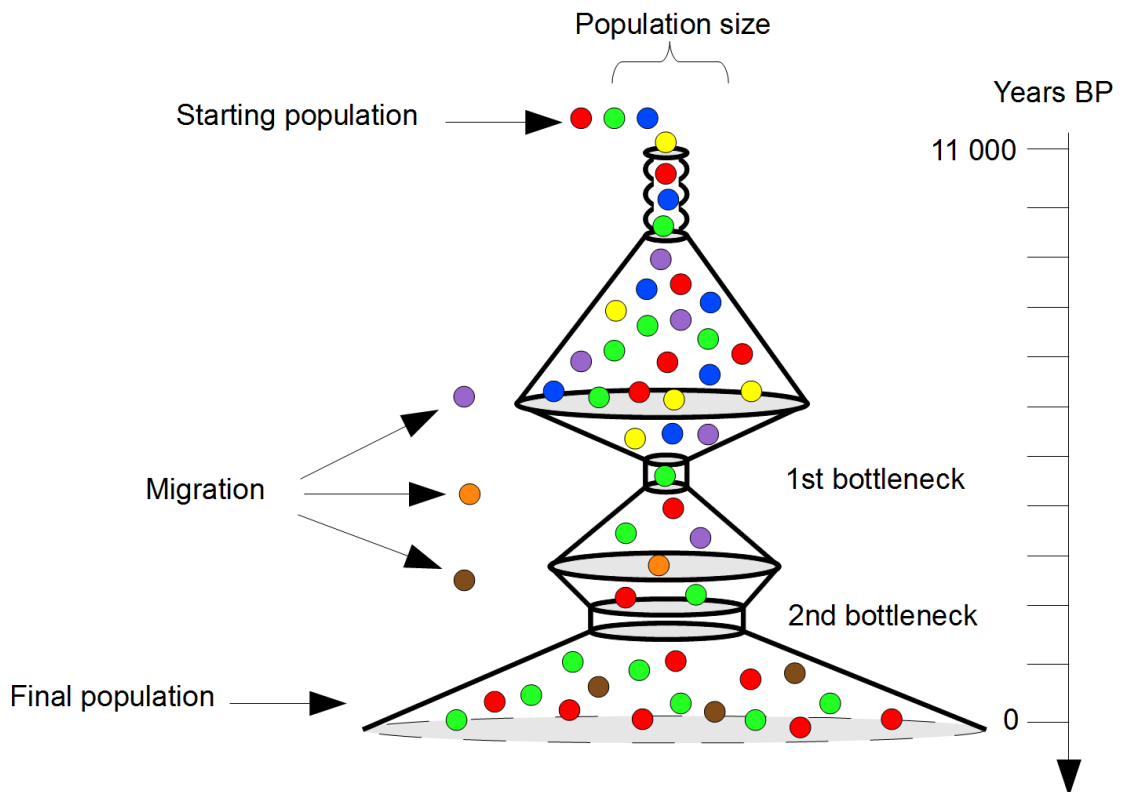


Figure 3 The general demographic model of Finnish population based on archeology, used in our simulations. The width of the cone represents the relative population size (not to scale) at the time before present (BP) depicted on the y axis. The coloured circles depict genetic variants and show what happens when they travel through bottlenecks: only a few variants make it through leaving the final population with reduced diversity.

The simulation starts with a small initial pioneer population of 250 females and 250 males that is divided into two subpopulations: Saami and Other Finland, referring to the North-eastern (NE) subpopulation and the South-western (SW) subpopulation, respectively. This is run for 2000 years with realistic life expectancy, birth and death rates and transmission of genes, to create two natural, highly interrelated small founder populations. We employed two different scenarios for this initial phase: stable starting populations versus a fluctuating start, reminiscent of serial founder effects. In the stable starting population models, the population size remains constant at 500 individuals (Saami 250, Other Finland 250), whereas in the serial founder effect models, the population fluctuates between 240-630 (Saami 120-315, Other Finland 120-315, respectively). Both subpopulations have the same size change rate, with population minima reached *c.* every 200 years. With Saami, here, we refer specifically to genetic Saami whose origins remain partially open.

The total population size is set to grow exponentially from 9000 BP onward until the first bottleneck occurs. Immediately after 7000 BP, and at a size of 15,000 individuals, the Other Finland subpopulation is divided into two: the SW subpopulation and the NE subpopulation. The respective sizes for these subpopulations are 5000 and 10,000. This split event is necessary to later enable internal migration and targeted migration from neighbouring background populations. In simulation scenarios with internal female-specific migration, the SW, NE and Saami populations are allowed to exchange female individuals, but SW and Saami subpopulations never directly exchange male individuals since these subpopulations are geographically too distant from each other.

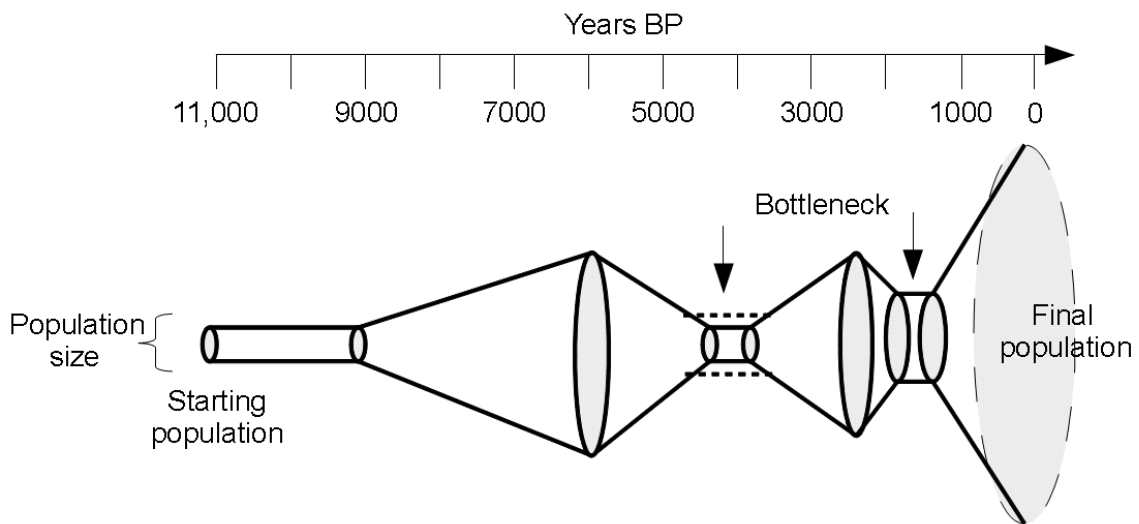


Figure 4 *A simplified demographic model used in simulations. The width of the cone represents the relative population size (not to scale) at the time before present (BP) depicted on the x axis. The dashed line depicts the different bottleneck sizes at the first bottleneck whereas the second bottleneck size remains the same in all simulation scenarios.*

Each individual carries a mitochondrial genome and males also a Y chromosome. These genomes are affected by similar evolutionary forces as nuclear chromosomes. Mitochondrial DNA (mtDNA) is maternally transmitted and simulated as a 631 bp DNA sequence, corresponding to HVS-I and HVS-II of the mtDNA control region. Hypervariable segments (HVS) of mtDNA are the segments that typically are sequenced when mtDNA lineages are studied. Any nucleotide at any position can mutate creating new variation. The mutation rates used in the simulations are based on values published in Heyer *et al.* 1997, 2001; Kayser & Sajantila 2001 and Sigurdardottir *et al.* 2000. In order to make the model even more realistic, the chromosomes were initialized with actual Finnish genetic data including distinct mutations (Hedman *et al.* 2007).

The genetic effects were measured with two basic indicators of genetic diversity: the number of haplotypes (A) and haplotype diversity (\hat{H}) in a sample. The first is simply a direct count of different haplotypes (differing in at least one nucleotide position or microsatellite locus in mtDNA and Y chromosomes, respectively). \hat{H} (Nei 1987) is based on population haplotype frequencies and measures the probability of observing two different haplotypes when sampling two random chromosomes or, as in this case, haploid individuals, from a population. Haplotype diversity is calculated with the formula $\hat{H} = n(1 - \sum x_i^2)/(n-1)$ where N is the number of individuals and x_i the haplotype frequency of the i th haplotype. When \hat{H} is low it is likely that two randomly drawn chromosomes are identical, and vice versa.

We formed three background populations to model external migration waves and minor gene flow into Finnish subpopulations: Archaic European, Archaic Scandinavian and Saami. We selected actual ancient mtDNA haplotypes to form the archaic European and Scandinavian background populations in the simulation (Haak *et al.* 2005; Rudbeck *et al.* 2005; Melchior *et al.* 2007, 2008; Bramanti *et al.* 2009; Malmström *et al.* 2009; Skoglund *et al.* 2012). In the absence of ancient Saami DNA, both the Saami background and subpopulations were initialized with an approximation of present day Saami haplotype composition. Technically, each background population evolves separately for 12,000 years and a snapshot is saved every 2000 years. The snapshot populations are then used as source pools for immigration in the bottleneck simulations. The Archaic European background population includes 50,000 individuals, the Archaic Scandinavian 25,000 individuals and the Background Saami 5000 individuals. The sizes here are suggestive as we merely wish to evaluate if gene flow from other populations would have dramatic effects on the variation of Finnish subpopulations.

We include two migration waves from neighbouring populations, the TCW migration wave at around 6000 BP and the CW migration wave at around 5200 BP. The newest research however suggests that CW migration might actually have taken place later, *c.* 4800 BP (Halinen *et al.* submitted). Due to coherence, we use the older age estimate in both our population genetic simulations to enable direct comparisons between the results.

Two sizes of migration waves were tested (Table 2, scenarios E1-L2): moderate and small. In the moderate migration waves, the Typical Comb Ware migration wave replaces *c.* 8% of the north-eastern subpopulation with Archaic European and the Corded Ware migration wave replaces *c.* 3% of the south-western subpopulation with Archaic European. With smaller migration waves the percentages are divided by four, *i.e.* approx. 2% and 0.8%, respectively. The migration rates were chosen based on our previous work (Publication I). In our previous study (I) we used higher gene flow rates and came to the conclusion that lower gene flow rates should be explored since the high rates forced the simulated populations' diversity to the background populations' which is not true even for the current Finnish population. When we started our population genetic simulations (I), there were no numerical estimates or references to migration or gene flow rates to prehistoric Finland. At that time, we had to base our rates on the prior knowledge we had on the population peak, followed by a decline, and enhanced with the best academic guess (see discussion, chapter 4.3).

In addition to the above-mentioned specific migration waves, moderate constant gene flow replaces 0.01% of the population with Archaic European every 10 years during the entire simulation. Additionally, 0.005% of the Saami and north-eastern subpopulation is replaced with Background Saami gene flow every 10 years and after 3500 BP 0.05% of the south-western subpopulation is replaced with Archaic Scandinavian every 10 years. Lower constant gene flow is one tenth of the above rates, *i.e.* 0.001%, 0.0005% and 0.005%, respectively (Table 2, scenarios E1-L2).

Total population is set to reach a maximum of 25,000 individuals shortly before 5750 BP, to correspond to the known Stone Age population peak (Tallavaara *et al.* 2010). The exponential growth slows down before this and the population size remains approximately the same for a brief period after this time point. After 5750 BP the population gradually begins to decline, which continues over 1600 years towards the Stone Age bottleneck. After the bottleneck the population slowly recovers, followed by a less severe second bottleneck at 1500-1300 BP. During the last 1300 years the population size is set to grow to 1,000,000 individuals, the final population size of the simulations. The current census of Finland is 5,4 million inhabitants, but it was not considered necessary to simulate the final population up to this size as the exponential growth preserves the variation present at the time when the exponential growth began. Also, at the time when this research was begun, it was not technically efficient to simulate populations including millions of individuals. Besides, the census size of Finland has reached one million as late as in the 1820s. Genetically this is a very short time for any mutations to occur in mtDNA and Y-chromosomes.

2.1.4 Simulation scenarios

We carried out 15 different simulation scenarios in our first publication (I) (Table 1) and 24 scenarios in the refined simulation publication (III) (Table 2).

