UPPSALA
UNIVERSITET

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology 2221

Divergence, admixture and continuity in the human past

Demographic inference using ancient and modern genomes

JAMES MCKENNA

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2022

ISSN 1651-6214
ISBN 978-91-513-1668-0
URN urn:nbn:se:uu:diva-489300

Dissertation presented at Uppsala University to be publicly examined in Lindahlsalen, EBC, Norbyvägen 16, Uppsala, Monday, 23 January 2023 at 13:15 for the degree of Doctor of Philosophy. The examination will be conducted in English. Faculty examiner: Associate Professor Ida Moltke (University of Copenhagen).

Abstract
McKenna, J. 2022. Divergence, admixture and continuity in the human past. Demographic inference using ancient and modern genomes. *Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology* 2221. 65 pp. Uppsala: Acta Universitatis Upsaliensis. ISBN 978-91-513-1668-0.

Demographic forces shaping the genetic variation we observe today can include population divergences, admixture events and continuity through time. The advancement of high-throughput sequencing technologies, together with developments in molecular and bio-informatics methods, mean the number of ancient genomes available for inference has risen steeply. To make effective use of aDNA however, inference tools need to be developed that account for temporal as well as geographic sampling of genomes. Here I have developed, evaluated and applied methods for estimating divergence times between ancient and modern populations. I used simulation to study the sensitivity of these approaches to violations of model assumptions, before applying them to study the history of population divergence between pairs of populations from a global panel. Non-tree-like demography is common in the human past, with evidence of ancestral structure in the form of archaic admixture in the genomes of all non-African modern humans. Using SNP-array data collected from 118 ethnic groups in the Philippines, I show that the highest levels of Denisovan ancestry are found among the Ayta Magbukon, further highlighting the complex history between modern human groups and the archaic hominins occupying Eurasia before our arrival. Among the most important contributions population genetics has made to the study of the human past is the demonstration that cultural transitions and spread of technologies were often associated with migrating groups of people. This can result in the admixture, displacement or replacement of populations, and aDNA provides us with the opportunity to assess these trends directly through time. I developed a statistical tool to detect population continuity through time, evaluating its performance using simulation. Applied to a dataset of ancient genomes from Early Neolithic Scandinavia, I demonstrate population continuity in the hunter-gathering Pitted Ware culture, despite these people overlapping both geographically and temporally with farmers of the Funnel Beaker culture. In another study of the hunter-gatherer ancestors of the San people of southern Africa, I show evidence that this group exhibited long-term population isolation, remaining unaffected by admixture from outside southern Africa until surprisingly recent times. Using these ancient genomes, I provide further evidence that all modern Khoe-San populations exhibit significant levels of admixture with people of non-Khoe-San ancestry, demonstrating the strong impact migrations in this region have had in the past ~2,000 years.

Keywords: Human evolution, population genetics, ancient genomes, population continuity, divergence

James McKenna, Department of Organismal Biology, Human Evolution, Norbyvägen 18 A, Uppsala University, SE-752 36 Uppsala, Sweden.

ISSN 1651-6214
ISBN 978-91-513-1668-0
URN urn:nbn:se:uu:diva-489300 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-489300)

“Our Pioneers keep striking
Inwards and downwards,
Every layer they strip
Seems camped on before."
Seamus Heaney - Bogland

List of papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.

I Sjödin, P., **McKenna, J.,** Jakobsson M. (2021) Estimating divergence times from DNA sequences. *Genetics*, 217(4)

II Larena, M.*, **McKenna, J.*,** Sanchez-Quinto, F., Bernhardsson, C., Ebeo, C., Reyes, R., Casel, O., Huang, JY., Hagada, KP., Guilay, D., Reyes, J., Allian, FR., Mori, V., Azarcon, LS., Manera, A., Terando, C., Jamero Jr, L., Sireg, G., Manginsay-Tremedal, R., Labos, MS., Vilar, RD., Latiph, A., Saway, RL., Marte, E., Magbanua, P., Morales, A., Java, I., Reveche, R., Barrios, B., Burton, E., Salon, JC., Kels, MJT., Albano, A., Cruz-Angeles, RB., Molanida, E., Granehäll, L., Vicente, M., Edlund, H., Loo, JH., Trejaut, J., Ho, SJW., Reid, L., Lambeck, K., Malmström, H., Schlebusch, C., Endicott, P., Jakobsson, M. (2021) Philippine Ayta possess the highest level of Denisovan ancestry in the world. *Current Biology*, 31:4219-4230

III **McKenna, J.,** Bernhardsson, C., Waxman, W., Jakobsson, M., Sjödin, P. (2022) Detecting population continuity and ghost admixture among ancient genomes. *biorxiv* (*submitted*)

IV **McKenna, J.*,** Hollfelder, N.*, Vicente, M., Edlund, H., Coutinho, A., Bernhardsson, C., Sjödin, P., Brink, J., Zipfel, B., Malmström, H., Lombard, M., Schlebusch, C., Jakobsson, M. (-) Ten thousand years of ancient genomes show long-term isolation of Stone Age southern Africans. (*manuscript*)

* These authors contributed equally to this study

I am also coauthor on the following article published during my graduate studies.

> Larena, M., Sanchez-Quinto, F., Sjödin, P., **McKenna, J.,** Ebeo, C., Reyes, R., Casel, O., Huang, JY., Hagada, KP., Guilay, D., Reyes, J., Allian, FP., Mori, V., Azarcon, LS., Manera, A., Terando, C., Jamero Jr, L., Sireg, G., Manginsay-Tremedal, R., Labos, MS., Vilar, RD., Latiph, A., Saway, RL., Marte, E., Magbanua, P., Morales, A., Java, I., Reveche, R., Barrios, B., Burton, E., Salon, JC., Kels, MJT., Albano, A., Cruz-Angeles, RB., Molanida, E., Granehäll, L., Vicente, M., Edlund, H., Loo, JH., Trejaut, J., Ho, SYW., Reid, L., Malmström, L., Schlebusch, C., Lambeck, K., Endicott, P., Jakobsson, M. (2021) Multiple migrations to the Philippines during the last 50,000 years. *PNAS*, 118(13)

Abbreviations

DNA	Deoxyribonucleic acid
aDNA	Ancient DNA
mtDNA	Mitochondrial DNA
BP	Before Present
calBP	Calibrated Before Present
kya	Thousand years ago
LGM	Last Glacial Maximum
SE	Standard error
bp	Base pairs
LD	Linkage Disequilibrium
PCA	Principal component analysis
SNP	Single nucleotide polymorphism

Contents

1. Introduction

"We carry within us the wonders we seek around us."
Sir Thomas Browne

When English polymath, Thomas Browne, wrote those words in the 17th century, he could not have known just how prophetic they were. Even Charles Darwin, publishing his "On the Origin of Species" (Darwin, 1859) over 200 years later, would still not understand the precise mechanism through which variation is passed through the generations allowing the full diversity of life to evolve. In the 19th century, the most widely accepted hypothesis was that the continuity of variation we observe in populations must arise from some similarly continuous basis, with offspring showing a blended inheritance of parental forms. Opponents of such a model of inheritance were quick to point out a seemingly critical flaw; any beneficial trait that arises would be diluted at a rate far exceeding natural selection's opportunity to favour it. It took the work of an Augustinian friar, Gregor Mendel, cross-breeding varieties of round and wrinkled peas in his monastery gardens, to demonstrate the first convincing evidence that the factors underlying parental traits are not blended during reproduction, but rather maintain their integrity across generations (Mendel, 1866). In the early 20th century, Morgan's experiments with mutant variants of Drosophila fruit flies (Morgan et al., 1915) lent additional support to what became known as "particulate", or Mendelian inheritance, showing that discrete Mendelian factors are capable of generating continuous variation in populations.

Despite mounting experimental evidence, at this time many in the physical sciences considered the largely narrative driven field of evolutionary biology to consist of little more than unscientific speculation about history. This began to change with the publication by Fisher in 1918 of a paper entitled "The Correlation between Relatives on the Supposition of Mendelian Inheritance" (Fisher, 1918). Here Fisher introduced the "infintismal model", showing mathematically how natural selection acting on a number of independently segregating alleles could result in the appearance of continuous variation in populations. This early work, along with major contributions by others including Wright, Haldane and Dobzhansky, was instrumental in resolving the decades-old puzzle of how Mendelian inheritance was consistent with Darwinian evolution by natural selection. It laid the foundations for the new field of population genetics, a discipline that firmly embraced the mathematical modelling and

controlled experimentation more common to the physical sciences. This ensured that the study of evolution was no longer so much unverifiable speculation, but was now grounded in a framework allowing it to become predictable, measurable and testable.

1.1 A new era

It is remarkable to consider that these early pioneers had established much of the theoretical underpinnings of population genetics before it was even suggested that deoxyribunucleic acid (DNA) was the molecule responsible for the transmission of genetic information (Griffith, 1928; Avery, MacLeod, and McCarty, 1944). It would be decades before the structure of DNA was discovered to be a macromolecule consisting of two strands, coiling around each other to form a double helix (Watson and Crick, 1953). Each strand is comprised of simpler units called nucleotides, with their specific sequence encoding the instructions necessary for the development and reproduction of all known living organisms. By the 1970s technologies such as "Sanger sequencing" had enabled researchers to determine the order of nucleotides (Sanger, Nicklen, and Coulson, 1977), resulting shortly thereafter in the sequencing of the first genome (bacteriophage ϕX174 (Sanger et al., 1977)). Rapid technological developments through the 1990s paved the way for several new "second-generation", or "next-generation" (NGS) sequencing methods. Highly scalable, these technologies typically rely on the fragmentation of a whole genome into shorter stretches of sequence (also called reads), before sequencing these reads in a massively parallel fashion. In this way entire genomes could be sequenced at once.

By 2001, drafts of the human genome were freely available (Lander et al., 2001; Venter et al., 2001). As sequencing technologies have continued to improve and the cost of sequencing has drastically reduced (figure 1.1), increasing numbers of genomes have been sequenced, with their reads mapped to the human reference genome and "variant" positions identified (Wheeler et al., 2008). Today, large-scale genome sequencing projects such as the 1000 Genomes Project (Siva, 2008) and Simons Genome Diversity Project (Mallick et al., 2016) describe genetic variation present among hundreds of human genomes sampled from diverse populations around the globe. From the handful of allozyme markers available to population geneticists fifty years ago, to the thousands of full genomes available today, there has never been richer opportunity to study patterns of genetic variation within and among populations, and especially to make inferences on the demographic history of those populations.