Table 1 *The simulation scenarios (A-O) used in publication I. The converting simulation parameters are bottleneck size at 4100-3800 BP, migration waves from Typical Comb Ware (TCW) and Corded Ware (CW) as well as constant gene flow from neighbouring populations.*

Scenario	Bottleneck size at 4100-3800 BP	Migration waves (TCW and CW)	Constant gene flow
A	5000	-	-
B	1000	-	-
C	200	-	-
D	5000	small	small
E	1000	small	small
F	200	small	small
G	5000	extensive	extensive
H	1000	extensive	extensive
I	200	extensive	extensive
J	5000	extensive	small
K	1000	extensive	small
L	200	extensive	small
M	5000	small	extensive
N	1000	small	extensive
O	200	small	extensive

Table 2 *The simulation scenarios (A1-L2) used in publication III. The converting simulation parameters are population size first 2000 years, bottleneck census size at 4100-3800 BP, internal migration between subpopulations, migration waves from Typical Comb Ware (TCW) and Corded Ware (CW) as well as constant gene flow from neighbouring populations.*

Scenario	Population size first 2000 years	Bottleneck census size at 4100-3800 BP	Internal migration between subpopulations	Migration waves (TCW and CW)	Constant gene flow
A1	stable	1000	no	-	-
A2	stable	200	no	-	-
B1	stable	1000	yes	-	-
B2	stable	200	yes	-	-
C1	fluctuating	1000	no	-	-
C2	fluctuating	200	no	-	-
D1	fluctuating	1000	yes	-	-
D2	fluctuating	200	yes	-	-
E1	stable	1000	no	small	moderate
E2	stable	200	no	small	moderate
F1	stable	1000	yes	small	moderate
F2	stable	200	yes	small	moderate
G1	fluctuating	1000	no	small	moderate
G2	fluctuating	200	no	small	moderate
H1	fluctuating	1000	yes	small	moderate
H2	fluctuating	200	yes	small	moderate
I1	stable	1000	no	moderate	small
I2	stable	200	no	moderate	small
J1	stable	1000	yes	moderate	small
J2	stable	200	yes	moderate	small
K1	fluctuating	1000	no	moderate	small
K2	fluctuating	200	no	moderate	small
L1	fluctuating	1000	yes	moderate	small
L2	fluctuating	200	yes	moderate	small

2.2 Spatial analyses (II, IV)

Spatial analyses refer to a set of techniques for analyzing spatial data. Spatial data may derive from diverse fields, from epidemiological registers to species distribution in ecology, or as in here, archaeological dataset including GIS-related coordinates. The techniques can include different analytical approaches, from random fields to spatial regression and Bayesian hierarchical models. Here we used Bayesian methods for assessing spatial distributions of archaeological artefacts in Finland in different time periods to see how human occupation has altered in time and space.

2.2.1 Bayesian spatial analysis

In Bayesian statistics the evidence about the true state of the world is expressed in terms of degrees of belief, Bayesian probabilities. Bayesian inference is a process of learning from data. A special feature of Bayesian way of thinking is taking into account the knowledge we have available before observing the data: the prior information (in the form of prior probability). The results obtained after observing the data and performing the analysis is called posterior information. The words “prior” and “posterior” are relative to the data under consideration. This can be summed as “Today’s posterior is tomorrow’s prior.” The use of prior information is very natural especially in archaeology – prior information often already exists or is to be found in new research and excavations.

The analyses were carried out within the framework of the Besag-York-Mollie (BYM) spatial model, based on the Bayesian hierarchical methodology for small area analysis (Besag *et al.* 1991). In the model the presence of archaeological find(s) in a cell results in a spatial signal in the cell and, additionally, assumes similarity of neighbouring regions. The BYM model, originally conceived as an application in image analysis has been developed into a tool for disease mapping (BYM 1991). Its use in archaeology is a novelty. Since the model easily allows for utilizing different types of data, the BYM-model fits the general context of this study well; the presence or lack of an archaeological find in a cell results correspondingly in a spatial signal in the cell (Figures 13a-c).

Finland and ceded Karelia were first divided into a grid of 10km to 10km square cells, resulting in a total of 3997 cells. Each cell was given an integer value zero if there were no Mesolithic or Neolithic stone artefacts in this cell and one if there were one or more archaeological finds. The model assumes that neighboring cells are more alike than cells located farther away. Finally, posterior means for the probability of making at least one find in a cell were plotted on a map of Finland and ceded Karelia. Besides showing highlighted areas of prehistoric occupation – or one might say – of human activity in the area, the sequence of visualizations from different time intervals outlines patterns for the overall temporal development of the settlement history.

2.2.2 Computational approaches

Previously we have used the Markov chain Monte Carlo (MCMC) approach as the computational method of estimating posterior probability distributions in the Bayesian spatial analyses (II). More recently, it has been suggested that posterior distributions can be approximated efficiently using Integrated Nested Laplace Approximations (INLA) (Simpson *et al.* 2011). The method was applied by using specialized software (R-INLA). As part of the methodical development (Figure 19), we compared R-INLA and MCMC methods (Kammonen *et al.* 2013). The INLA approach proved computationally faster and less memory consuming. Thus, we decided to apply it in our following publication (IV).

2.3 Data sets for spatial analyses (II, III, IV)

Typologically dated archaeological finds, with or without radiocarbon (^{14}C) datings, can be used as proxies for ancient human activity and occupation. Radiocarbon dates of course gives more accurate information of the period when artefacts and features were deposited.

In publication II we concentrate on the period of 4000-3500 cal BC, as it represents the most prominent era of the TCW ceramics in the prehistory of Finland. Additionally, the population reached a peak at that time (Tallavaara et al. 2010). We used three types of data falling into the time period of 4000-3500 cal BC associated with Typical Comb Ware (TCW): 1) 187 radiocarbon datings, 2) 676 sites with TCW ceramics and 3) 347 finds of leaf-shaped arrowheads. The latter two classes are typologically dated to the same period. Several datings, ceramics and arrowheads derive from the same find context, yielding 728 separate locations in total. The radiocarbon datings used in this study belong to the database which has been collected and augmented by the Laboratory of Chronology (Finnish Museum of Natural History, University of Helsinki). The dataset has been extended to cover the other published archaeological radiocarbon dates from eastern Fennoscandian territory measured elsewhere. In addition, the data contains also those unpublished dates that have been kindly released for our use by the Laboratory of Chronology customers (Oinonen *et al.* 2010).

2.4 Stone Artefact Database (IV)

In publication IV we inventory the new Stone Artefact Database (SAD) for further analysis of assumed Neolithic demographic patterns. SAD is a major subset of the Ancient Monuments Register (AMR), in which Finland's archaeological heritage is documented. AMR is maintained by the National Board of Antiquities, and includes up to two million artefacts altogether. The database is being updated by the Argeopop project (<http://www.helsinki.fi/bioscience/argeopop/>).

In total, we utilize 7506 typologically defined stone artefacts in the database divided into 160 distinct artefact types. This is the first time the Stone Artefact Database has been data mined and disclosed to large extent including thousands of objects belonging to several Mesolithic and Neolithic time periods. Previously it has been used to study singular artefacts and artefact types exclusively.

We categorized single artefacts into artefact types characteristic to the time periods: M1= Pioneering Stage (8850-8000 BC), M2= Ancylyus Mesolithic (8000-6800 BC), M3= Litorina Mesolithic (6800-5100 BC), N1= Early Neolithic (5100-4000 BC), N2= Middle Neolithic (4000-2800 BC) and N3= Late Neolithic (2800-1900/1800 BC). Next, the numbers of artefacts were calculated and spatial analyses performed to find out temporal differences in the intensity of the archaeological signal.

3 RESULTS

3.1 Population genetic simulations (I, III)

3.1.1 Publication I: Effects of population bottlenecks on genetic diversity

Our simulations show that a severe prehistoric bottleneck can affect current genetic diversity even today after thousands of years and even in the presence of moderate immigration from other populations (Figure 6 in Publication *I*, Figure 4 in Publication *III*). The difference of the effect of bottleneck narrowness is visualized in Figure 5. Especially the most severe bottleneck (simulation model C, bottleneck size 200 individuals) drastically reduces genetic diversity. This is, obviously, in compliance with population genetic principles. The effects of migration, overlaid on the bottleneck models, can be seen by comparing Figure 5 to Figures 6 and 7 (note the different scale in Y axis).

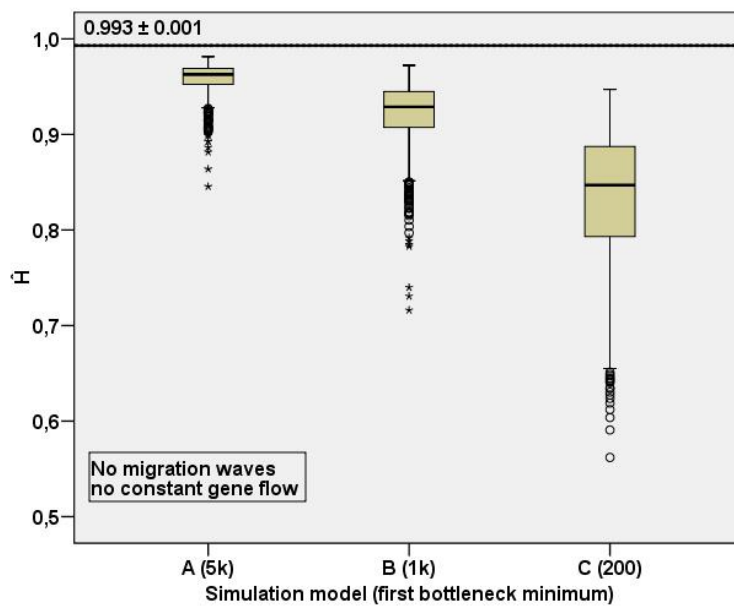


Figure 5 The mtDNA haplotype diversity in the final population with scenarios A-C with 1000 simulation replicates, each with a sample of 832 individuals. The number of individuals in a sample corresponds to the sample size in Palo et al. 2009 enabling direct comparison between the simulations and the real world. The first bottleneck minimum is 5000 (A), 1000 (B) and 200 (C) individuals, respectively. There are neither migration waves nor continuous gene flow to the population, that is, the population is completely isolated. Also note the greater variance of \hat{H} with the most severe bottleneck scenario (C). The simulation result datasets were visualized with IBM SPSS's PASW Statistics 18, boxplot utility (SPSS Inc. 2009). The boxes contain 50% of the observed values. The vertical lines that end in a horizontal stroke, whiskers, give information about the spread of the data: approximately 95% of the data lie between the inner fences. The stroke in the middle implies the median of the observations.

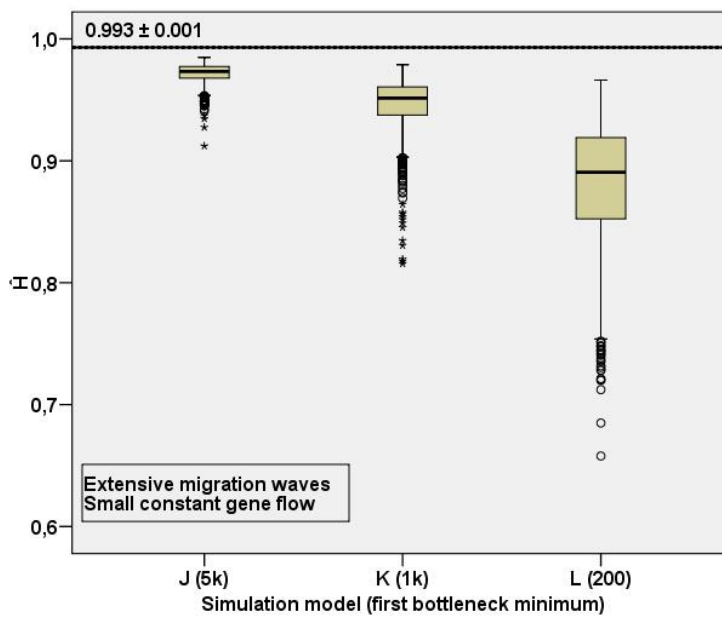


Figure 6 The mtDNA haplotype diversity in the final population with scenarios J-L with 1000 simulation replicates, each with a sample of 832 individuals. The first bottleneck minimum is 5000 (J), 1000 (K) and 200 (L) individuals, respectively. Migration waves are extensive and constant gene flow is small.

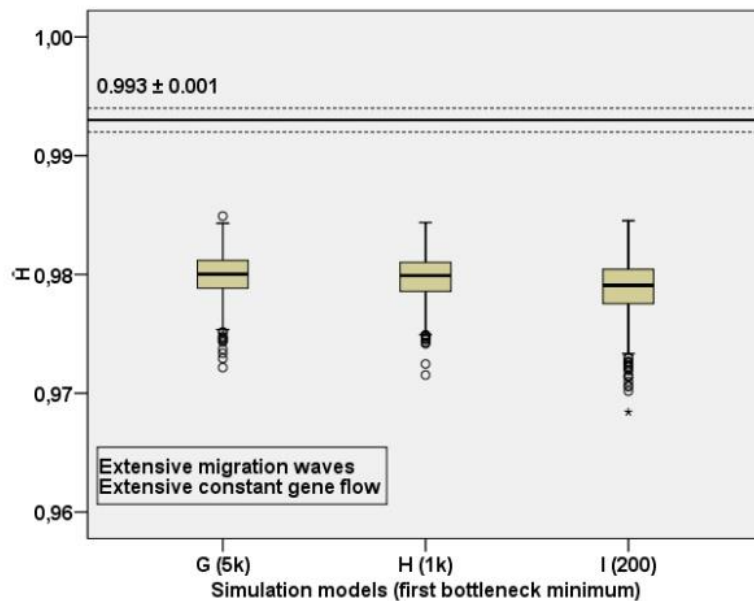


Figure 7 *The mtDNA haplotype diversity in the final population with scenarios G-I with 1000 simulation replicates, each with a sample of 832 individuals. The first bottleneck minimum is 5000 (G), 1000 (H) and 200 (I) individuals, respectively. Both migration waves and constant gene flow are extensive. When the gene flow is extensive all the bottleneck sizes produce similar genetic diversity in the final population.*

The chronologically later bottleneck (1500-1300 BP, size 10,000 individuals) obviously has a much smaller effect on genetic diversity. The overall development of the number and diversity of mitochondrial and Y-STR over time with models A-C are visualized in Figures 7-10 in Publication I). Somewhat surprisingly, a constant small gene flow seems to be a much more important factor than few larger migration waves. While the migration waves have barely any effect, moderate constant gene flow over millennia can cause clear differences to genetic diversity (I, III). Extensive constant gene flow (simulation models G-I, Figure 7 and M-O, Figure 8) also completely eclipses the effects of bottleneck severity in the mitochondrial data, in other words the simulated populations appear nearly identical in the final generation independent of the first bottleneck size (5000, 1000 and 200 individuals). Also, migration waves have barely any effect if constant gene flow is extensive, at any setting. A comparison of all migration wave and constant gene flow variants with the bottleneck size of 1000 individuals is shown in Figure 9.