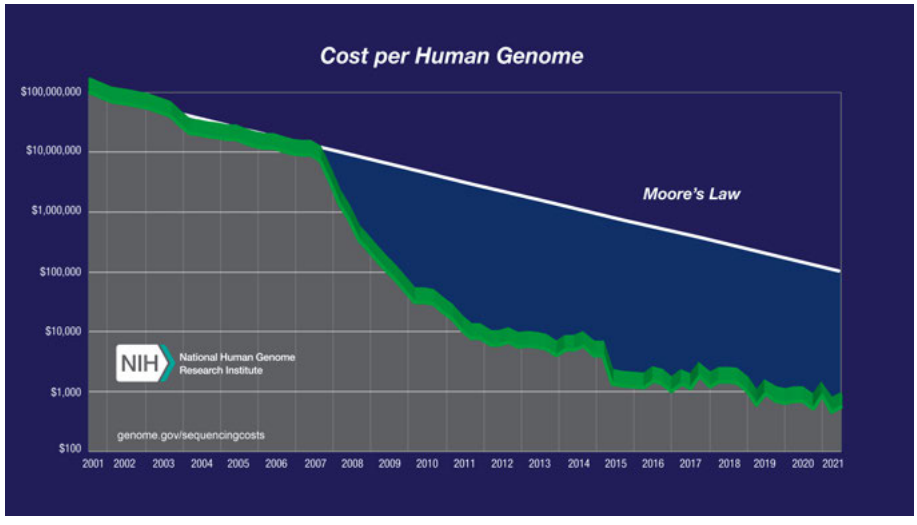


Figure 1.1. The faster-than-exponential reduction in approximate costs of sequencing a human genome since 2001. (Image from National Human Genome Research Institute (NHGRI) is in the public domain).

1.2 ancient DNA

The rapid pace of development in NGS discussed above also revolutionized the capacity to retrieve DNA from ancient remains (Stoneking and Krause, 2011; Llamas, Willerslev, and Orlando, 2017). Since the first extraction and successful sequencing of DNA from the preserved remains of a quagga (a South African Zebra, hunted to extinction in the late 19th century) in 1984, ancient DNA (aDNA) studies have hugely increased our knowledge of past and present genetic diversity. Many complete genomes of extinct species have been fully sequenced, from the passenger pigeon (Murray et al., 2017) to the woolly mammoth (Palkopoulou et al., 2015). For population geneticists, fundamentally interested in questions surrounding the evolutionary forces shaping modern populations, aDNA offers windows into evolutionary time, direct snapshots of patterns of genetic diversity in different places and times in history. Ancient DNA studies have repeatedly revealed aspects of human history that would have been impossible to infer using samples of modern genomes alone (Llamas, Willerslev, and Orlando, 2017). This is because large-scale migrations of people throughout human history has often led to admixtures and population replacements that effectively obscure ancient populations from us. For instance, the first complete genome from an ancient human was successfully sequenced from the permafrost-preserved hair of an individual in Greenland $\sim$4 kya (Rasmussen et al., 2010). This sample showed strong genetic affinity to present-day north-east Siberian people rather than present-day Inuit and Native Americans of the New World Arctic, revealing a previously

undetected ancient migration into the New World. Shortly after this study, draft genomes of two archaic hominins were published (Green et al., 2010; Reich et al., 2010). Not only did these Neanderthal and Denisovan genomes reveal the existence of previously unknown branches of the human tree, and that these archaic hominins lived contemporaneously to anatomically modern humans (AMH) in Eurasia $\sim$50 kya, but they also provided the first genetic evidence of admixture between archaic hominins and our species. Since this time, aDNA studies have continued to refine and deepen our understanding of the demographic changes and evolutionary forces that have contributed to modern humans. Despite this great utility, there are major challenges associated with the retrieval and processing of aDNA.

1.2.1 aDNA damage

As soon as an organism dies, the enzymatic repair mechanisms that maintain the integrity of DNA in the living cell cease to function. Microorganisms spread through decaying tissues and intracellular nucleases begin to degrade DNA (Pääbo, 1989; Willerslev and Cooper, 2005; Sawyer et al., 2012). Hydrolytic depurination and β-elimination break the backbone of the DNA molecule, causing it to quickly become fragmented to ultra-short lengths of typically $<$100 bp (Dabney et al., 2013). A host of post-mortem chemical alterations to the DNA structure and sequence act to drastically reduce the amount of retrievable DNA. One of the most important of these is the deamination of cytosine nucleotides into uracil, which is in turn read as thymine during DNA sequencing (Pääbo, 1989; Briggs et al., 2010; Sawyer et al., 2012), leading to artefactual C to T transitions in aDNA (and G to A in double stranded libraries). These misincorporations are more prevalent at the ends of aDNA fragments, likely due to the exposed overhanging single strands (Dabney et al., 2013). It is possible to reduce the impact of artefact C to T transitions in aDNA samples through treatment of DNA extracts with uracil-DNA-glycosylase (UDG), which removes uracil before the building of DNA libraries (Briggs et al., 2010; Briggs and Heyn, 2012).

Although aDNA damage correlates with the age of the sample, these damage patterns have been found in samples only a couple of decades old (Sawyer et al., 2012; Orlando et al., 2021). Ultimately it is local environmental conditions that primarily contribute to levels of aDNA damage. Environments with frequently changing temperatures, hot and humid conditions, and acidic or alkaline soil pH have all been associated with an increase in post-mortem modifications in the DNA of ancient remains. Since aDNA preservation varies with environmental conditions, many regions are under-represented in aDNA studies. This is particularly true of the hot and humid equatorial regions, or much of sub-Saharan Africa which is characterized by acidic soils. Although the damage patterns characteristic of aDNA were initially seen as a problem

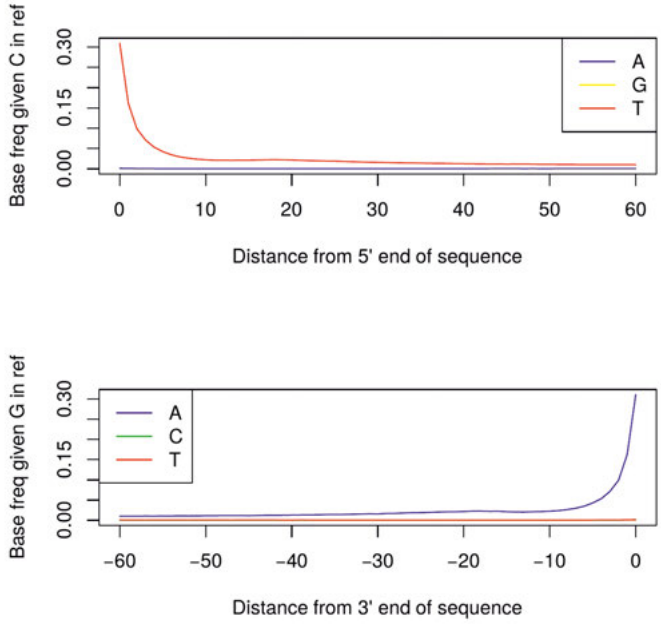


Figure 1.2. Damage plot for genetic data of ancient southern African individual "flo011", dating to 5600-5480 calBP from Matjes River Rock Shelter, South Africa. Sharp increases in deamination-caused C-T nucleotide misincorporations on the 5' end of DNA fragments, and increases in A on the 3' end, characteristic of genuinely ancient DNA sequence.

since they may be misinterpreted as genuine genetic variation, as we will see in the forthcoming sections, they have been instrumental for authenticating endogenous ancient DNA sequences.

1.2.2 Contamination

One result of post-mortem modifications is that lower quantities of DNA are present in ancient tissues compared to modern samples. This means that even tiny amounts of present-day DNA can exist in greater concentration than endogenous DNA. This contaminating DNA can come from a number of sources; microorganisms in the soil surrounding the sample, the hands of the archaeologist or museum staff handling the remains, even on laboratory surfaces and reagents used during DNA extraction and sequencing. For this reason, the first step in combating contamination is to observe laboratory protocols that reduce the likelihood of further contamination (Dabney et al., 2013; Orlando et al., 2021). Commonly observed procedures include the wearing of protective clothing, use of negative pressure rooms and strict cleanroom protocols at each step of DNA extraction. Certain tissues from ancient remains are preferentially sampled due to the increased likelihood of obtaining high

amounts of endogenous DNA; in particular the petrous bone, tooth cementum and dentine. Prior to DNA extraction, the surfaces of samples are treated with denaturing compounds or UV light to remove potential contaminants, and mechanically ground bone powder usually sampled from the inner layers of the remains. Despite these extensive precautions, contamination is often present in aDNA samples and has the potential to severely bias downstream genetic analyses.

1.2.3 Downstream analyses

Due to the high risk of contamination, it is essential to establish authenticity of endogenous DNA in ancient samples. A typical first step is to align sequenced DNA to the human reference genome in order to exclude sequences from non-human contaminating sources. However, filtering based on sequence divergence does not exclude contamination from modern or ancient humans. Various bioinformatic and computational tools exist to quantify post-mortem damage patterns and estimate levels of contamination. Because the rate of cytosine deamination increases towards the ends of aDNA fragments (figure 1.2), elevated frequencies of C to T transitions in non-UDG treated libraries indicates the presence of genuinely ancient DNA (Krause et al., 2010; Dabney et al., 2013; Skoglund et al., 2014). Contamination levels in a sample can also be assessed using post-mortem damage patterns (Renaud et al., 2015; Meyer et al., 2016), or from deviations in expected ploidy using mitochondrial data (Fu et al., 2013), the X-chromosome in males (Rasmussen et al., 2011; Rasmussen et al., 2015; Moreno-Mayar et al., 2020) and autosomal DNA (Green et al., 2010; Nakatsuka et al., 2020; Racimo, Renaud, and Slatkin, 2016). Another issue concerns population genetic methods themselves. Many techniques commonly applied to datasets that include aDNA sequences have often not been created with temporally distributed DNA samples in mind. For this reason the analysis of such datasets can sometimes lead to contradictory or misleading results. The genetic changes that are expected to have occurred in the time separating samples from ancient and modern populations can make ancient samples seem more genetically similar to each other, and more distant to modern samples (Skoglund et al., 2014). However the surge in available ancient genomes in recent years has motivated the development of increasing numbers of tools for demographic inference that explicitly account for the expected differences between ancient and modern samples (Racimo, Renaud, and Slatkin, 2016; François et al., 2019; François and Jay, 2020; Sjödin, Skoglund, and Jakobsson, 2014; Schraiber, 2017).