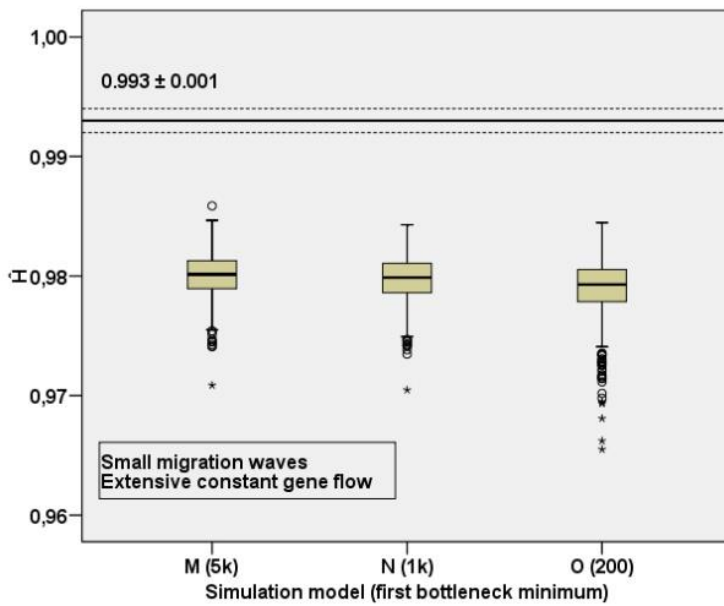


Figure 8 The mtDNA haplotype diversity in the final population with scenarios M-O with 1000 simulation replicates, each with a sample of 832 individuals. The first bottleneck minimum is 5000 (M), 1000 (N) and 200 (O) individuals, respectively. Migration waves are small and constant gene flow is extensive. When the gene flow is extensive all the bottleneck sizes produce a similar genetic diversity in the final population (see Figure 7).

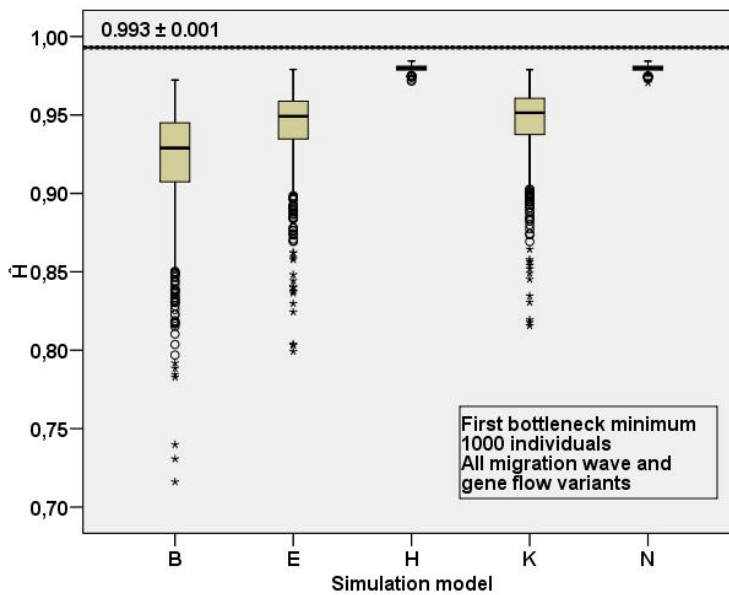


Figure 9 The mtDNA haplotype diversity in the final population with scenarios B, E, H, K and N with 1000 simulation replicates, each with a sample of 832 individuals. The first bottleneck minimum is set at 1000 individuals and all migration wave and constant gene flow variants are included (see Figure 4 in Publication I). Scenarios where the gene flow is extensive produce a genetic diversity closer to the value observed in the present-day population (horizontal line).

3.1.2 Publication III: Effects of higher female-specific migration rate

Given that the observed mtDNA diversity is higher than Y-specific, it has been postulated that in the past, females have moved more than men, as they do in patrilocal systems. Therefore, we wanted to test the effect of elevated female-specific migration in the simulation context. Overall, it brings the simulated mitochondrial genetic diversity on the subpopulation level closer to the observed contemporary genetic diversity in Finland (Figure 10), while keeping the Y chromosomal variation at the lower level.

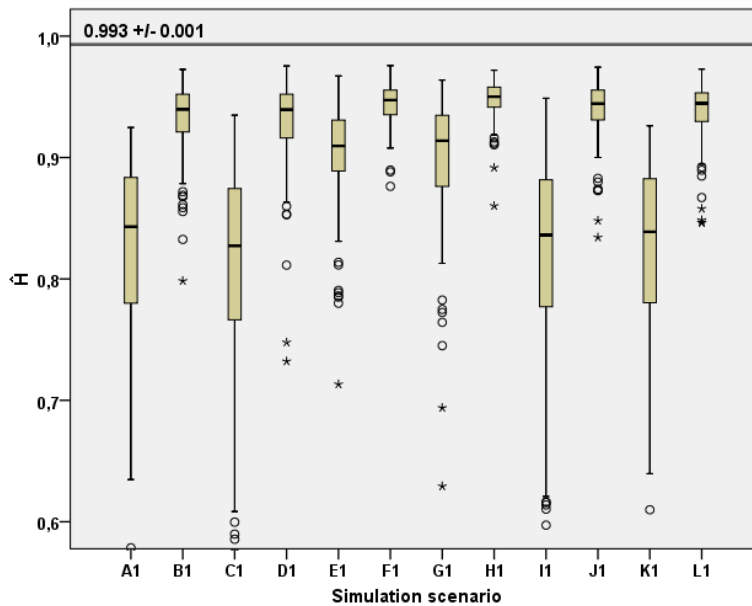


Figure 10 Mitochondrial haplotype diversity (\hat{H}) in the final population (0 BP) with higher migration rates for females (scenarios B1, D1, F1, H1, J1 and L1). The bottleneck size was 1000 individuals. All migration and gene flow variations are included (see Chapter 2.1.4, Table 2).

Second, our results indicate that simulation models beginning with serial founder effects reduce genetic diversity at the first checkpoint after the initial phase, as it should (Figure 4 in Publication III).

Finally, the simulation scenarios with moderate constant gene flow produce Y-chromosomal diversity measures similar to those observed in present-day Finnish population (Figures 11 and 12 below).

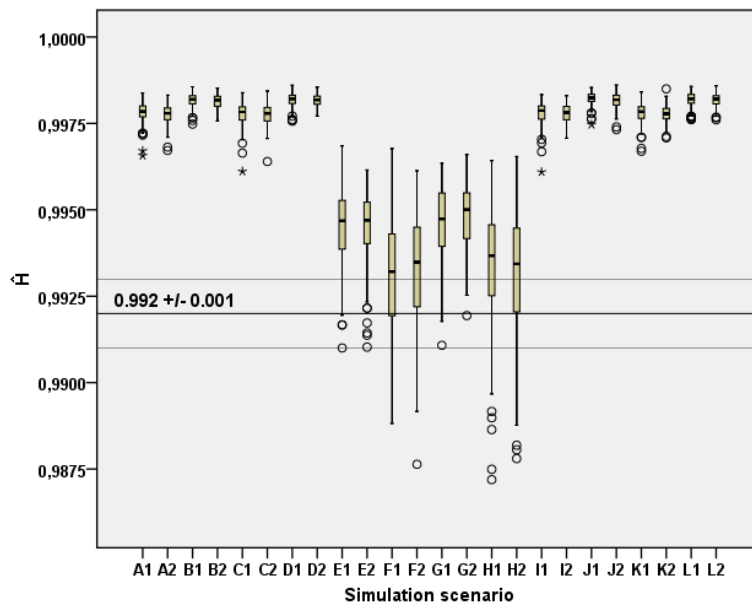


Figure 11 *Y-STR haplotype diversity (\hat{H}) in final population (0 BP) in samples of 907 individuals per scenario. The horizontal reference line represents the value observed in the present (Palo et al. 2009). The details of simulation scenarios A1-L2 are explained in Table 2, chapter 2.1.4.*

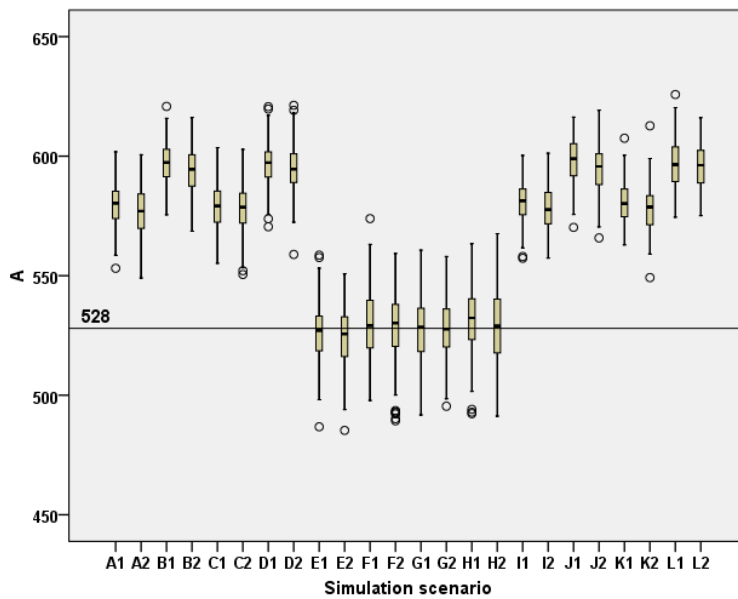
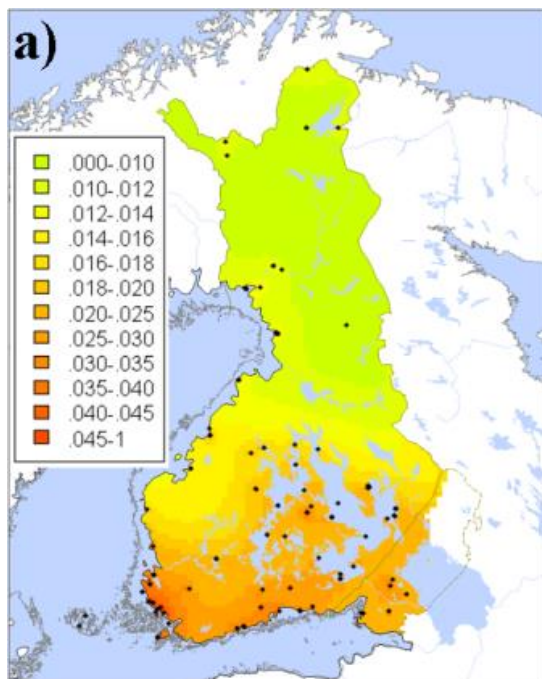


Figure 12 *Number of Y-STR haplotypes (A) in final population (0 BP) in samples of 907 individuals from each of 1000 replicates per scenario. The reference line represents the value observed at the present (Palo et al. 2009). Scenarios E1-H2 shows exactly the same number of haplotypes as observed today in Palo et al. 2009. The advanced simulation model (Publication III) brings the differences in Y chromosomal diversity clearly visible.*

3.2 Spatial distribution of archaeological finds and their relevance in the evaluation of population changes (II, IV)

The first set of the maps (Figures 13a-c, Publication II) concern the time periods of 4000-3500 cal BC, the time of the population peak. Mostly they comply with pre-existing archaeological understanding. Geographically, most of the radiocarbon datings are concentrated in southwestern coastal areas and in the Saimaa Lake district, as well as the Kemi region in the North (Figure 13a). Additionally, the TCW finds and leaf-shaped arrowhead finds (Figures 13b and 13c) show a clear signal in the Kainuu region in the East. Concerning the TCW ceramics, the posterior distribution (Figure 13b) can be interpreted to correspond with the diffusion of the ceramics from the South-East, where the origins of this type are found in the Valdai region and along the upper reaches of River Volga in Russia (*e.g.* Carpelan 1999). This comb- and pit-decorated ceramics style along with many new material and cultural manifestations eventually spread as far as the Arctic Circle. However, it did not reach northern Lapland where the geographical distribution of radiocarbon dates may be indicative of an indigenous population in the area (Figure 13a). Leaf-shaped arrowheads also highlight the southern coastal areas and the Saimaa Lake district (Figure 13c).