1.3 Inferring human history

1.3.1 Human origins in a nutshell

The study of human origins combines evidence from paleoanthropology, archaeology, linguistics and genetics. A cross-disciplinary approach has resulted in a broad consensus that anatomically modern humans (AMH) evolved in Africa (Conroy and Pontzer, 2012; Harpending and Rogers, 2000; Atkinson, 2011). Most evidence concerning the deep history of our species is found in the fossil record, with our own genus, *Homo*, appearing $\sim$2 mya (Wood and Collard, 1999). *H. ergaster*/*H. erectus* is characterized by distinct anatomical features including flatter faces, smaller teeth and changes in body size and shape suggesting a more erect posture than its predecessors. The early discovery of fossils such as Peking Man in China, or Java Man in Indonesia, show that *H. erectus* was also present in Eurasia from as early as $\sim$1.8 mya (Lordkipanidze et al., 2007). It is possible that its more gracile descendent *H. heidelbergensis* evolved in Africa $>$1 mya and gave rise to both modern humans and Neanderthals. Alternatively, an earlier split in the *Homo* phylogenetic tree may have given rise to a Eurasian branch (*H. heidelbergensis*), and an African branch (*H. rhodesiensis*), evolving into Neanderthals and modern humans respectively.

Although there is now broad acceptance that modern humans evolved in Africa, the specific location of *where* AMH evolved within Africa remains a topic of active debate (Jobling, Hurles, and Tyler-Smith, 2019; Scerri et al., 2018). The relative abundance of hominin fossils from East African sites historically suggested East Africa as the place of origin. However, more recent discoveries of early modern human remains dating up to 300 kya at Middle Stone Age sites as distant as Jebel Irhoud, Morocco (Hublin et al., 2017), Herto, Ethiopia (White et al., 2003) and Klasies River, South Africa (Lam, Pearson, and Smith, 1996), have led to suggestions of a "mosaic" model of human evolution, with the history of modern humans characterised by deep ancestral structure (Scerri et al., 2018; Ragsdale et al., 2022). Others argue that the genetic evidence is more in favour of a single origin of AMH, combined with either admixture from an unsampled archaic species within Africa (Durvasula and Sankararaman, 2020), or an ancient "back-to-Africa" migration introducing archaic hominin ancestry related to *H. erectus* into some African populations (Chen et al., 2020; Cole et al., 2020).

The "mosaic vs single origin" theories of human evolution within Africa echoes an earlier contentious debate in paleoanthropology; whether modern humans in Eurasia evolved under a "multiregional" model or an "Out-of-Africa" model. Proponents of the former hypothesized that the transition from *H. erectus* to *H. sapiens* took place in several regions of the Old World, including Europe and Asia. Proponents of the latter suggested this transition took place within Africa, with a subsequent spread of modern humans outside Africa at some

point ∼100 kya, replacing the archaic hominin species previously inhabiting Eurasia. Genetic evidence proved instrumental in resolving this question. The demonstration that sub-Saharan African populations have the highest genetic diversities of all modern human groups, combined with the fact that genome-wide homozygosities and linkage disequilibrium increases with geographic distance from Africa, strongly supports the Out-of-Africa model (Rosenberg et al., 2002; Jakobsson et al., 2008; Li et al., 2008; Prugnolle, Manica, and Balloux, 2005; Ramachandran et al., 2005). It is likely that population genetics will prove similarly useful in resolving questions regarding the evolutionary history of modern humans within Africa.

Our understanding of hominin diversity between 50-500 kya has continued to develop. The exciting discovery of remains of small hominin species on the islands of Flores, Indonesia (*H. florensiensis* (Brown et al., 2004)) and Luzon, in the Philippines (*H. luzonensis* (Détroit et al., 2019)), suggests surprising recent diversity among archaic hominins in Eurasia. The sequencing of the remains of an unknown juvenile hominin excavated in Denisova Cave, Siberia, and dated to 74-82 kya, revealed the existence of yet another archaic hominin distinct from Neanderthals (Meyer et al., 2012; Reich et al., 2010). It has been suggested that all these archaic hominins can trace their ancestry to *H. erectus*. Though the details of their phylogenetic relationships to *H. erectus*, or indeed to each other, are still far from clear, ancient DNA studies have been instrumental in providing evidence of archaic admixture into modern human populations.

1.3.2 Archaic admixture

Paleontologists first detected evidence of archaic admixture into modern humans with the puzzling discovery that a ∼25 kya skeleton of a boy from the Iberian peninsula exhibited a mixture of Neanderthal and modern human morphologies (Duarte et al., 1999). However, it was the sequencing of the Neanderthal genome in 2010 that provided the strongest evidence of such archaic introgression into modern humans (Green et al., 2010). Allele-sharing tests demonstrated that all modern non-African populations share ∼2.5% of their genome with Neanderthals, indicating gene-flow between the ancestors of all non-Africans and Neanderthals. Green *et al* (2010) suggested the most parsimonious explanation for this shared Neanderthal-related ancestry among modern non-Africans was an admixture event occurring between these two groups in the Middle East, sometime after the Out-of-Africa migration.

Not long after the publication of these findings, the discovery of an unknown archaic hominin with the sequencing of the Denisovan genome (figure 1.3) provided additional evidence that present-day Australasian populations harbour up to 5% of Denisovan-related ancestry not shared with European or African populations. Again, this is most likely explained by an admixture

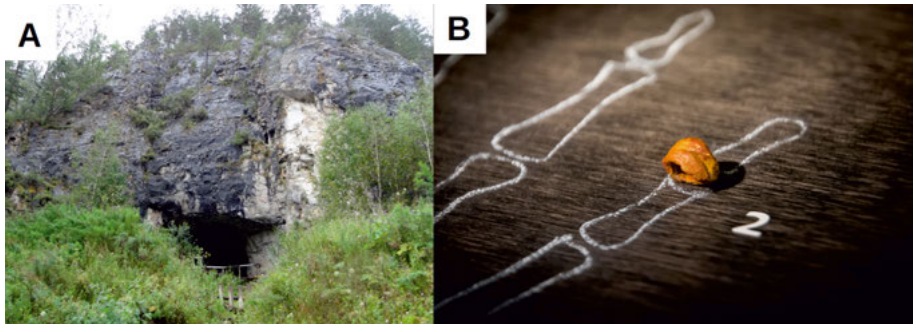


Figure 1.3. (A) Denisova Cave in the Altai Mountains, Siberia, where the first reported remains of Denisovans were found. (B) Replica bone fragment of fifth distal finger phalanx from which ancient DNA was extracted, providing the first whole genome sequence of a Denisovan (Reich et al., 2010), dated to 76.2-51.6 kya (Douka et al., 2019). (Images distributed by Thilo Parg/Wikipedia Commons under licence CY-BY-SA-3.0).

event between the ancestors of these modern populations and a Denisovan or Denisovan-like hominin (Reich et al., 2010; Reich et al., 2011; Larena et al., 2021). However the large geographic distance between the location of these modern day Australasian populations and Denisova Cave in Siberia raised the obvious question of when, where, and how often admixture between Denisovans and modern humans took place. In fact a much more complex admixture history has been suggested, with Denisovan ancestry detected (though to lesser degrees) among present-day Siberian, Native American and South Asian populations, and sequenced Denisovans likely representing a deeply structured population (Sankararaman et al., 2016; Skoglund, Götherström, and Jakobsson, 2011; Qin and Stoneking, 2015; Prufer et al., 2014; Browning et al., 2018). Recent work has provided evidence that the ancestors of modern human populations admixed with Denisovans at least twice in the past (Browning et al., 2018), and that some of that ancestry is exclusively found among Papuan-related populations (Jacobs et al., 2019).

1.3.3 Neolithic Scandinavia

The genetic diversities of contemporary Europeans can be modelled as mixtures of three main ancestries. These components are populations of hunter-gatherers inhabiting Europe after the Last Glacial Maximum (LGM, ∼15-5 kya), an expansion of agriculturalists from present-day Anatolia spreading the "Neolithic package" of agriculture and domestication throughout Europe (∼9-3.5 kya), and a migration of pastoralist peoples from the Yamnaya culture of the Pontic steppe (∼5 kya) (Lazaridis et al., 2016; Haak et al., 2015; Allentoft et al., 2015). Scandinavia was the last region of Europe to become free of ice with the retreat of the Fenno-Scandian ice sheet towards the end

of the LGM ∼11.7 kya. It also harboured some of the last remaining populations of hunter-gatherers in Europe, with the arrival of the Funnelbeaker culture (TRB) to southern Scandinavia ∼5.5 kya signalling a relatively late emergence of Neolithic material culture in the region (Midgley, 1992; Lillie et al., 2000; Skoglund et al., 2012; Fraser et al., 2018). Studies of ancient genomes reveal clear genetic differences between these early farmers and earlier Scandinavian hunter-gatherers, with strong genetic affinity to other Early and Middle Neolithic farmers of Europe (Malmström et al., 2015; Malmström et al., 2019).

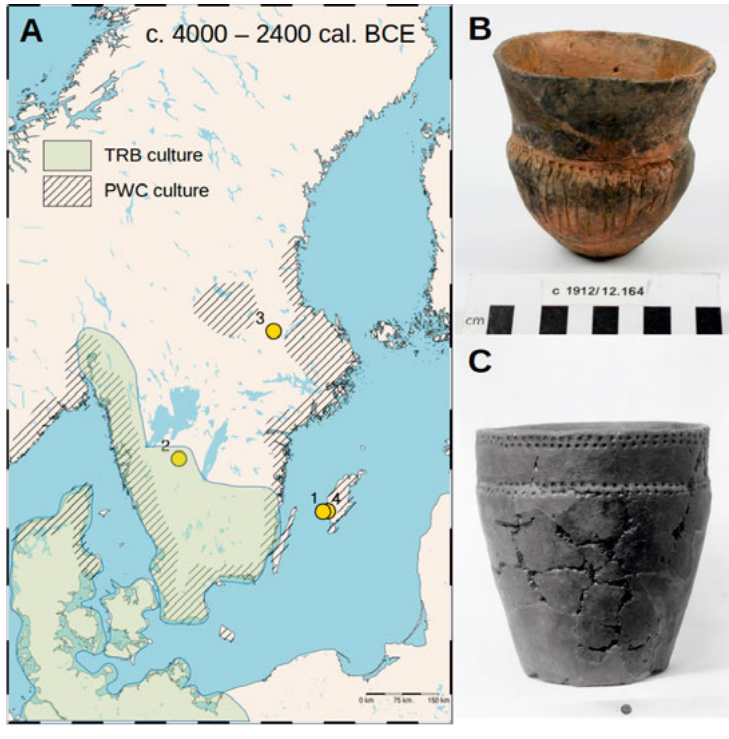


Figure 1.4. (A) Map of Early Neolithic Scandinavia, showing regions inhabited by Pitted Ware culture (PWC) and Funnelbeaker culture (TRB) groups. Markers depict sampling sites of ancient individuals used in Paper III. (B) Ceramic pot with "funnel-shaped" neck characteristic of TRB culture, and (C) ceramic pot with "pits" characteristic of the PWC.