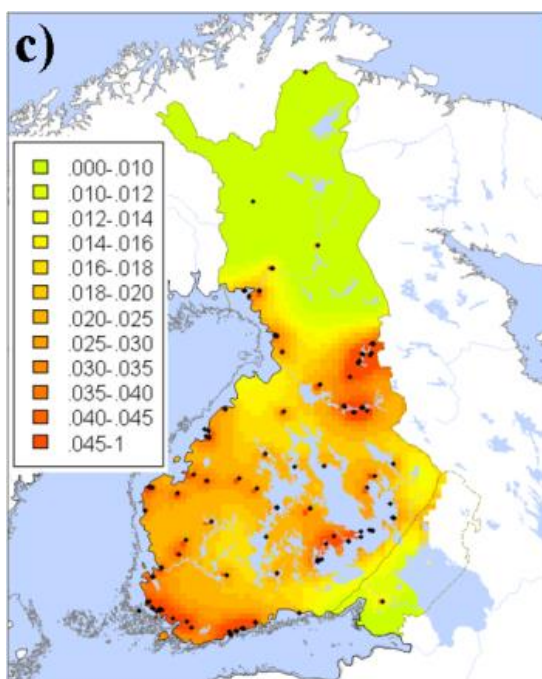
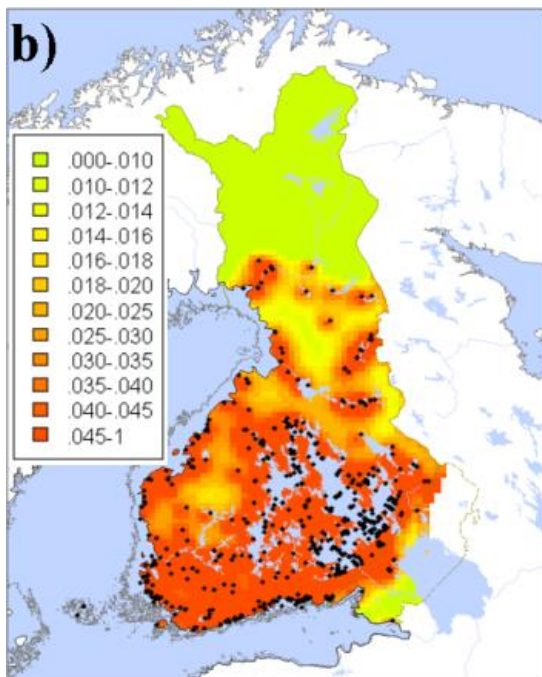


Figure 13 Posterior density (green-yellow-red color scale) of **a)** radiocarbon dates ($N=187$), **b)** Typical Combed Ware (TCW) ceramics ($N=676$) and **c)** leaf-shaped arrowheads ($N=347$) from the time period of 4000-3500 cal BC. Actual find locations are represented by black diamonds. The shoreline of the central lakes area corresponds to that of c. 4000 cal BC estimated by Jouko Vanne / Geological Survey of Finland (GTK), whereas the shoreline of the Baltic Sea corresponds to that of 3500 cal BC estimated by Johan Daniels / Geological Survey of Sweden (SGU).

To find out whether our spatial model is biased by the size of areal units we also ran a test with a larger grid cell size, 20km x 20km, which increases the area of each cell by fourfold. The loss of detailed resolution is inevitable (see Figure 7 in Publication II). Despite the roughness entailed by the larger grid cells, the overall results remain the same reinforcing our assumption that our model is not biased by the size of the areal units (see Publication II for details).

3.3 Analyses of stone artefacts (IV)

3.3.1 Quantitative analysis of the Stone Artefact Database

Altogether the Stone Artefact Database consists of 7506 stone artefacts divided into 160 distinct artefact types. The artefacts in the Database currently lack precise dating (other than being associated with the Stone Age). Therefore we had to manually determine a time period for each of the stone artefact type.

First, we categorized the artefacts to six time periods: Early, Middle and Late Mesolithic (M1, M2, M3) as well as Early, Middle and Late Neolithic (N1, N2, N3) according to the archaeological knowledge of the artefacts' typology. We then calculated two summary statistics: *i*) the absolute number of stone artefacts falling to each time period and *ii*) the number of different artefact types associated with each time period.

The counts of the stone artefacts and stone artefact types through the Mesolithic and Neolithic periods are depicted in Figures 14 and 15. The numbers of stone artefacts and stone artefact types show clear differences when turning from the Middle Neolithic (N2) to the Late Neolithic (N3). There is a distinct peak in the number of stone artefacts from N2 period followed by a clear reduction to the N3 period (Figure 14). Furthermore, there is an apparent rise in the number of stone artefact types from N1 until the beginning of the N3 period (Figure 15). Interestingly, the decline in the artefact numbers is greater than the decline in the artefact types. This is intuitively logical as new innovative types should persist in case of a population continuation.

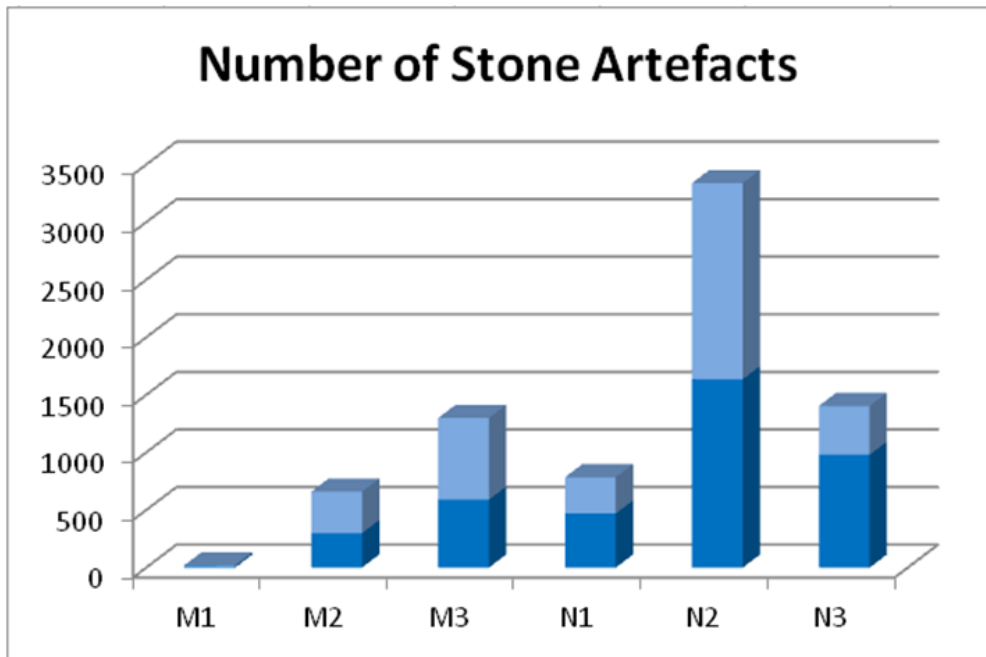


Figure 14 *The number of stone artefacts found in Mesolithic and Neolithic archaeological contexts. The proportions of typologically long-lasting artefacts are depicted by the lighter shade of colour in the columns.*

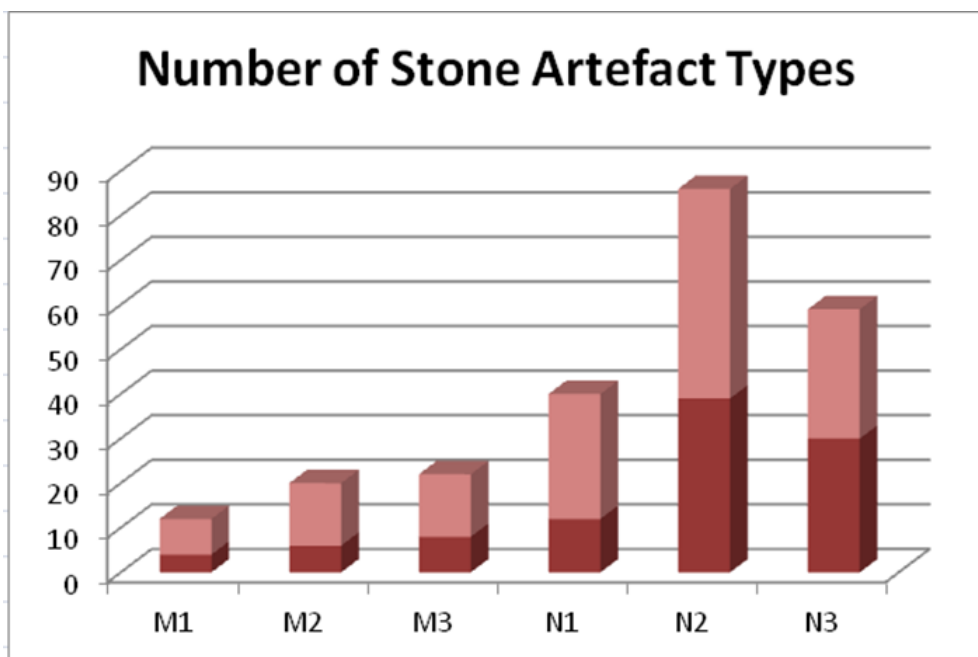


Figure 15 *The number of stone artefact types found in Mesolithic and Neolithic archaeological contexts. The proportions of typologically long-lasting artefacts types are depicted by the lighter shade of colour in the columns.*

We next performed Bayesian spatial analysis of the artefacts in order to get an overall picture of the distribution of the artefacts in the area of Finland. Here, we concentrate only on Neolithic artefacts, as this is closer to the assumed bottleneck period and secondly, the Mesolithic data was too sparse to show any differences in the spatial analyses.

A total of 5237 out of 5528 Neolithic artefacts had location information (longitude and latitude) associated with them and thus could be utilized for the spatial analyses. There are a number of artefacts that cannot be explicitly categorized as belonging to a specific Mesolithic or Neolithic sub period. These typologically more long-lasting artefacts associated with multiple time periods were divided to the three Mesolithic and Neolithic periods in proportion to the total number of stone artefacts known to belong to each time period in question.

3.3.2 The spatial distribution of stone artefacts

The spatial distributions of stone artefacts are visualized in Figures 16-18. The map of the N2 period (Figure 17) shows the densest distribution of artefacts across Finland. There is an apparent change when turning from the N2 period to the N3 period (Figure 18). The south-western coastal area shows higher artefact intensity than the two earlier periods. In contrast, the intensity lowers in many northern, eastern and inland areas.

Besides showing highlighted areas of plausible prehistoric occupation, the sequence of visualizations outline patterns for the overall temporal development of the settlement history. There is an overall rise in posterior probabilities across the area of Finland when moving from N1 period to N2 period. Moreover, the posterior probabilities decrease in specific areas in eastern Finland when proceeding from N2 period to the later N3 period.

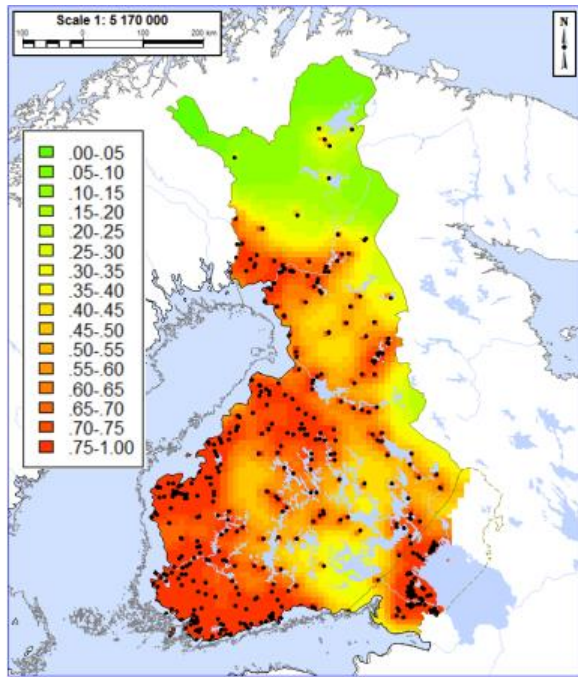


Figure 16 Intensity (posterior density) of stone artefacts from the Early Neolithic (N1-period). The shoreline of the Baltic Sea corresponds to that of 4000 cal BC estimated by Johan Daniels/Geological Survey of Sweden, SGU.