The sudden emergence of the Pitted Ware Culture (PWC) ∼5 kya along the coastal regions of southern Scandinavia and islands of the Baltic Sea has long mystified researchers. Named after their ceramic style of pitted pots (figure 1.4), these people exhibited a hunter-gathering economy; seal hunters with dietary patterns and funerary customs distinct from the contemporaneous TRB people (Malmström et al., 2009; Fraser et al., 2018). Genetic studies have shown that unlike the TRB, the PWC were largely descendent from the earlier

Mesolithic hunter-gatherer populations of Scandinavia (Skoglund et al., 2012; Skoglund et al., 2014). The expansion of farming peoples throughout Europe during the Neolithic was typically characterized by major cultural transformation and the repeated absorption or replacement of previous hunter-gathering populations. Yet in southern Scandinavia, archaeological and genetic evidence demonstrates that despite overlapping, these two groups maintained some degree of reproductive isolation for several hundred years, resulting in distinct patterns of genetic diversity and material cultures (Malmström et al., 2009; Fraser et al., 2018; Coutinho et al., 2020; Malmström et al., 2019). An offshoot of the Corded Ware culture, the Battle Axe culture (BAC), later emerged in southern Scandinavia $\sim$4.8 kya. This group absorbed and replaced the agricultural TRB, continuing to coexist alongside the PWC for some centuries before eventually absorbing the population. The fusion of Battle Axe and Pitted Ware cultures is considered to have led to the development of the Nordic Bronze Age in Scandinavia.

1.3.4 Southern Africa in the Holocene

Despite Africa being the birthplace of humanity, and sub-Saharan African populations holding the greatest genetic diversities of all humans groups, Africa is extremely underrepresented in genetic studies (Need and Goldstein, 2009). This heavy Euro-centric bias is evidenced by the fact that individuals of African ancestry represent $<$3% participation in genome-wide association studies (Martin et al., 2018). The deepest divergence events in the tree of modern humans are captured by sub-Saharan hunter-gatherer populations including the rainforest hunter-gatherers of central Africa and the Khoe-San of southern Africa (Tishkoff et al., 2009; Gronau et al., 2011; Gronau et al., 2011; Schlebusch et al., 2012; Veeramah et al., 2012). Present-day Khoe-San populations live in South Africa, Botswana, Angola and Namibia. They are grouped together on the linguistic basis that they speak languages characterized by "click-consonants", though these languages are only distantly related to each another (Guldemann, 2008). Research has shown that populations within the Khoe-San are genetically distinct, and can be loosely grouped as the traditionally hunter-gathering San and the traditionally pastoralist Khoe peoples (speakers of Khoekhoe branch of the Khoe-Kwadi language family languages). Archaeological evidence suggests pastoralism reached southern Africa $\sim$2 kya (Pleurdeau et al., 2012; Sadr, 1998; Smith, 1992), and studies of ancient genomes from South Africa have demonstrated that the Khoe were formed with the arrival of pastoralist herders of East African ancestry in southern Africa $\sim$1.3 kya, admixing with the ancestors of modern San populations (Coutinho et al., 2020; Vicente et al., 2019). This admixture predates the later arrival of Bantu-speaking peoples of West African ancestry in southern Africa, a migration associated with agriculture and iron-working techologies. This, together

with the arrival of European colonists in the 16th century, had tremendous demographic impacts on the ancient hunter-gatherer population of southern Africa, displacing them from much of their former homeland. In 2017 the remains of a Late Stone Age hunter-gatherer from Ballito Bay were sequenced (Schlebusch et al., 2017). This individual predated the large-scale movements of people of the past 2 millenia, and provided incontrovertible evidence that all modern Khoe-San groups exhibit relatively high levels of admixture with non-Khoe-San populations. Even the Ju|'hoansi, often used as an outgroup in studies of human evolution due to their genetic distance to all other populations, were shown to possess significant admixture from sources of East African ancestry. One important result of this finding was to push back the time of the oldest divergence in the tree of modern humans to $\sim$200-300 kya. The history of the Khoe-San in southern Africa still leaves many outstanding questions: how were the hunter-gathering ancestors of the San related to present-day populations? How long were those ancestors inhabiting southern Africa? How were the ancient hunter-gatherers of southern Africa related to other hunter-gatherers sequenced from other regions of Africa (e.g. Shum-Laka in West Africa $\sim$8 kya and the genome of an ancient hunter-gatherer from Mota, Ethiopia, dated to $\sim$4.5 kya (Lipson et al., 2020; Lipson et al., 2022; Llorente et al., 2015; Skoglund et al., 2017)). Just as aDNA has proven to be an invaluable resource in reconstructing human history in Eurasia, so ancient genomes still have critical role to play in deepening our understanding of the ancient past in Africa.

2. Methods

2.1 Genetic variation

Population genetics makes use of the variation present within and between populations. This genetic variation arises as a result of errors in DNA replication during meiosis. When not repaired by the cell's DNA repair mechanisms, these errors (now mutations) are passed on to the gametes and offspring. There are many possible types of mutation, ranging from the very large (chromosomal duplications), through medium-sized (structural variants such as insertions, deletions, translocations and copy number variants), to the smallest mutations possible, Single Nucleotide Polymorphisms (SNPs). SNPs are variants at a single base pair position, and in this thesis, I focus only on the genetic variation arising from SNPs. SNPs are the most common form of mutation, occurring in the human genome about once every 1000 base pairs (Prado-Martinez et al., 2013). This is largely because smaller errors are less likely to cause significant developmental problems for the offspring inheriting them. SNPs are particularly useful as genetic markers because of their relatively low mutation rate. This allows us to make assumptions that simplify demographic inference, including that a mutation at a site can only have occurred once. The average mutation rate of SNPs in the human genome has been subject to revision - when estimated from the number of mutational differences between chimpanzees and humans (and a split of 6-7 mya based on the fossil record), the rate is estimated to be 2.5×10^{-8} (Ségurel, Wyman, and Przeworski, 2014). When estimated instead from the number of observed mutations across generations in pedigree studies, the mean mutation rate drops to $\sim1.25\times10^{-8}$ to 1.5×10^{-8} (Scally and Durbin, 2012).

There are different methods available for measuring genetic diversity. As discussed above, the rapid advancement of high-throughput sequencing technologies has resulted in a steep reduction in cost of whole genome sequencing. While the sequencing of the human reference genome in 2001 had a "cost-per-genome" of $\sim$\$100,000,000, the cost today now stands below \$1,000 (National Human Genome Research Institute, 2021). Despite this reduction, the cost of whole genome sequencing may prove prohibitive for studies involving thousands of genomes. SNP-arrays represent a powerful and much more economic method of quantifying genetic diversity in populations (Gibbs et al., 2003; Novembre and Ramachandran, 2011). They consist of an array of SNPs genotyped in a panel of individuals. In an effort to capture as much diversity among populations as possible, SNPs are often selected to have intermediate

frequencies in the panel. Two issues arise from this which can cause bias in downstream population genetic analyses. Firstly, if the sample of individuals used in the panel is not a random representation of the population as a whole, ascertainment bias can occur (Clark et al., 2005). This results in genotyped individuals showing greater diversity when they are from a population similar to those represented in the panel. Genotyped individuals of more diverged populations will look less diverse, and variation at sites in those individuals not present in the SNP-array will not be detected at all (Pagani et al., 2015). This is the case for African populations, heavily underrepresented in many SNP-arrays, resulting in African samples showing lower genetic diversities in some studies (Lachance and Tishkoff, 2013), despite whole genome sequencing revealing the opposite to be true. Secondly, because intermediate frequency SNPs are specifically selected for the array, there is a bias against rare alleles (Albrechtsen, Nielsen, and Nielsen, 2010). These alleles have been shown to be particularly informative about fine structure in populations and recent demographic history (Gravel et al., 2011; Mathieson and McVean, 2014), information that can be lost in SNP-array studies. Despite these disadvantages, SNP-arrays can be an economical alternative to whole genome sequencing, particularly suited to assaying levels of genetic diversity in many individuals across many populations. In Paper II for instance, 118 ethnic groups from the Philippines were genotyped on an array of $\sim$2.3 million SNPs, enabling us to assess a breadth of genetic diversity across a large set of populations that would not have been economically feasible using whole genome sequencing.

2.2 Genetic drift and N_e

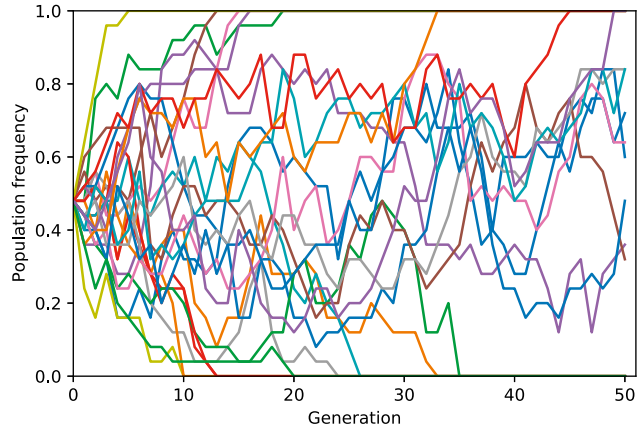
Before we can examine the methods used in this thesis to explore patterns of diversity among ancient and modern populations, we first need to understand some key concepts in populations genetics. The first of these is *genetic drift*. Once a mutation gives rise to a variant at a particular locus in the genome, there now exist two forms of that locus in the population, also called alleles. Genetic drift describes the random fluctuation in an allele's frequency over time that results from random sampling in finite populations. Drift can lead to the eventual fixation or loss of genetic variants, and therefore a loss of genetic diversity within a population. This was an important finding of classical population genetics (Fisher, 1922; Wright, 1931); that even in the absence of natural selection favouring one allele over another, genetic diversity will be lost in finite populations. In reality, this loss in diversity is countered by the process of mutation, which continuously produces new genetic variants. The so-called Wright-Fisher model makes several simplifying assumptions to ensure that genetic drift is the only evolutionary force affecting allele frequencies, with random mating, no mutation, no migration, and

discrete, non-overlapping generations. This simple model represents an extremely useful tool for assessing the effect of various evolutionary forces on population allele frequencies. In particular, the model makes it apparent that the rate at which genetic diversity is removed from a population is inversely proportional to population size (Kimura and Ohta, 1969). Simulations can be used to show that the rate of genetic drift is much greater in small populations, (figure 2.1). This early theoretical expectation was borne out in classic experimental work by Buri (1956) who demonstrated strong fluctuations in allele frequencies in small populations of Drosophila fruit files. An empirical example closer to home is the reduced genetic diversity in non-African human populations. This loss of diversity is thought to have occurred as a result of the population bottleneck, or series of bottlenecks, associated with the Out-of-Africa event (Jakobsson et al., 2008; Henn, Cavalli-Sforza, and Feldman, 2012; Schlebusch et al., 2020).