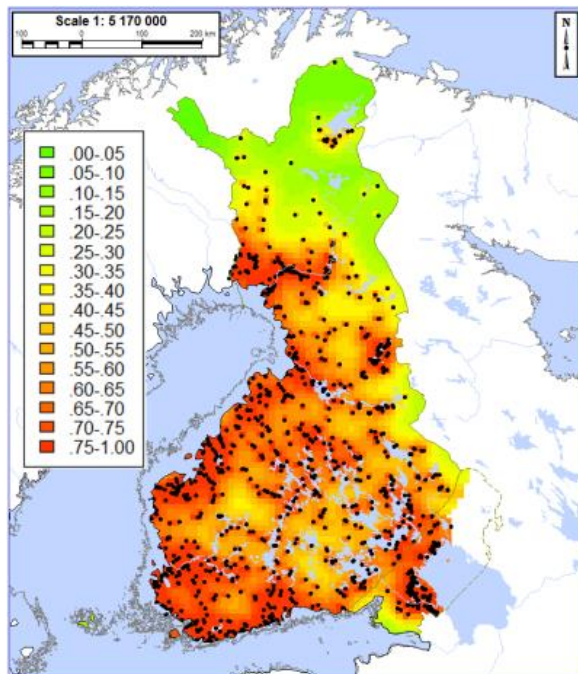


Figure 17 Intensity (posterior density) of stone artefacts from the Middle Neolithic (N2-period). The shoreline of the Baltic Sea corresponds to that of 3500 cal BC estimated by Johan Daniels/Geological Survey of Sweden, SGU.

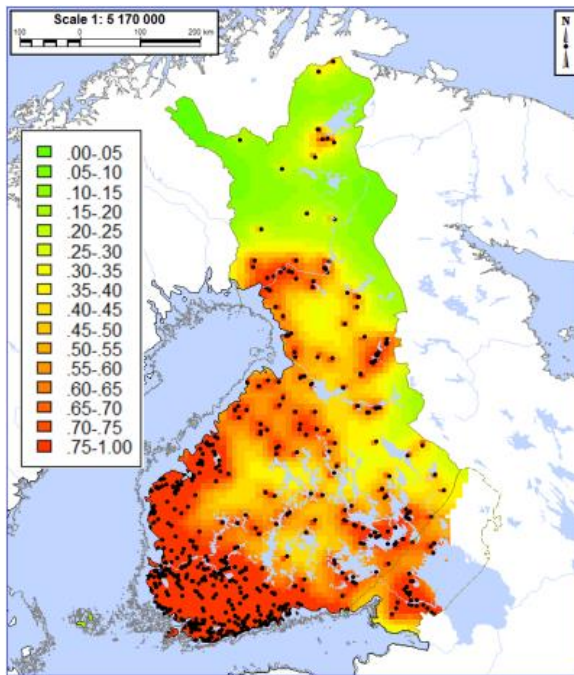


Figure 18 Intensity (posterior density) of stone artefacts from the Late Neolithic (N3-period). The shoreline of the Baltic Sea corresponds to that of 2500 cal BC estimated by Johan Daniels/Geological Survey of Sweden, SGU.

4 DISCUSSION

4.1 Summary of the results (I, II, III, IV)

4.1.1 Simulations

The prehistoric Finnish population has not been previously simulated to this extent. Our first simulation necessarily used a simplified layout where Finland was simulated as a whole unit and no internal migration was yet included. After three years of gradually building and improving the layout and adding attributes such as internal female-specific migration and targeted migration from neighbouring populations, the simulation environment is now capable of simulating more complicated and refined settings.

Our simulations indicate that:

- The constant small gene flow seems to be a much more important factor than few larger migration waves. While the migration waves have barely any effect, moderate constant gene flow can cause great differences to genetic diversity
- Interestingly, the simulation scenarios with a moderate constant migration from neighbouring populations produce genetic diversity measures similar to those observed in present day Finnish population. Consistently, the scenarios without migration induce considerable deviation from these measures
- The first, tight Neolithic bottleneck substantially reduces genetic diversity, which according to our simulations could be detected even today
- Higher female-specific migration brings the simulated mitochondrial genetic diversity closer to the observed contemporary genetic diversity in Finland

When we started our population genetic simulations (I), there were no numerical estimates or references to migration rates, let alone, to gene flow rates to prehistoric Finland. We had to start with some values to narrow down the likely amount of gene flow into the country. We tested different possible migration and gene flow rates to find out the levels that would actualize the contemporary genetic diversity in Finland. In addition, we refined our simulation parameters after noticing for example that the mildest bottleneck used in our first simulation (I), 5000 individuals, was too mild and did not produce genetic diversity measures close to the contemporary Finnish values. We ran the next simulations (III) with the bottleneck sizes better suited to produce the observed values of today.

The number of possible simulation settings is almost limitless. Our choice simulation scenarios were based on archaeological (Lavento 1997, 2001; Saipio 2008; Tallavaara *et al.* 2010) and genetic (Sajantila *et al.* 1996) data. In other words the archaeological signal indicating the prehistoric population size changes – as well as the genetically justified population bottlenecks - are used in the simulations as prior information. We decided to

choose the scenarios which can be argued for using current scientific knowledge. Naturally, archaeological and genetic research continuously produces new data that can be used in our future work.

Simulations are a useful tool in studying population processes unreachable by other means. Nevertheless, they can never be an exact replication of the complex reality, but instead, help us to reject the most infeasible models. It must be noted that the similarities between the modelled and observed diversities do not directly prove causality. One of the principal benefits of simulation models is that you can begin with a simple approximation of a process and gradually refine the model as your understanding of the process improves.

4.1.2 Spatial analyses and stone artefacts

This study applies a Bayesian statistical model (BYM) to build spatial distributions of various archaeological datasets in Finland. The model is based on image analysis (Besag 1986) and assumes similarity of neighbouring areas in geospatial applications. Obviously, the model used here is not exhaustive. Major water systems and other geographical formations could not be included in the applied model. These all can be considered as undisputed geographical factors affecting the drift of people and thereby cultural influences. The methodology presented here is one of the first efforts of Bayesian spatial analysis with different types of archaeological data in Finland.

This is the first time the Database including thousands of objects belonging to several Mesolithic and Neolithic time periods has been data mined to large extent. Previously it has exclusively been used to study singular artefacts and artefact types. The analyses of stone artefacts shows evidence for a marked increase in the number of stone artefacts and stone artefact types coincident with the appearance of TCW culture in 4000-3500 cal BC. This population peak period was followed by a distinct weakening in the archaeological signal, clearly seen in our stone artefact analyses. Thus, the stone artefact analyses further contributes to the line of evidence for a Neolithic population bottleneck in Finland. In addition, spatial and temporal analysis of radiocarbon dates implies a similar decline in the archaeological signal and consequently population size after the peak period (Oinonen *et al.* 2010; Tallavaara *et al.* 2010; Onkamo *et al.* 2012).

4.2 Development of the approach used in this thesis

We have performed two forward-time population genetic simulations (Publications I, III). Our first simulation model (I) was further developed into the refined simulation model (III). The refined model has several improved components which enabled us to build a model of putative prehistoric population events in Finland. Concurrently, we performed Bayesian spatial and temporal analysis of radiocarbon dated archaeological artefacts,

which provided more evidence in support of prehistoric population bottlenecks (Publication II; Pesonen *et al.* 2011; Onkamo *et al.* 2012). This evidence was then integrated into a refined version of the simulations (Publication III). In addition, as part of the methodical development, we also compared R-INLA and MCMC methods (Kammonen *et al.* 2013). The INLA approach proved computationally faster and less memory consuming. Thus, we decided to apply it in our following publication (IV). Here the thesis presents the current developmental state of these approaches and their mutual interactions (Figure 19) including new archaeological data from the Stone Artefact Database.

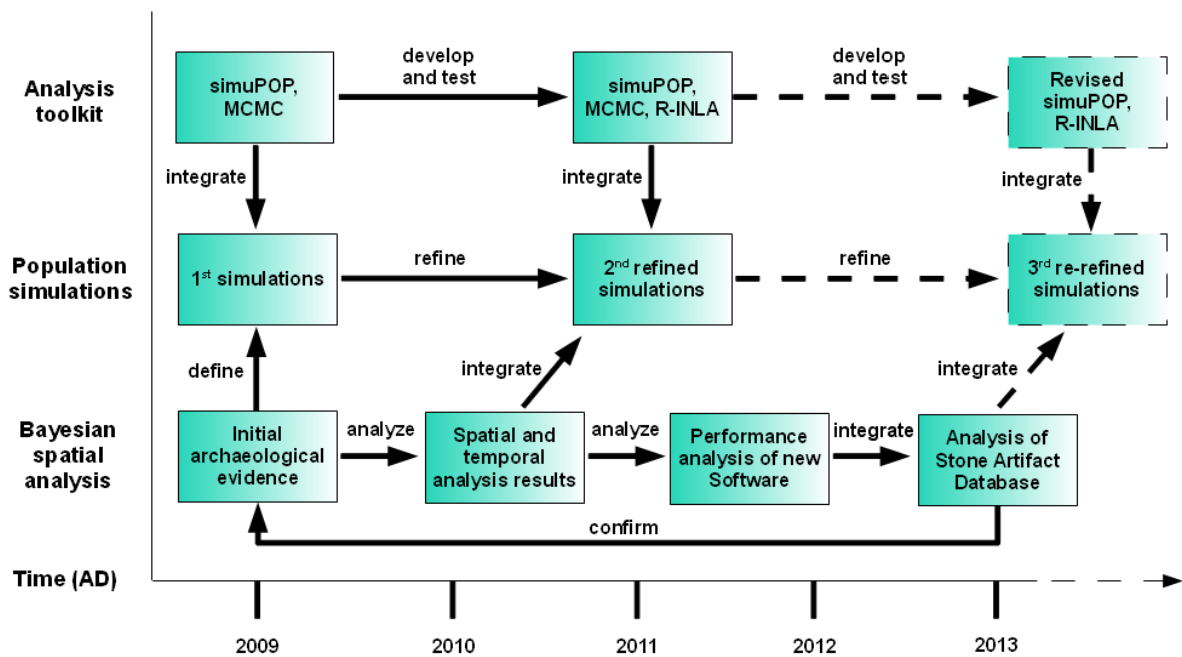


Figure 19 A flowchart depicting interactions between the stages of our present and previous work, showing the overall methodological development. The dashed lines in the chart refer to possible directions in our forthcoming research.

The sequence of our approach is following: Our first simulation publication (I) was followed by a second, refined model (III). The spatio-temporal publication (II) rationalizes the archaeologically justified population peak used as prior information in the simulations. Finally, the new Stone Artefact analyses (IV) add another dimension to the chain of proof for prehistoric population bottleneck(s) in Finland.

4.3 Future prospects

4.3.1 Incorporating geographical data

In the future, the analyses could be expanded to span multiple time periods and utilize recently updated archaeological records. It is possible to incorporate for example geographical data into the simulations and spatial analyses and study possible reasons as well as directions of the spread of human activity and cultural influences across time and space. Especially the addition of a temporal dimension into the analyses would be useful. Furthermore, the inclusion of large waterways, lakes and rivers and the evaluation of their importance as means of transportation would be interesting. Also notable eskers, which traditionally have been considered as barriers for mobility should, rather, be considered as easing the movements of populations and perhaps be added to the model.

4.3.2 Stone artefact analyses

Currently the artefacts in the Stone Artifact Database lack precise dating (other than being associated with the Stone Age). Therefore when working on our publication IV, we had to manually go through all the different typological stone artefact groups and determine a time period for each of them. At the moment, the Stone Artefact Database is being updated by our project Argeopop. In the future, the refinements made in the database will enable more accurate temporal analyses. The data will be also made publicly available.

4.3.3 Simulation of autosomal markers

In this thesis, we concentrated on mitochondrial DNA and Y chromosomal microsatellites which are transmitted maternally and paternally as whole haplotypes without recombination. In principle, simulation of autosomal data is also possible, as the process of recombination is implemented in simuPOP, but we did not yet proceed to that due to the lack of both human and computational resources. Simulation of autosomal markers would be a valuable supplement that should be done in the future. Particularly, simulating the specific Finnish Disease Heritage would be highly interesting.

Our simulation runs notably slowed down with the 631 bp mitochondrial DNA segment and 16 Y chromosomal microsatellites when the population size was growing towards the higher population size. Thus, adding a number of autosomal markers would have been impractical at this stage. The simulations were run in the Murska supercomputer of the Finnish IT Center for Science (CSC). It is clear that additional autosomal markers would have increased the memory use and slowed down the simulation even further, resulting in a higher simulation failure rate. Recently Kammonen (2013) has suggested memory

improvements in the simuPOP workflow, which are also likely to reduce the computational load of simulations, thereby facilitating autosomal simulations in the future.

4.3.4 The coalescent approach

Coalescent based simulators have the advantage of being computationally more efficient than forward time simulators. Nevertheless, the constraints of the coalescent sometimes make it more difficult to model complex evolutionary phenomena (Heled *et al.* 2013), especially gene flow and migration. For example mutations are usually applied separately to the entire simulated genealogy (Excoffier 2000). Another challenge yet to be solved is the modelling of mitochondrial DNA and Y-STR microsatellites with the selected coalescent simulator. Most coalescent simulators only allow binary (sequence of 0's and 1's) representation of individual genomes for segregating and preserved sites. The coalescent simulation scheme could be a valuable supplement to the simulation modelling of Finnish prehistory. Still, it is clear that comprehensive experimental design is needed in order to maintain comparability of coalescent simulation results with those of the forward approach. Considering the framework of this study, we had to narrow down the approaches and leave the coalescent simulations for the future.