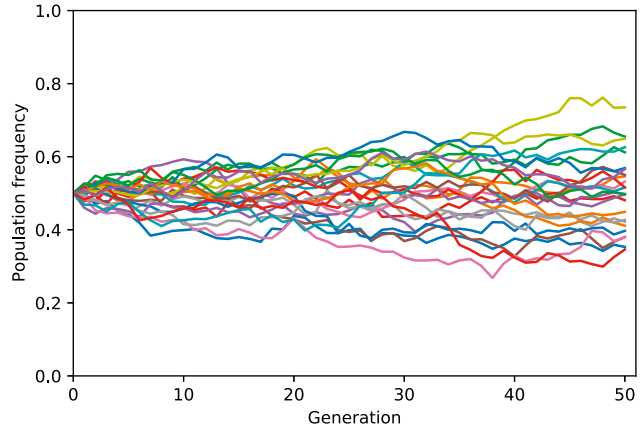
The population of the Wright-Fisher model is a highly idealized one following several simplifying assumptions. For example, it has equal numbers of males and females, individuals mate randomly, and there is no variation in fitness. All individuals in a generation are born together, reproduce and die simultaneously, with no changes in population size. These ensure that genetic drift is the only evolutionary force driving changes in allele frequencies. Clearly, however, it would be rare for a natural population to behave in precisely the same manner as a Wright-Fisher population. The fact that a real population can differ considerably in its dynamics from an idealized population led Wright to develop the concept of an "effective population size" (N_e) (Wright, 1932). In Wright's original definition, the N_e of a real population is the size of the idealized population with the same changes in allele frequency from one generation to the next. A more general definition would be that the N_e of a real population is the number of individuals a Wright-Fisher population would have, if it displayed the same level of genetic diversity as the real population. It then becomes clear that the rate of genetic drift need not be inversely proportional to a real population's census population size. Rather any process that increases the variance in reproductive success in a natural population (including unequal sex ratios, overlapping generations, selection, changes in population size) will decrease effective population size and therefore increase the rate of genetic drift.

2.3 Measuring genetic diversity

As the previous section makes clear, the genetic diversity of human populations will have been influenced by their different demographic histories. Major historical migrations of people, geographic distance and barriers, varying levels of admixture between neighbouring groups; all impact levels of genetic differentiation. A major goal of population genetics is to use these patterns



(a) Population size = 25



(b) Population size = 1000

Figure 2.1. Simulating random genetic drift under the Wright-Fisher model, with 25 replicates per simulation and an initial allele frequency of 0.5. The trajectory of frequency changes for a single given allele is measured across 50 generations. The effect of genetic drift is shown for populations of a) 25, and b) 1000 haploid individuals. Alleles drift to loss or fixation much more rapidly at lower population sizes.

of diversity in an attempt to reconstruct past demography. However, disentangling the complex interplay of demographic and evolutionary forces influencing genetic variation in large datasets is a challenging task. In order to achieve this, we need means of measuring and interpreting patterns of genetic diversity. A common starting point in many genetic studies is to summarize genetic diversity using tools capable of visualizing genetic structure.

2.3.1 Visualizing Population structure

A popular tool in genetics studies for over half a century, principal components analysis (PCA), is a useful means of visualizing the genetic structure in a dataset (Pearson, 1901; Menozzi, Piazza, and Cavalli-Sforza, 1978; Cavalli-Sforza et al., 1994). Each genetic marker (e.g. a SNP) in a dataset represents a dimension along which individuals in the dataset can vary. PCA is a statistical method for reducing such high-dimensional datasets to a lower number of dimensions, increasing interpretability while maintaining as much variation as possible. For instance, the genetic variation present among a dataset containing ancient and modern African populations is summarised in the PCA shown in figure 2.2. New axes have been constructed that maximize the variance in the dataset, and individuals have been projected onto those axes. PCA has been used to show major demographic patterns among large datasets, for instance that the genetic variation among European populations correlates with geography (Menozzi, Piazza, and Cavalli-Sforza, 1978; Novembre et al., 2008). Although PCA projections can be distorted by high rates of missingness (as can be the case with low-coverage ancient samples), it is possible to construct the major components of variation using a high quality subset of the data. In aDNA studies it is commonplace to use high-coverage modern samples to build the principal components and then project the ancient samples onto those axes. It ought to be remembered however that genuine variation present only in the ancient samples will not be captured in these projections, as their position in the PCA will be constrained by the variation in modern diversity.

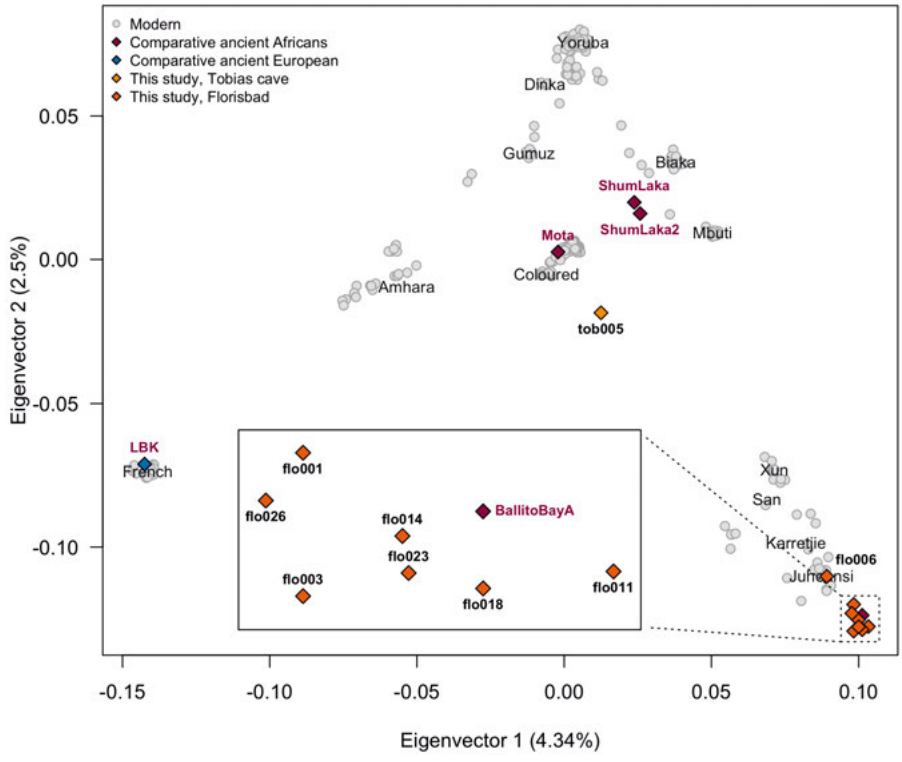


Figure 2.2. An unprojected PCA comparing ancient southern Africans against a global comparative dataset of ancient and modern populations (Paper IV).

The model-based clustering algorithms STRUCTURE (Pritchard, Stephens, and Donnelly, 2000) and ADMIXTURE (Alexander, Novembre, and Lange, 2009) are fundamentally related to PCA (McVean, 2009; Engelhardt and Stephens, 2010). These techniques fit the allele frequencies in a dataset to a model in which a fraction of each individual is assigned to one of a pre-defined number of clusters. These clusters are commonly interpreted as representing hypothesized ancestral populations. Similarities in clustering between individuals then represents shared demographic history, enabling the quick detection of patterns of genetic relatedness across a dataset. Caution needs to be exercised not to over-interpret the results of such analyses (Novembre and Peter, 2016; Lawson, Dorp, and Falush, 2018), and it should be remembered that very different demographic models can result in similar PCA or ADMIXTURE results (McVean, 2009; Lawson, Dorp, and Falush, 2018). Although PCA and ADMIXTURE are useful for exploring patterns of population structure in large datasets, their results are difficult to relate to explicit demographic models. They are best used to help formulate hypotheses, and complemented by formal tests of these hypotheses.

2.3.2 f-statistics

Reich's f-statistics represent a framework for quantifying levels of shared genetic drift and formally testing hypotheses of admixture between populations (Reich et al., 2010; Patterson et al., 2012). Closely related to Wright's F_{ST}, a measure of population differentiation due to genetic structure (Wright, 1931; Wright, 1949), f-statistics are today among the most popular methods for investigating relationships between ancient and modern populations (Peter, 2016). They are built upon the covariance in allele frequencies between pairs of populations, and the premise that this covariance implies shared evolutionary history. As such, they find particular application as formal tests of questions relating to the "tree-ness" of populations. Reich (2009) originally proposed the $f3$-population and $f4$-population tests of admixture. The $f3$-population statistic tests the null hypothesis that in an unrooted topology relating three populations (A, B and X), the allele frequency changes occurring in the branch leading to X should be uncorrelated with the allele frequency changes occurring in either of the other two population branches. In the presence of a significant skew in allele frequency changes towards one of either A or B, this null hypothesis is rejected in favour of a more complex topology involving admixture. Since this topology is unrooted, the test makes no assumptions regarding the order in which populations A, B and X diverged, though this statistic does have the limitation that strong genetic drift in the branch leading to population X can mask the signal of admixture. Significance is assessed with standard errors estimated using a block-jackknife approach (Kunsch, 1989).

Raghavan et al. (2014) proposed a variant of the $f3$ admixture test that is rooted in the interpretation of f-statistics as measures of shared drift. Using an outgroup (O), the statistic $f3(O; X, Y)$ estimates the shared drift between O and two populations X and Y, where X represents a target population, and Y represents one of a panel of populations. Repeating the test with each of the panel populations, the outgroup-$f3$ test aims to find the population most closely related to target population X. Outgroup-$f3$ tests are frequently used in aDNA studies to find the modern population most closely related to an ancient sample. They are used in Paper IV to investigate which modern Khoe-San population shares greatest genetic affinity to different ancient southern African hunter-gatherers.

More powerful than the f_3-population test of admixture, but requiring more prior assumptions regarding the relationships among studied populations, is Reich et al. (2009) f_4-population test. Closely related to the ABBA-BABA test (Green et al., 2010; Durand et al., 2011), the f_4-population statistic is used to test for deviations from a simple tree-like topology relating four populations W, X, Y and Z. In the unrooted topology $((W, X), (Y, Z))$, correlations in allele frequency differences between W and X, and Y and Z, are inconsistent with the

Figure 2.3. (A) Map showing distribution of Asia-Pacific populations with (B) levels of Denisovan ancestry estimated using f4-ratio test of form f4(Mbuti, Neanderthal; East-Asian, X) / f4(Mbuti, Neanderthal; East-Asian, Denisovan), where X=tested population. (Paper II).

proposed topology. This inconsistency is interpreted as evidence of gene flow between branches, with the sign of the statistic indicating whether gene flow has occurred between W and Y (or X and Z), or W and Z (or X and Y). A major strength of the f_4 population test is that it provides a framework with which to estimate admixture fractions. Calculating the ratio of two f_4-statistics where O represents an outgroup:

$$\alpha = \frac{f_4(O,W;A,Z)}{f_4(O,W;X,Z)}$$

provides an estimate α of the fraction of X related ancestry in target population A. f_4-ratio tests of admixture are used in Paper III to estimate fractions of archaic admixture into present-day Australasian populations (figure 2.3).