4.3.5 Population genetic measures

New research on natural and sexual selection in a historical Finnish population has been published recently (Courtiol *et al.* 2012). The paper suggests that in historical times, males have had higher variance in reproductive success than females. According to the paper the probability of surviving to reproductive age was lower in males than in females. Moreover, among individuals who survived to reproductive age, the probability of marrying at least once was also lower in men. This could be added to our mating model to evaluate the effect on Y-chromosomal diversity. Nevertheless, it has to be carefully considered whether historical reproduction models can be applied to prehistoric times as such.

Linkage disequilibrium (LD) is a phenomenon in which alleles at different but closely situated loci tend to be co-inherited more often than expected. The amount of LD in a population measures the recombination level of that population. By investigating the amount and division of LD it is possible to make inferences regarding the population size and, possibly, genetic population bottlenecks. However, the dating of the bottlenecks is very difficult. In order to study the amount of LD in a population, autosomes have to be incorporated into the simulations.

The effective population size (N_e) depicts the number of individuals in a given generation whose genes contribute to the next generations. Previously, we have compared the actual total population size with theoretical effective population size in publication III. In the future, other ways of counting this population genetic measure will be investigated.

F_{ST} serves as a measure of genetic distance (variation) among subpopulations. The smaller the F_{ST} between the subpopulations is the more similar are the subpopulation. F_{ST} can be estimated as the deviation between the average number of pairwise differences pooled from all subpopulations and the heterozygosity estimated from each subpopulation with relation to the pooled differences. This ratio was not calculated in the first version of the simulation model (Publication I) as there was no subpopulation structure. In publication III we had divided Finland into three subpopulations, however, the F_{ST} ratio was not calculated since the contemporary genetic data available (Palo *et al.* 2009) did not include values which could have been compared to our simulations. The calculation of the F_{ST} statistic becomes more relevant in case of numerous simulated subpopulations, as for example is the case in the Savonian expansion simulation performed in our project (Heger 2011).

5 CONCLUSIONS

This dissertation comprises four original publications combining information acquired from both archaeological and genetic research, introducing new genuinely multidisciplinary results on top of the previous archaeological work that has been done before in Finland. The first simulation publication was followed by a refined model. The spatio-temporal publication rationalizes the archaeologically justified population peak used as a proxy in the simulation publications. Finally, the last publication summarizes the previous three publications and adds another dimension to the chain of proof by introducing the Stone Artefact Database analyses which provide more evidence in support of a Neolithic population bottleneck.

This contribution of combining analyses of detailed archaeological datasets with population genetic simulations is the first of its kind. When starting our research project it became clear that this type of research has not been done anywhere else. Our methodology used here has been shown to be appropriate for evaluating prehistoric demographic events in a genetically and culturally continuous population. This same methodology could potentially be applied to any such population.

Current genetic diversity provides only indirect evidence about the demographic events of past populations. Instead, population genetic simulation tools can be used to study the effects of population demography on genetic diversity over thousands of years. Within our simulated framework we have been capable of testing several scenarios including *e.g.* migrations and bottleneck sizes comparing those to the present day genetic diversity in Finland. Furthermore, additional evidence and current datasets can be integrated in the model, *e.g.* new aDNA results from neighbouring areas can be added to the background population pools to refine the results.

We describe and utilize the coverage of the Ceramic as well as the new Stone Artefact Database and employ a Bayesian spatial model to analyze stone artefact dispersion in prehistory. The methodology presented here is one of the first efforts of applying Bayesian spatial analysis with different types of archaeological data in Finland. The Stone Artefact and Ceramic Databases, and our analyses derived from them, offer unique opportunities to compare observations derived from genetic evidence. With this kind of approach, we can evaluate the evidence in a more detailed fashion compared to solely an archeological or genetic approach. Previously, the Stone Artefact Database has been used to study singular artefacts and a few artefact types exclusively. This is the first time the Database including thousands of objects belonging to several Mesolithic and Neolithic time periods has been data mined to a large extent.

The spread of Typical Comb Wear has been associated with the arrival of Finno-Ugric languages into Finland and the neighbouring areas. However, the timing of the arrival of Finno-Ugric languages has been an enduring issue among the linguistics. This topic, which is beyond the scope of this study, inevitably continues to arouse vivid discussions. Additionally, the question of prehistoric migration into the area has been continuously debated in Finnish archaeology. Our population simulations show that the scenarios including small migration waves and moderate constant gene flow from neighbouring areas are those where the strongest similarity to present day genetic diversity can be observed.

This achievement of combining information acquired from both archaeological and genetic research is crucial for the understanding of the Finnish prehistoric population fluctuations as well as the contemporary genetic composition of the inhabitants of Finland. Information required by this research generates a greater awareness of demographic processes *per se* and increases our understanding of the overall prehistoric population dynamics. These accomplishments represent a major contribution to the multidisciplinary research of Finnish prehistory in all its complexity.

REFERENCES

- Alenius, T., Mökkönen, T.O. & Lahelma, A. 2013. Early Farming in the Northern Boreal Zone: Re-assessing the History of Land Use in South-Eastern Finland through High-resolution Pollen Analysis. *Geoarchaeology* 28, 1: 1-24.
- Argeopop. <http://www.helsinki.fi/bioscience/argeopop>
- Besag, J. 1986. On the Statistical Analysis of Dirty Pictures. *Journal of the Royal Statistical Society. Series B* 48: 259-302.
- Besag, J., York, J., and Mollié, A. 1991. Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics* 43:1-59.
- Best, N., Richardson, S., and Thomson, A. A comparison of Bayesian spatial models for disease mapping. *Statistical Methods in Medical Research*, 14: 35-59, 2005.
- Beaumont, M. A., Zhang, W. & Balding, D. J. 2002. Approximate Bayesian Computation in Population Genetics. *Genetics* 162: 2025-2035.
- Bramanti, B., Thomas, M.G., Haak, W., Unterlaender, M., Jores, P., Tambets, K., Antanaitis-Jacobs, I., Haidle, M.N., Jankauskas, R., M.N., Kind, C.-J., Lueth, F., Terberger, T., Hiller, J., Matsumura, S., Forster, P. & Burger, J. 2009. Genetic Discontinuity Between Local Hunter-Gatherers and Central Europe's First Farmers. *Science* 326: 137-40.
- Brotherton, P., Haak, W., Templeton, J., Brandt, G., Soubrier, J., Adler C. J., Richards, S. M., Der Sarkissian, C., Ganslmeier, R., Friederich, S., Dresely, V., van Oven, M., Kenyon, R., Van der Hoek, M.B., Korfach, J., Luong, K., Ho, S.Y.W., Quintana-Murci, L., Behar, D.M., Meller, H., Alt, K.W., Cooper, A. & The Geographic Consortium (2013). Neolithic mitochondrial haplogroup H genomes and the genetic origins of Europeans. *Nature communications* 4:1764 doi: 10.1038/ncomms2656:1-11.
- Carpelan, C. 1999a. Käännekohtia Suomen esihistoriassa aikavälillä 5100...1000 eKr. Pohjan poluilla. Suomalaisten juuret nykytutkimuksen mukaan. In P. Fogelberg (ed.), *Bidrag till kännedom av Finlands natur och folk* 153. Societas Scientiarum Fennica. Helsinki. 249-80.
- Carpelan, C. 1999b. On the Postglacial Colonisation of Eastern Fennoscandia. In M. Huurre (ed.), *Dig it all. Papers dedicated to Ari Siiriäinen*: 151-71. The Finnish Antiquarian Society & The Archaeological Society of Finland, Helsinki.
- Carpelan, C. & A. Parpola. 2001. Proto-Indo-European, Proto-Uralic and Proto-Aryan. *The Earliest Contacts between Uralic and Indo-European: Linguistic and Archeological Considerations*, in C. Carpelan, A. Parpola & P. Koskikallio (ed.). *Mémoires de la Société finno-ougrienne* 242.

- Cavalli-Sforza, L.L., Menozzi, P. & Piazza, A. 1994. *The History and Geography of Human Genes*. Princeton.
- Costopoulos, A., Vaneeckhout, S., Okkonen, J., Hulse, E., PaberzYTE, I. & Wren, C.D. 2012. Social Complexity in the Mid-Holocene Northeastern Bothnian Gulf. *European Journal of Archaeology*, Volume 15, Number 1: 41-60.
- CourtIol, A., Pettay, J. E., Jokela, M., Rotkirch, A. & Lummaa, V. Natural and sexual selection in a monogamous historical human population. 2012. *Proc Natl Acad Sci USA*.
- Cramp, L.J.E, Evershed, R.P., Lavento, M., Halinen, P., Mannermaa, K., Oinonen, M., Kettunen, J., Perola, M., Onkamo, P. & Heyd, V.. (Submitted). Prehistoric Dairying in Fennoscandia at the Northerly Extreme of Agriculture Above the 60th Parallel.
- De la Chapelle, A. & Wright, F. A. 1998: Linkage disequilibrium mapping in isolated populations: the example of Finland revisited. *Proc Natl Acad Sci USA* 95: 12416-23.
- De la Chapelle, A. 1999. Geenimutaatiot Suomen väestön juurten tutkimuksissa. Pohjan poluilla. Suomalaisten juuret nykytutkimuksen mukaan. In P. Fogelberg (ed.), *Bidrag till kännedom av Finlands natur och folk* 153. Societas Scientarum Fennica. Helsinki. 307-10.
- Edgren, T. 1999: Käännekohtia Suomen kivikaudessa. Pohjan poluilla. Suomalaisten juuret nykytutkimuksen mukaan In P. Fogelberg (ed.), *Bidrag till kännedom av Finlands natur och folk* 153. Societas Scientarum Fennica. Helsinki. 281-93.
- Edgren, T. 2007. On the non-megalithic mortuary practices in Finland. *Innovation and Continuity – Non-Megalithic Mortuary Practices in the Baltic. New Methods and Research into the Development of Stone Age Society*, edited by L. Larsson, F. Lüth, and T. Terberger. Bericht der Römisch-Germanischen Kommission 88:501-20.
- EUROEVOL. <http://www.ucl.ac.uk/euroevol/>.
- Excoffier, L., Novembre, J. & Schneider, S. 2000. SIMCOAL: a general coalescent program for the simulation of molecular data in interconnected populations with arbitrary demography. *Journal of Heredity*. 91: 506–509.
- FDH. <http://www.findis.org>
- Finnilä, S., M.S. Lehtonen & K. Majamaa. 2001. Phylogenetic network for European mtDNA. *The American Journal of Human Genetics*. 68: 1475-84.
- Forsberg, L. 1996. The Earliest Settlement of Northern Sweden - Problems and Perspectives. In L. Larsson (ed.), *The Earliest Settlement of Scandinavia and its Relationship with Neighbouring Areas*. Acta archaeologica Lundensia Series in 8°, No. 24. Stockholm. 241-50.

- Gamble, C., Davies, W., Pettitt, P., Hazelwood, L., and Richards, M. 2005. The Archaeological and Genetic Foundations of the European Population during the Late Glacial: Implications for 'Agricultural Thinking'. *Cambridge Archaeological Journal* 15:193–223.
- Haak, W., Forster, P., Bramanti, B., Matsumura, S., Brandt, G., Tänzer, M., Villems, R., Renfrew, C., Gronenborn, D., Alt, K.W. & Burger, J. 2005. Ancient DNA from the First European farmers in 7500-Year-Old Neolithic Sites. *Science* 310: 1016-8.
- Halinen, P. 1999a. Saamelaiset – arkeologinen näkökulma. Pohjan poluilla. Suomalaisten juuret nykytutkimuksen mukaan In P. Fogelberg (ed.), *Bidrag till kännedom av Finlands natur och folk* 153. Societas Scientiarum Fennica. Helsinki. 121-130.
- Halinen, P. 1999b. Burial Practices and the Structure of Societies during the Stone Age in Finland, in M. Huurre (ed.), *Dig it all, papers dedicated to Ari Siiriäinen*. The Finnish Antiquarian Society and The Archaeological Society of Finland. Helsinki. 173-179.
- Halinen, P. 2005. Prehistoric Hunters of Northernmost Lapland. Settlement patterns and subsistence strategies. *Iskos 14*. Vammala.
- Halinen, P. 2011. Arkeologia ja saamentutkimus in I. Seurujärvi-Kari, P. Halinen & R. Pulkkinen (eds.), *Saamentutkimus tänään. Tietolipas* 234. Helsinki, Suomalaisen Kirjallisuuden Seura. p. 130-176.
- Halinen, P., Heyd, V., Lavento, M. & Mannermaa, K. (Submitted). "When south meets north - Corded Ware in Finland". *Cordeds Days in Kraków* (ed.) Włodarczyk P (Instytut Archeologii i Etnologii Pan, oddział w Krakowie).
- Hannelius, U., Salmela, E., Lappalainen, T., Guillot, G., Lindgren, C. M., von Döbeln, U., Lahermo, P. & Kere, J. 2008: Population substructure in Finland and Sweden revealed by the use of spatial coordinates and a small number of unlinked autosomal SNPs. *BioMedCentral Genetics* 9(54): 1-12.
- Hedman, M., Brandstätter, A., Pimenoff, V., Sistonen, P., Palo, J. U., Parson, W. & Sajantila, A. 2007. Finnish mitochondrial DNA HVS-I and HVS-II population data. *Forensic Science International* 172: 171-8.
- Held, L., Natário, I., Fenton, S. E., Rue, H. and Becker, N. Towards joint disease mapping. *Statistical Methods in Medical Research*, 14: 61-82, 2005.
- Heger, M. 2011. *Population Genetic Simulations of the Savonian Expansion in Finland*. MSc thesis. Department of Biosciences, University of Helsinki.
- Heled, J., Bryant, D. & Drummond, A.J. 2013. Simulating gene trees under the multispecies coalescent and time-dependent migration. *BMC Evolutionary Biology*, 13:44, doi:10.1186/1471-2148-13-44.