The formal admixture tests detailed above are useful for testing simple topologies relating 2, 3 or 4 populations. But increasingly, population genetic datasets are characterised by much larger numbers of populations. The goal of investigating historical relationships among large numbers of populations has led to the development of tree or graph-based approaches that aim to fit more complex models of demographic history to genetic data. Methods including qpgraph (Reich et al., 2009), TreeMix (Pickrell and Pritchard, 2012) and Momi2 (Kamm et al., 2020) fit allele frequency data to complex demographies that involve population divergences, changes in genetic drift along branches, and admixture events in the history of populations. These approaches represent powerful tools for demographic inference, but rely on allele frequencies at a particular set of loci assumed to be independent, and therefore do not take advantage of the rich source of information that exists in the correlation between

linked loci along a genome. To do so however, requires the difficult task of modelling recombination along the genome.

2.3.3 Multiple Sequentially Markovian Coalescent

Building on earlier theoretical work (Wiuf and Hein, 1999; McVean and Cardin, 2005), many modern population genetic methods aim to model recombination as a stochastic process along the genome. For instance, the "Sequentially Markovian Coalescent" model fits a Hidden Markov Model (HMM) along a DNA sequence, a technique underlying several powerful methods used today, including the PSMC (Li and Durbin, 2011), MSMC (Schiffels and Durbin, 2014), MSMC-IM (Wang et al., 2019), SMC++ (Terhorst, Kamm, and Song, 2017), and ARGWeaver (Rasmussen et al., 2014). Spurred on by the ever-increasing numbers of whole genome sequences available, some of these tools are capable of scaling to large population datasets (Albers, 2020; Kelleher et al., 2019; Speidel et al., 2019). The general principle underlying these approaches is that, rather than finding an exact solution to a simple model, they instead find an approximate solution to an arbitrarily complex (and presumably more realistic) model of demographic history. They have been used to estimate trajectories of historical population size change, track the history of gene flow between populations and date population divergence events. Given that the evolutionary history of human populations has been shown to be complex and "non tree-like" (Reich et al., 2010; Green et al., 2010; Gunz et al., 2009; Schlebusch et al., 2012; Bergström et al., 2020; Wang et al., 2020), the allowance for greater complexity may improve demographic inference.

It should be remembered however that there are good arguments for keeping models conceptually simple, beyond obvious reasons of mathematical tractability. Simple models usually come with more transparent assumptions, and evaluating the performance of these models when assumptions are not met can be much more straightforward than with complex models. As an example of the kind of bias that may occur as a result of model mis-specification with even highly sophisticated statistical methods, we can look to the case of the Pairwise Sequentially Markovian Coalescent (PSMC) (Li and Durbin, 2011). The publication of this method marked a watershed moment in the history of population genetics. Capable of estimating changes in effective population size through time using just a single diploid genome as input, this tool (and the closely related MSMC and MSMC2) has found application in many studies of demographic history. In many cases this trajectory is interpreted straightforwardly as changes in census population size. As we have seen however, there is no direct relationship between effective population size and census population size. A recent study highlighted this by reconstructing typical PSMC trajectories using classical models of population structure, with

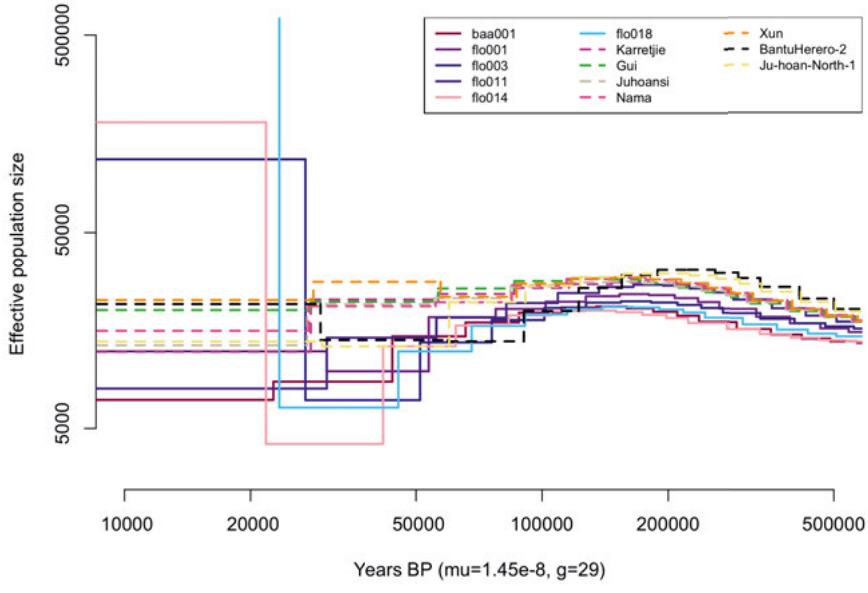


Figure 2.4. Trajectories of effective population size through time estimated for ancient and modern African populations using MSMC. Assumed mutation rate = $1.45x10^{-8}$ and generation time = 29 years (Paper IV).

varying migration rates and no changes in population size at all (Mazet et al., 2016).

2.3.4 Runs of homozygosity

Runs of homozygosity (ROH) occur when an individual inherits two copies of an ancestral haplotype. To understand how this can happen, consider the fact that without past inbreeding, an individual alive today would have a billion ancestors 30 generations ago, greatly outnumbering the estimated global population at that time of $\sim$300 million (McEvedy and Jones, 1978). This apparent genealogical paradox is resolved when it is realized that some degree of inbreeding has occurred in the history of all human populations. Because our parents are descendent from a common ancestor, it is possible to inherit two copies of an ancestral allele at a locus, leading to higher rates of homozygosity than would be expected under a panmictic population. This deviation from panmictia in the history of populations is captured by distributions of ROH (Ceballos, Hazelhurst, and Ramsay, 2019). Although genome-wide data demonstrate that ROH is common to all human populations (Ceballos, Hazelhurst, and Ramsay, 2019), differing demographic histories will have different

impacts on the distributions of ROH observed in populations. In outbred populations the number and length of ROH tend to increase with decreasing effective population size. Bottlenecks in the history of populations result in elevated ROH (Jakobsson et al., 2008), and communities that have experienced strong inbreeding depression as a result of consanguinity display the greatest burden of ROH. The genome analysis tool PLINK can perform fixed-window scans of chromosomes to find stretches of consecutive homozygous SNPs, designated as ROH when the number of these SNPs exceeds user-defined thresholds. Inbreeding coefficients can then be estimated from the proportion of the genome designated as ROH (McQuillan et al., 2012; Ceballos et al., 2018; Fortes-Lima et al., 2022). Patterns of ROH and inbreeding coefficients are studied in Paper IV to better understand the demographic histories of ancient and modern African populations (figure 2.5).

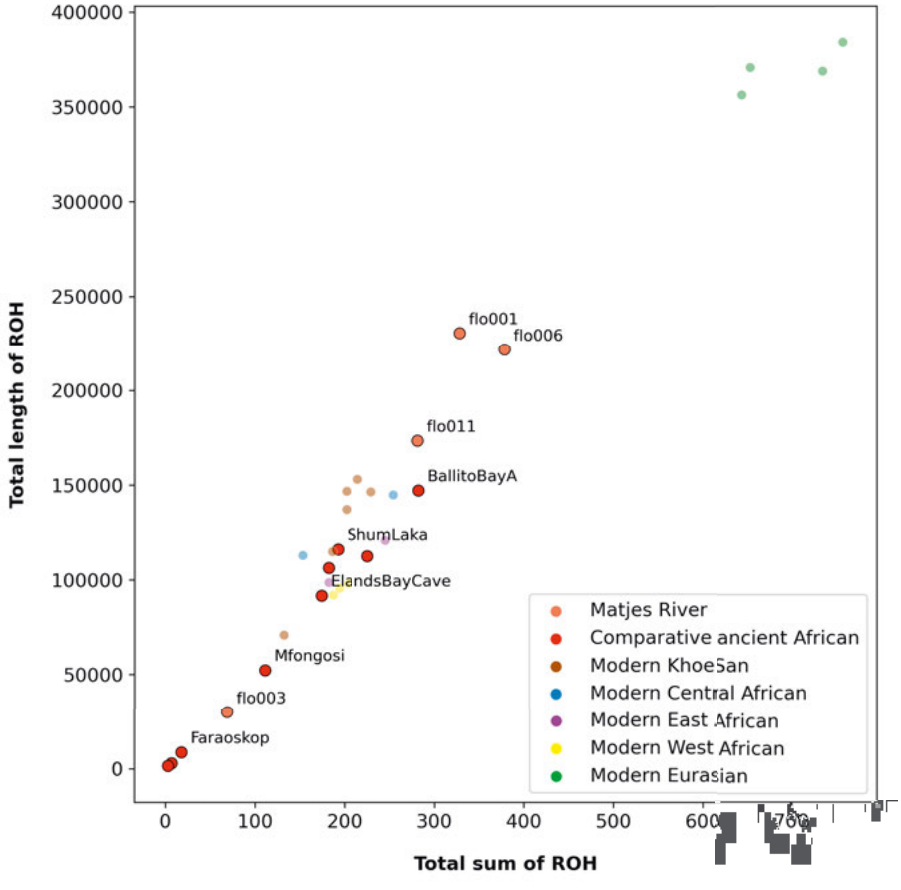


Figure 2.5. Patterns of runs of homozygosity (ROH) among studied populations of Paper IV, showing the relationship between total sum of ROH and total length of ROH for each group.

2.4 Simulating genomes

In recent decades the field of population genetics has shifted from being a discipline primarily concerned with "prospective" theoretical questions, to one concerned with the empirical study of observed patterns of genetic variation and "retrospective" demographic inference. This has largely been driven by the ever-increasing availability of genetic data (Siva, 2008; Mallick et al., 2016; Fan et al., 2019; Bergström et al., 2020; Bycroft et al., 2018; Le et al., 2022), but also results from our increasing capacity to simulate genetic data. Simulations provide us with the ability to evaluate the likelihood that observed variation results from a given demographic model. They also enable us to "ground-truth" and assess the performance of new statistical methods on data simulated under known demographic parameters.

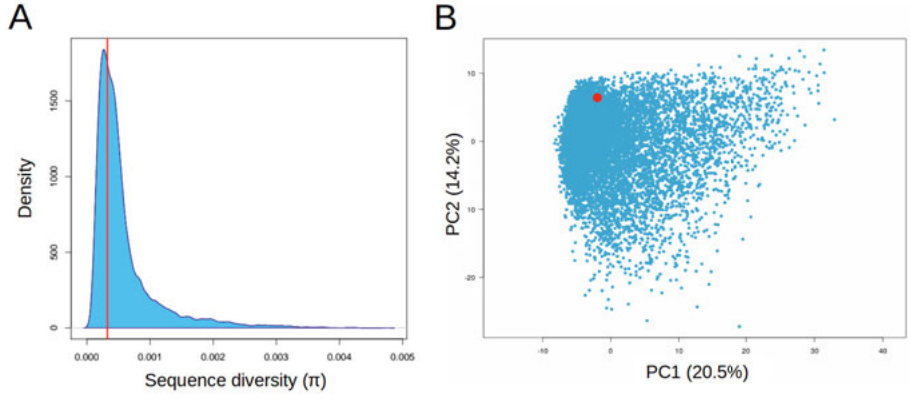


Figure 2.6. Comparing simulated and empirical data to evaluate model fit. (A) Individual summary statistics calculated for observed data (here sequence diversity (π) indicated by the red line) can be compared to the distribution of that summary statistic calculated across large numbers of datasets simulated under a particular demographic model (n=20,000). (B) PCAs can be used to summarize model fit for large numbers of summary statistics. Here principal components have been built using 500 summary statistics calculated for 20,000 simulated datasets under a particular demographic model. The summary statistics calculated for an observed dataset are then projected (red point) onto these axes to compare model fit.

There are two main classes of simulation software available; backwards-in-time simulators based on the coalescent model (Hudson, 2002; Laval and Excoffier, 2004; Hellenthal and Stephens, 2007; Kelleher, Etheridge, and McVean, 2016; Kelleher and Lohse, 2020; Baumdicker et al., 2022) and forwards-in-time simulators (Peng and Kimmel, 2005; Messer, 2013; Haller and Messer, 2017). Each have their advantages and disadvantages. The coalescent model, of central importance to much of modern population genetics (Wakeley, 2009; Nordborg, 2019), seeks to provide a mathematical description of the branching process by which the genealogies relating individuals are made (Kingman,

1982a; Kingman, 1982b; Hudson, 1983). Coalescent-based approaches such as msprime (Kelleher, Etheridge, and McVean, 2016; Kelleher et al., 2019; Baumdicker et al., 2022) have made rapid advancements in recent years as frameworks for highly efficient simulation of sequence data under models of almost arbitrary complexity. They do however fundamentally depend on an assumption of neutrality. This inability to simulate the processes of natural selection means that in some cases they may poorly reflect the evolutionary forces impacting real populations (Johri et al., 2021a; Johri et al., 2021b). Forwards-in-time simulators such as SLiM (Messer, 2013; Haller and Messer, 2017) simulate an entire population and track the ancestral process for every individual through time. Although this means they are capable of generating data under selection models, this can be computationally burdensome, limiting scalability. In recent years these two approaches have been combined such that efficient simulation under more realistic biological models combining both demographic and selection processes is possible (Haller and Messer, 2019; Kelleher et al., 2019). The emergence of tools capable of introducing errors into "perfect" simulated data, so that these datasets better reflect the damage patterns, sequencing errors and low-coverages characteristic of ancient genomes (Renaud et al., 2017) has made simulation even more applicable to ancient genomes. Finally, the growth of community maintained libraries of commonly used population genetic models (Adrion et al., 2020; Gower et al., 2022) has meant it is easier than ever to simulate data under complex demographic models. In this thesis I have used msprime to simulate data, both as means to evaluate the performance of new statistical methods (Papers I and III), and as a tool for evaluating various models of archaic admixture into the ancestors of modern human populations (Paper II).

3. Research Aims

The main objectives of this thesis were to develop and apply novel methods capable of utilizing ancient and modern genomes to gain a deeper understanding of the demographic forces shaping human populations. Specific aims were to:

I. Develop a method for estimating population divergence times using genetic data. Evaluate the robustness of this approach using simulation, and apply the method to a global panel of populations to better understand demographic history.

II. Investigate levels and patterns of admixture from archaic hominins into the ancestors of the present-day Ayta population of the Philippines.

III. Develop a tool to assess population continuity through time using a temporal sample of genomes. Evaluate the power and robustness of this approach using simulation and apply to ancient genomes of hunter-gatherers and farmers from the Scandinavian Early Neolithic.

IV. Describe the genetic diversity among ancient southern African genomes and investigate the relationships of these individuals to modern Khoe-San groups and ancient hunter-gatherers from across Africa.

4. Summary of the papers

4.1 Paper I

When studying the evolutionary and demographic histories of populations, we are often interested in estimating the times at which populations diverge. Genetic drift accumulates along the branches of a phylogeny as a function of effective population size and the number of generations. Therefore, if we can estimate the drift along two branches of a general population split model, (assuming a particular generation time), we can estimate population divergence time. One simple approach is to assume a fixed effective population size in each of the two daughter populations of the model, but demographic changes in real populations violate this assumption and can bias resulting divergence time estimates. In this paper we present a novel method for estimating population divergence times. This approach requires only a diploid genome from each of two populations and requires no assumptions regarding the effective population size of daughter populations. Assuming the ancestral and derived variants are known at bi-allelic sites, sampling two gene copies from each of two populations results in one of nine possible sample configurations. Restricting the data in this way makes it possible to derive solutions for all possible sample configurations. These equations can then be solved to obtain analytic estimates of model parameters, including population divergence times.

The advantages of this simple approach are speed and computational efficiency, together with the transparency of the model underlying the method. The model assumes no migration between daughter populations, and a constant ancestral population size. Because it is important to understand the kinds of bias we can expect when such assumptions are violated, we include a simulation study to evaluate the power and robustness of the method in cases where these core assumptions are violated. For purposes of comparison, we also evaluated the performance of an alternative method of demographic inference; the Generalized Phylogenetic Coalescent Sampler, or G-PhoCS (Gronau et al., 2011). This powerful, computationally intensive Bayesian method allows users to infer ancestral population sizes, population divergence times, and migration rates from individual genome sequences.

We found that ancestral population bottlenecks can result in a severe downward bias on divergence time estimates. This makes intuitive sense given that a sufficiently severe ancestral bottleneck results in lineages coalescing in the

ancestral population earlier than would be expected under a model of constant ancestral size. Therefore, such bottleneck events in the ancestral population are interpreted by the "Two-Two" (TT) method as a reduced population divergence time. Surprisingly, ancestral bottlenecks result in comparable levels of bias with G-PhoCS (figure 4.1 A-C), suggesting that such biases may be common in methods of divergence time estimation. Application of the TT method to empirical datasets of diverse human populations support these findings; comparisons between non-African populations can result in biologically nonsensical negative divergence times. The fact that a similar pattern is not observed for those comparisons involving African populations, suggests that a significant reduction in divergence time estimates is being driven by the Out-of-Africa bottleneck. This finding led us to propose and test an outgroup ascertainment scheme capable of relaxing the assumption of a constant ancestral population size (TTo). Both simulations and empirical findings suggest that the ascertainment scheme reduces much of the downward bias caused by ancestral bottlenecks (figure 4.1 D-F). We further demonstrate that the TT method is relatively robust to low levels of gene flow between diverged populations, but can exhibit increasing bias when levels of mean migration between daughter populations exceed 10%. The usefulness of this approach is then illustrated by estimating divergence times between a global panel of populations and archaic humans, both with and without outgroup ascertainment.

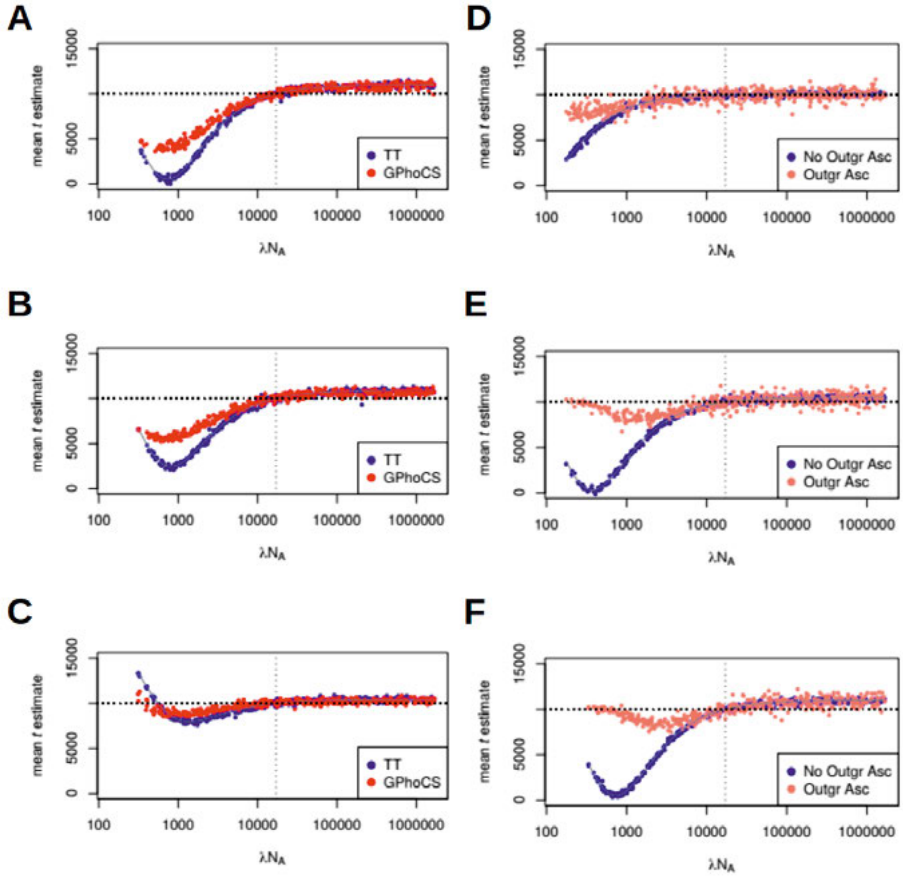


Figure 4.1. (A-C) Comparisons of the effect of ancestral population size changes (δN_A) on TT method and GPhoCS divergence time estimates when time between population divergence and δN_A is (A) 0, (B) 2000, and (C) 10,000 generations. (D-F) Comparing divergence time estimates vs δN_A with (TTo), and without (TT) outgroup ascertainment. The duration of alternative ancestral population size is (D) 100, (E) 500, and (F) 1000 generations. In all cases, true divergence time is 10,000 generations.