- Heyer, E., Puymirat, J., Dieltjes, P., Bakker, E. & de Knijff, P. 1997. Estimating Y Chromosome Specific Microsatellite Mutation Frequencies Using Deep Rooting Pedigrees, *Human Molecular Genetics*. 6 (5):799-803.
- Heyer, E., Zietkiewicz, E., Rochowski, A., Yotova, V., Puymirat, J. & Labuda D 2001. Phylogenetic and Familial Estimates of Mitochondrial Substitution Rates: Study of Control Region Mutations in Deep-Rooting Pedigrees. *American Journal of Human Genetics*. 69(5):1113-1126
- Huurre, M. 1990. *9000 vuotta Suomen esihistoriaa*. Keuruu.
- Huurre, M. 2001. *Kivikauden Suomi*. Keuruu.
- Jakkula, E., Rehnström, K., Varilo, T., Pietiläinen, O.P.H., Paunio, T., Pedersen, N. L., deFaire, U., Järvelin, M-R., Saharinen, J., Freimer, N., Ripatti, S., Purcell, S., Collins, A., Daly, M.J., Palotie, A. & Peltonen, L. 2008: The Genome-wide patterns of Variation Expose significant Substructure in a Founder Population. *The American Journal of Human Genetics* 83: 787-94.
- Jobling, M. A., Hollox, E., Hurles, M.E., Kivisild, T. & Tyler-Smith, C. 2013. *Human Evolutionary Genetics second edition*. New York.
- Kammonen, J. 2013. Improving the throughput of the forward population genetic simulation environment simuPOP. MSc thesis, Department of Computer Science, University of Helsinki.
- Kammonen, J., Sundell, T., Pesonen, P., Oinonen, M. & Onkamo, P. Bayesian Spatial Modelling of Radiocarbon Dated Archaeological Artefacts Using R-INLA. 2013. *CAA 2012: Proceedings of the 40th Annual Conference on Computer Applications and Quantitative Analysis in Archaeology*. 414-419
- Kayser & Sajantila. 2001. Mutations at Y-STR loci: implications for paternity testing and forensic analysis. *Forensic Science International*.
- Sigurdardottir et al. 2000. The Mutation Rate in the Human mtDNA Control Region, *American Journal of Human Genetics* (?)
- Kere, J. 2001: Human population genetics: Lessons from Finland. *Annual Review of Genomics and Human Genetics* 2: 103-28.
- Kerosuo, L. 2010. Höläyksiä. *Aforismeja vuosilta 1993-2008*. Into Kustannus. Vantaa.
- Kittles, R. A., Perola, M. & Peltonen, L. 1998: Dual origins of Finns revealed by Y chromosome haplotype variation. *The American Journal of Human Genetics* 62: 1171-9.
- Kittles, R. A., Bergen, A. W., Urbanek, M., Virkkunen, M., Linnoila, M., Goldman, D. & Long, J. C. 1999: Autosomal, Mitochondrial and Y Chromosome DNA Variation in

- Finland: Evidence for a Male-Specific Bottleneck. *American Journal of Physical Anthropology* 108: 381-99.
- Lahermo, P., Sajantila, A., Sistonen, P., Lukka, M., Aula, P., Peltonen, L., Savontaus, M-L. 1996. The Genetic Relationship between the Finns and the Finnish Sami (Lapps): Analysis of Nuclear DNA and mtDNA. *The American Journal of Human Genetics* 58: 1309-22.
- Lahermo, P., Savontaus, M-L., Sistonen, P., de Knijff, P., Aula, P., Sajantila, A. 1998. Y chromosomal polymorphisms reveal founding lineages in the Finns and the Sami.
- Lahermo, P. 1998: Mitochondrial and Y Chromosomal Variations in the Finno-Ugric-Speaking Peoples. *Turun yliopiston julkaisu*, sarja D, osa 315. Medica-Odontologia.
- Lao, O., Lu, T.T., Nothnagel, M., Junge, O., Freitag-Wolf, S., Caliebe, A., Balascakova, M., Bertranpetit, J., Bindoff, L.A., Comas, D., Holmlund, G., Kouvatsi, A., Macek, M., Mollet, I., Parson, W., Palo, J., Ploski, R., Sajantila, A., Tagliabracci, A., Gether, U., Werge, T., Rivadeneira, F., Hofman, A., Uitterlinden, A.G., Gieger, C., Wichmann, H.E., Ruther, A., Schreiber, S., Becker, C., Nurnberg, P., Nelson, M.R., Krawczak, M. & Kayser, M. 2008. Correlation between Genetic and Geographic Structure in Europe. *Current biology, CB*, vol. 18, no. 16, 1241-1248.
- Lappalainen, T., Koivumäki, S., Salmela, E., Huoponen, K., Sistonen, P., Savontaus, M-L. & Lahermo, P. 2006. Regional differences among Finns: A Y-chromosomal perspective. *Gene* 376: 207-15.
- Lappalainen, T., Laitinen, V., Salmela, E., Andersen, P., Huoponen, K., Savontaus, M-L. & Lahermo, P. 2008: *Migration Waves to the Baltic Sea Region. Annals of Human Genetics* (337-48).
- Lavento, M. 1997. Geoarchaeological Observations on the Early Metal Period dwelling sites in the Ancient Lake Saimaa area. *Slavjane i Finno-ugri. Arheologija, istorija, kultura. St. Petersburg.* 38-46.
- Lavento M. 1998. Sisämaan vanhemman metallikauden väestö tutkimusongelmana. *Muinaistutkija* 4/1998: 46-55.
- Lavento, M. 2001. Textile Ceramics in Finland and on the Karelian Isthmus. Nine Variations and Fugue on a Theme of C. F. Meinander. *Suomen Muinaismuistoyhdistyksen Aikakauskirja* 109. Doctoral Dissertation, Department of Archaeology, University of Helsinki. Vammala.
- LeCHE. <https://sites.google.com/a/palaeome.org/leche/>
- Leonardi, M., Gerbault, P., Thomas, M.G., & Burger, J. 2012. The evolution of lactase persistence in Europe. A synthesis of archaeological and genetic evidence. *International Dairy Journal* 22: 88-97.

- Lunn, D. J, Thomas, A., Best, N., and Spiegelhalter, D. 2000. WinBUGS – a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing* 10:325-337.
- Malmström, H., Gilbert, M.T.P., Thomas, M.G., Brandström, M., Storå, J., Molnar, P., Andersen, P.K., Bendixen, C., Holmlund, G., Götherström, A. & Willerslev, E. 2009. Ancient DNA Reveals Lack of Continuity between Neolithic Hunter-Gatherers and contemporary Scandinavians. *Current Biology* 19: 1758-62.
- Manninen, M. A., Tallavaara, M. and Hertell, E. Subneolithic Bifaces and Flint Assemblages in Finland: Outlining the History of Research and Future Questions. *Uniting Sea. Occasional Papers in Archaeology* 33: 161-179. Uppsala Universitet. Uppsala, 2003.
- McEvoy, B. P., Montgomery, G. W., McRae, A. F., Ripatti, S., Perola, M., Spector, T. D., Cherkas, L., Ahmadi, K. R., Boomsma, D., Willemsen, G., Hottenga, J. J., Pedersen, N. L., Magnusson, P. K. E., Kyvik, K. O., Christensen, K., Kaprio, J., Heikkilä, K., Palotie, A., Widen, E., Muilu, J., Syvänen, A-C., Liljedahl, U., Hardiman, O., Cronin, S., Peltonen, L., Martin, N. G. and Visscher, P. M. 2009: Geographical structure and differential natural selection among North European populations. *Genome Research* 19: 804-14.
- Meinander, C. F. 1984: Kivikautemme väestöhistoria. In Gallén, J. (ed). *Bidrag till kännedom av Finlands natur och folk* H. 131. Suomen väestön esihistorialliset juuret. Societas Scientiarum Fennica. Helsinki. 21-48.
- Melchior, L., Gilbert, M.T.P., Kivisild, T., Lynnerup, N. & Dissing, J. 2007. Rare mtDNA Haplogroups and Genetic Differences in Rich and Poor Danish Iron-Age Villages. *American Journal of Physical Anthropology* 135: 206-15.
- Melchior, L., Kivisild, T., Lynnerup, N. & Dissing, J. 2008. Evidence of Authentic DNA from Danish Viking Age Skeletons Untouched by Humans for 1,000 Years. *The Public Library of Science, PLOSone*, 3(5).
- Mökkönen, T. 2002. Chronological Variation in the Locations of Hunter-Gatherer Occupation Sites vis-à-vis the Environment. In H. Ranta (ed.), *Huts and Houses. Stone Age and Early Metal Age Buildings in Finland*: 53-64. Finnish National Board of Antiquities, Helsinki.
- Mökkönen, T. 2011. Studies on Stone Age housepits in Fennoscandia (4000-2000 CAL BC). Changes in ground plan, site location, and degree of sedentism. Helsinki.
- Nei, M. 1987. *Molecular evolutionary genetics*. Columbia University Press.
- Nevanlinna, H. R. 1972. The Finnish population structure. A genetic and genealogical study. *Hereditas* 71: 195-236.
- Nevanlinna, H. R. 1973: *Suomen väestörakenne: geneettinen ja genealoginen tutkimus*. Kansaneläkelaitoksen julkaisuja, sarja A:9. Vammala.

- Nevanlinna, H. R. 1984: Suomalaisten juuret geneettisen merkkiominaisuus-tutkimuksen valossa. *Bidrag till kännedom av Finlands natur och folk* H. 131. Suomen väestön esihistorialliset juuret. Societas Scientiarum Fennica. Helsinki. (157-174).
- Norio R, Nevanlinna H. R, Perheentupa J. 1973. Hereditary diseases in Finland; rare flora in rare soul. *Annals of Clinical Research*. 5(3): 109-141.
- Norio, R. 2000. *Suomi-neidon geenit*. Helsinki.
- Norio, R. 2003a. Finnish Disease Heritage I: characteristics, cause, background. *Human Genetics*. Vol. 112, no. 5-6, pp 441- 456.
- Norio, R. 2003b. Finnish Disease Heritage II: population prehistory and genetic roots of Finns. *Human Genetics*. Vol. 112, no. 5-6, pp 457-469.
- Norio, R. 2003c. Finnish Disease Heritage III: the individual diseases. *Human Genetics*. Vol. 112, no. 5-6, pp 470-526.
- Norio, R. 2004. Suomalaisten mutkalliset geenijuuret. In R. Grünthal (ed.), *Ennen, muinoin. Miten menneisyyttämme tutkitaan?* Helsinki. 174-87.
- Nuñez, M. 1987: A Model for the Early Settlement of Finland. *Fennoscandia archaeologica* IV. 3-18.
- Nuñez, M. 1997a. Finland's settling model revisited. Varhain Pohjoisessa. Maa. Varhain Pohjoisessa -hankkeen artikkelit. The Land. Reports of the Early in the North Project, in E-L. Schulz & C. Carpelan (ed.) *Helsinki papers in archaeology* 10. 1997: 93-102.
- Nuñez, M. 1997b. Uusi katsaus Suomen asuttamismalliin. Itämerensuomi – eurooppalainen maa. With Estonian, Hungarian and English Summaries, in K. Julku (ed.). *Studia Historica Fenno-ugrica* 2: 47-62.
- Nuñez, M. & Okkonen, J. 1999. Environmental Background for the Rise and Fall of Villages and Megastructures in North Ostrobothnia 4000-2000 calBC. In M. Huurre (ed.), *Dig it all. Papers dedicated to Ari Siiriäinen*: 173-9. The Finnish Antiquarian Society & The Archaeological Society of Finland, Helsinki. 105-15.
- Oinonen, M., P. Pesonen & M. Tallavaara. 2010. Archaeological radiocarbon dates for studying the population history in eastern Fennoscandia. *Radiocarbon* 52: 393–407.
- Okkonen, J. 2003. Jättiläisen hautoja ja hirveitä kiviröykkiöitä – Pohjanmaan muinaisten kivirakennelmien arkeologiaa. Oulun Yliopisto. Taideaineiden ja antropologian laitos. Oulu. <http://herkules oulu.fi/isbn951427170X/isbn951427170X.pdf>.
- Onkamo P., J. Kammonen, P. Pesonen, T. Sundell, E. Moltchanova, M. Oinonen & M. Haimila. 2012. Bayesian spatio-temporal analysis of radiocarbon dates from Eastern Fennoscandia. *Radiocarbon* 54: 649-59.

- Ortman, S. G., Varien, M. D., and Gripp, T. L. 2007. Empirical Bayesian Methods for Archaeological Survey Data: An Application from the Mesa Verde Region. *American Antiquity* 72:241-272.
- Palo, J. U., Ulmanen, I., Lukka, M., Ellonen, P. & Sajantila, A. 2009. Genetic markers and population history: Finland revisited. *European Journal of Human Genetics*: Volume 17: 1336-46.
- Parpola, A. 2012. Formation of the Indo-European and Uralic (Finno-Ugric) language families in the light of archaeology: Revised and integrated 'total' correlations. In Grünthal, R. & Kallio, P. (eds.) *A Linguistic Map of Prehistoric Northern Europe. Suomalais-Ugrilaisen Seuran Toimituksia – Mémoires de la Société Finno-Ougrienne* 266. Helsinki. 119-84.
- Peng B. & Kimmel M. 2005. SimuPOP: a forward-time population genetics simulation environment. *Bioinformatics* 21: 3686-7.
- Peltonen, L., Palotie, A. & Lange, K. 2000. Use of Population Isolates for Mapping Complex Traits. *Nature*, 1: 182-90.
- Pesonen, P. 1999. Radiocarbon dating of birch bark pitches in the Typical Comb Ware in Finland. In: Huurre, M, editor. *Dig it all. Papers dedicated to Ari Siiriäinen*. The Finnish Antiquarian Society, the Archaeological Society of Finland. 191-199.
- Pesonen, P. 2002. Semisubterranean Houses In Finland – A Review. In H. Ranta (ed.), *Huts and Houses. Stone Age and Early Metal Age Buildings in Finland*: 53-64. Finnish National Board of Antiquities, Helsinki.
- Pesonen P. Neolithic pots and ceramics chronology – AMS-datings of Middle and Late Neolithic ceramics in Finland. In Uino P, editor. *Fenno-Ugri et Slavi 2002: Dating and Chronology*. Museoviraston arkeologian osaston julkaisuja n:o 10. Helsinki: National Board of Antiquities, 87-97, 2004.
- Pesonen, P. 2005: Sarvingin salaisuus – Enon Rahakankaan varhaismesoliittinen ajoitus. *Muinaistutkija* 2/2005. Helsinki. 2-13.
- Pesonen P. & T. Sundell T. 2011. Argeopop-projektin arkeologiset tietokannat. *Muinaistutkija* 4/2011: 11-18.
- Pesonen, P. and Tallavaara, M. 2008. *Suomen esihistoriallinen väestömäärä 8500 calBC – 1000 calAD*. University of Helsinki. Helsinki.
- Pesonen P, J. Kammonen, E. Moltchanova, M. Oinonen & P. Onkamo. 2011. *Archaeological Radiocarbon Dates and Ancient Shorelines - Resources and Reservoirs*. In Ikonen, A.T.K. & Lipping, T. (eds.). Proceedings of the seminar "The Sea Level Displacement and Bedrock Uplift", Pori, Finland. Posiva Working Report 2011-07: 119-129.

Pesonen, P., M. Oinonen, E. Holmqvist-Saukkonen, S. Kivimäki, T. Sundell. & P. Onkamo. Submitted. The Vuoksi breakthrough – A natural disaster triggering a large-scale cultural change in Neolithic eastern Fennoscandia.

Python Programming Language. <http://www.python.org>

Rankama, T. & Kankaanpää, J. 2008. Eastern arrivals in post-glacial Lapland: The Sujala site 10000 cal BP. *Antiquity*, Vol. 82 Issue 318, 884-899.

Ray, N. & Excoffier, L. 2009. Inferring Past Demography Using Spatially Explicit Population Genetic Models. *Human Biology*, 81(3):141–157.

R-INLA project. <http://www.r-inla.org>

Rudbeck, L., Gilbert, M.T.P., Willerslev, E., Hansen, A.J., Lynnerup, N., Christensen, T. & Dissing, J. 2005. MtDNA Analysis of Human Remains from an Early Danish Christian Cemetery. *American Journal of Physical Anthropology* 128: 424-9.

Saipio, J. 2008. Hävisikö sisämaan asbestikeraaminen väestö pronssikaudella? *Muinaistutkija* 2/2008: 2-18.

Sajantila, A., Lahermo, P., Anttinen, T., Lukka, M., Sistonen, P., Savontaus, M-L., Aula, P., Beckman, L., Tranebjaerg, L., Gedde-Dahl, T., Isser-Tarver, L., DiRienzo, A. & Pääbo, S. 1995: Genes and Languages in Europe: An Analysis of Mitochondrial Lineages. *Genome Res* 5: 42-52.

Sajantila, A., Salem, A-H., Savolainen, P., Bauer, K., Gierig, C., Pääbo, S. 1996. *Paternal and maternal DNA lineages reveal a bottleneck in the founding of the Finnish population. Proceedings of the National Academy of Sciences USA*,93: 12035-9.

Salmela, E., Taskinen, O., Seppänen, J. K., Sistonen, P., Daly, M. J., Lahermo, P., Savontaus, M-L. & Kere, J. 2006: Subpopulation difference scanning: a strategy for exclusion mapping of susceptibility genes. *Journal of Medical Genetics*, 43(7): 590-7.

Salmela, E., Lappalainen, T., Fransson, I., Andersen, P. M., Dahlman-Wright, K., Fiebig, A., Sistonen, P., Savontaus, M-L., Schreiber, S., Kere, J. & Lahermo, P. 2008: Genome-Wide Analysis of Single Nucleotide Polymorphisms Uncover Population Structure in Northern Europe. *The Public Library of Science, PLoS ONE* 3:e.

Savontaus, M-L. & Lahermo, P. 1999. Uralilainen muinaisuutemme väestögenetiikan valossa. Pohjan poluilla. Suomalaisten juuret nykytutkimuksen mukaan. *Bidrag till kännedom av Finlands natur och folk* 153. Societas Scientiarum Fennica. Helsinki. 60-63.

Service, S., DeYoung, J., Karayiorgou, M., Roos J.L., Pretorius, H., Bedoya, G., Ospina, J., Ruiz-Linares, A., Macedo, A., Palha, J.A., Heutink, P., Aulchenko, Y., Oostra, B., van Duijn, C., Jarvelin, M.R., Varilo, T., Peddle, L., Rahman, P., Piras G., Monne, M., Murray, S., Galver, L., Peltonen, L., Sabatti, C., Collins, A. & Freimer, N. 2006:

- Magnitude and distribution of linkage disequilibrium in population isolates and implications for genome-wide association studies. *Nature Genetics* 38: 556-60.
- Shennan, S. & K. Edinborough. 2007. Prehistoric population history: from the Late Glacial to the Late Neolithic in Central and Northern Europe. *Journal of Archaeological Science* 34: 1339–45.
- Shennan, S., Downey, S.S., Timpson, A., Edinborough, K., Colledge, S., Kerig, T., Manning, K. & Thomas, M.G. Regional population collapse followed initial agriculture booms in mid-Holocene Europe. *Nature Communications* 4(2486). doi:10.1038/ncomms3486.
- Siiriäinen, A. 1981. On the cultural ecology of the Finnish Stone Age. *Suomen Museo* 87: 5-40.
- Simpson, D., Lindgren, F. & Rue, H. Fast approximate inference with INLA: the past, the present and the future. 2011. Cornell University Library. arXiv:1105.2982
- Skoglund, P., Malmström, H., Raghavan, M., Storå, J. Hall, P., Willerslev, E., Gilbert, M.T.P., Götherström, A., & Jakobsson, M. 2012. Origins and Genetic Legacy of Neolithic Farmers and Hunter-Gatherers in Europe. *Science* 336, 466.
- SPSS Inc. 2009. PASW Statistics for Windows, Version 18.0. Chicago.
- Sundell, T. 2008. Populaatiopullonkaulat kivikauden Suomessa – arkeologinen ja geneettinen tulkinta. MA thesis, Department of Archaeology, University of Helsinki.
- Sundell, T. 2009. Kadonneet geenit: Suomen kivikautisten populaatiopullonkaulojen tutkiminen geneettisin menetelmin. *Muinaistutkija* 4/2009: 10-18.
- Sundell, T. & Onkamo, P. 2010. Argeopop: Keramiikasta geenivirtoihin? *Muinaistutkija* 1/2010: 3-8.
- Surovell, T. A. & Brantingham, P. J. 2007. A note on the use of temporal frequency distributions in studies of prehistoric demography. *Journal of Archaeological Science* 34: 1868-77.
- Surovell, T. A., Finley, J. B., Smith, G. M., Brantingham, P. J. & Kelly, R., 2009. Correcting temporal frequency distributions for taphonomic bias. *Journal of Archaeological Science* 36: 1715-24.
- Takala, H. 2004: *The Ristola Site in Lahti and the Earliest Postglacial Settlement of South Finland*. Jyväskylä.
- Tallavaara, M. & H. Seppä. 2012. Did the mid-Holocene environmental changes cause the boom and bust of hunter-gatherer population size in eastern Fennoscandia? *Holocene*. 22, 2: 215-25.

- Tallavaara, M., P. Pesonen & M. Oinonen. 2010. Prehistoric population history in eastern Fennoscandia. *Journal of Archaeological Science* 37: 251-260.
- Tumonggor M.K., T.M. Karafet, B. Hallmark, J.S. Lansing, H. Sudoyo, M.F. Hammer & M.P. Cox. 2013. The Indonesian archipelago: an ancient genetic highway linking Asia and the Pacific. *Journal of Human Genetics*. doi: 10.1038/jhg.2012.154
- Uimari, P., Kontkanen, O., Visscher, P. M., Pirskanen, M., Fuentes, R. & Salonen, J. T. 2005: Genome-Wide Linkage Disequilibrium from 100,000 SNP's in the East Finland Founder Population. *Twin Research and Human Genetics*,8(3):185-197.
- Varilo, T. 1999. The age of the mutations in the Finnish disease heritage; a genealogical and linkage disequilibrium study. Doctoral Dissertation, Medical Faculty, University of Helsinki.
- Varilo, T., Laan, M., Hovatta, I., Wiebe, V., Terwilliger, J. D. & Peltonen, L. 2000: Linkage disequilibrium in isolated populations: Finland and a young sub-population of Kuusamo. *European Journal of Human Genetics* 8: 604-612.
- Varilo, T., Paunio, T., Parker, A., Perola, M., Meyer, J., Terwilliger, J. D. & Peltonen, L. 2003. The interval of linkage disequilibrium (LD) detected with microsatellite and SNP markers in chromosomes of Finnish populations with different histories. *Human Molecular Genetics* 12: 51-59.
- Vuorinen, J.H.T. 1982. *Piikivi ja Suomen kampakeraaminen piikauppa*. HYAL study material 30, University of Helsinki.

APPENDIX 1: GLOSSARY

Allele: one of two or more alternative forms (variants) of a gene or DNA sequence.

Allele frequency: the frequency of a particular allele at a genetic locus in a population. Allele frequencies are used to depict the amount of genetic diversity at the individual and population level.

Autosome: one of the 22 biparentally inherited (human) chromosomes, each present in two copies in each cell.

Effective population size (N_e): the number of individuals in a given generation, whose gametes contribute to next generations; almost always considerably lower than the actual census size.

Founder effect: a small number of individuals from a larger base population colonize a new site and no significant gene flow occurs thereafter between these two. The genetic consequences are very similar to a bottleneck.

F_{ST} : a measure of genetic distance among subpopulations.

Gamete: egg cell or sperm cell. In humans (or mammals) egg contains an X chromosome and sperm contains either an X or a Y chromosome.

Haplotype: the sequence of alleles of a set of polymorphic markers, located near each other on the same chromosome.

Linkage disequilibrium (LD): non-random association between closely linked loci due to their tendency to be co-inherited.

Mitochondrion: a cellular organelle primarily concerned with energy generation, contains its own circular genome (mtDNA), maternally inherited.

Nucleotide: The molecular component of the polymers DNA or RNA. DNA and RNA sequence length is measured in nucleotides (also known as base pairs, bp)

Population bottleneck: an event in which a considerable part of the population is prevented from reproduction. The population undergoes a considerable decrease in size (or the number of reproducing individuals), which may happen as a sudden incident or over a long time period. A sudden bottleneck may occur due to famine, epidemic, war or other reason. Only the survivors will pass their genes on to the next generations, which leads to reduction of genetic diversity compared to the situation which prevailed before the bottleneck.

Y chromosome: one of the sex chromosomes, only present in males (in mammals).