4.2 Paper II

All non-African human populations today show traces of archaic admixture in their genomes (Green et al., 2010; Reich et al., 2010). Whereas Eurasians possess uniform levels of Neanderthal ancestry, elevated levels of Denisovan ancestry relative to any other population is unique to Australasians (a collective term for the shared genetic ancestry between Philippine Negritos and Australopapuans). Recent findings (Sankararaman et al., 2016; Skoglund, Götherström, and Jakobsson, 2011; Prufer et al., 2014) suggest considerable structure in the Denisovan lineage and that the history of admixture between Deniso-

vans and the ancestor of Australasians was more complicated than originally thought (Meyer et al., 2012; Reich et al., 2011; Browning et al., 2018; Jacobs et al., 2019). To deepen our understanding of past Denisovan-Australasian interactions in Island Southeast Asia (ISEA) we comprehensively investigated the archaic ancestry of 1,107 individuals from 118 distinct ethnic groups of the Philippines. Following their divergence from Papuans ∼53 kya, the Philippine Negritos are regarded as the earliest modern human inhabitants of Sundaland, a contiguous biogeographical region largely exposed during the last glacial period (Larena et al., 2021; Jinam et al., 2017; Reid, 2013). Using PCA and ADMIXTURE we explored patterns of structure among ISEA populations, finding evidence that Negritos lie on a cline between two hypothesized ancestral sources: an Australasian-related and an East-Asian-related ancestry. They also show considerable genetic diversity, with at least five distinct clusters within this group. Using formal tests of admixture (*D* tests) we show that all sampled Negrito groups exhibit varying degrees of admixture with East-Asians, attributable to recent admixture with Austronesian-speaking migrants from the South China-Taiwan greater area (Larena et al., 2021; Jinam et al., 2017). Interestingly, though Neanderthal ancestry is uniform across Philippine populations, PCA suggests greater genetic affinities between Negrito groups and Denisovans than is present in other ISEA populations. We detect a clear correlation between Negrito ancestry and Denisovan ancestry, highest for Northern Negritos (particularly the Ayta Magbukon). This group possess the highest estimate of Denisovan ancestry recorded to date, despite their significant levels of recent admixture with people of East Asian ancestry. This becomes especially clear after masking regions of East-Asian-related ancestry from the genomes of Negritos. These masked Negrito genomes show significantly higher (34-40%) Denisovan ancestry than Papuans and these results are confirmed using whole genomes sequences (mean depth of $\sim 37x$). Supporting these findings, we used the S' framework to identify putative archaic regions in the genomes of Ayta Magbukon and Papuans. We find the average amount of Denisovan sequence per individual in Ayta Magbukon to be significantly higher than that of Papuans (51.94 Mb [95% CI: 44.62 - 59.25 Mb] versus 41.96 Mb [95% CI: 39.54 - 44.37 Mb]; $p = 3.7 \times 10^{-3}$). The Denisovan segments in both populations share only moderate affinity to the reference Altai Denisovan, (match rate = ∼50%), suggesting that the archaic groups introgressing into the Ayta Magbukon and Papuans are likely to be distantly related to the Altai Denisovan of Siberia.

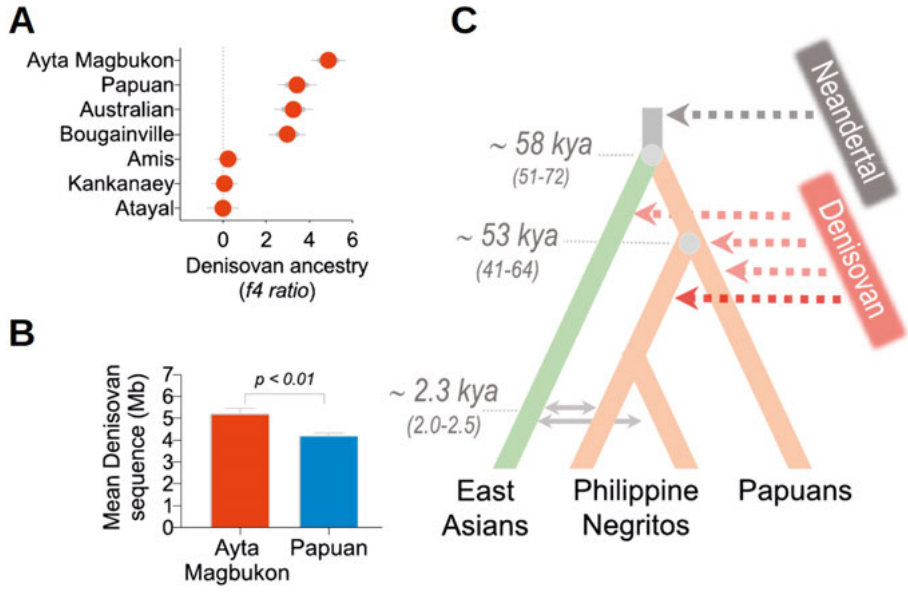


Figure 4.2. (A) Using high-coverage genomes, the percentages of Denisovan ancestry were estimated using $f4$-ratios of the form $f4$(Mbuti, Neanderthal; Han, X)/$f4$(Mbuti, Neanderthal; Han, Denisovan). (B) Mean plus SEM of Denisovan sequence detected per individual using S' approach. (C) Topology of East Asians and Australasians indicating the pulses of Denisovan introgression into Papuans and East Asians as previously reported in Browning et al., 2018 and Jacobs et al., 2019, and an independent pulse of Denisovan introgression into Philippine Negritos reported here. Time of divergence between East Asians and Australasians was previously reported in Malaspinas et al., 2016. Divergence time between Philippine Negritos and Papuans was estimated using the TT method (Sjödin, McKenna, and Jakobsson, 2021). Mean date of admixture between Philippine Negritos and East Asians was inferred using Malder. Values between brackets represent 95% confidence intervals.

The significantly higher level of Denisovan ancestry in Ayta Magbukon relative to Papuans highlights the possibility of separate introgression events between Denisovans and both Negritos and Australopapuans. We used qpGraph, (an f-statistics-based framework used to model population topologies including admixture), to demonstrate that the observed data is consistent with a model of an independent introgression event into the ancestral Negrito population. An alternative explanation could be a dilution of Denisovan ancestry in Papuans through admixture with populations containing little Denisovan ancestry. However, neither we nor previous studies (Malaspinas et al., 2016; Bergström et al., 2017) find evidence for recent gene flow from East Asians or Europeans into highland Papuans. On the contrary, Negritos do display evidence for recent admixture with East Asians (who carry very little Denisovan ancestry). A simulation study performed using the coalescent-based simulator msprime (Kelleher, Etheridge, and McVean, 2016; Kelleher et al., 2019)

showed that models including multiple Denisovan admixture events produce patterns of Denisovan ancestry more consistent with the observed data than the null hypthothesis of a single event in the shared ancestral population of Negrito and Papuans.

Taken altogether, our results demonstrate the highest levels of Denisovan ancestry yet detected in Negrito groups, and suggest two independent introgression events between Denisovans and the ancestors of Ayta Negritos and Papuans. It is interesting to note that the populations with the highest levels of Denisovan ancestry are found in the regions of ISEA and Near Oceania, yet no Denisovan fossils have yet been discovered in this region. However the majority of our knowledge on Denisovans is based on genomic data from the Altai Denisovans of Siberia (Meyer et al., 2012; Reich et al., 2010). The only other direct evidence of Denisovans outside of Siberia is from the Baishiya Karst Cave site of the Tibetan Plateau, (Chen et al., 2019; Zhang et al., 2020) where an ancient proteome and mtDNA analysis revealed the Xiahe hominin to be phylogenetically affiliated with the Altai Denisovan. This has led to suggestion that other previously discovered archaic humans in the region may in fact be Denisovans, including the Xujiayao and Penghu remains Chen et al., 2019; Dennell et al., 2020; Ao et al., 2017. Physical evidence for a previously undescribed hominin in Luzon $\sim$67 kya (Mijares et al., 2010; Détroit et al., 2019) (where present-day Negritos reside), is further suggestive of the intriguing possibility that Denisovans may have comprised deeply structured populations with considerable genetic and phenotypic diversity, enabling them to adapt to a wide variety of environments and thus inhabit a broad geographic range across the Asia-Pacific region (Teixeira et al., 2021; Choin et al., 2021; Huerta-Sánchez et al., 2014; Gittelman et al., 2016)

4.3 Paper III

A major contribution of the field of genetics to the study of the human past has been the finding that the spread of technologies, languages and culture have often been associated with large-scale movements of people, rather than cultural diffusion (Haak et al., 2015; Patin et al., 2017; Tishkoff et al., 2009; Reich et al., 2009). Underpinning this is the fact that genetic evidence of historical admixture, population displacements, and population replacements frequently accompanies archaeological evidence of the regional spread of material culture. Ancient genomes have proven to be a particularly powerful resource in this regard as they offer us direct windows onto patterns of genetic diversity through time. With them we can answer questions of demographic change that would be difficult or impossible to resolve using modern genomes alone (Llamas, Willerslev, and Orlando, 2017). A challenge associated with the effective use of ancient genomes however, is that many popular population genetic tools do not account for temporal sampling of DNA sequences

(Skoglund et al., 2014; Sjödin, Skoglund, and Jakobsson, 2014; Schraiber, 2017). If neglected, the genetic drift expected to separate ancient from modern samples has the potential to cause misinterpretation. For instance it can cause ancient samples to cluster together on PCA projections, looking more genetically related than they really are (François and Jay, 2020). Therefore it is important that statistical tools are developed that account for this additional temporal component of genetic drift, and that are capable of taking full advantage of the ever-increasing numbers of ancient genomes available for demographic inference.

In this paper, we introduce a new method to detect population continuity. The statistic it is based on is a simple one. If we take the oldest individual in a sample of ancient genomes, and identify heterozygote ancestral/derived sites in that individual (whom we identify as the "anchor"), then we condition on those sites and count the proportion derived occurring in more recent individuals. We show that genetic drift affects this statistic differently forwards and backwards in time from the population the anchor is sampled from. Specifically, the mean proportion derived does not change forwards-in-time, but is reduced backwards-in-time at a rate depending on the effective size of the population in the branch leading to the anchor. This means that a reduction in proportion derived in anchor-conditioned sites, in individuals more recent than the anchor, can signal population discontinuity. In particular, it can signal population admixture events and population replacements among a temporal DNA sample. Importantly, this is true even for admixture from an unsampled (also known as "ghost") population. Identifying such "ghost admixture" events in the absence of suitable reference genomes represents a long-standing challenge in population genetics. For instance, definitive genetic evidence of archaic admixture from Neanderthal and Denisovans into non-African human populations was not detectable until genomes from those species were sequenced (Reich et al., 2010; Meyer et al., 2012). Using msprime simulations, we evaluate the performance of this statistic, showing that it is capable of rejecting population continuity in the presence of ghost admixture. We also show that in a case with >1 anchor, it can be possible to estimate these admixture fractions, with estimation accuracy primarily depending on the time separating the two individuals used as anchors.

As a use case for this test of population continuity, we study Early Neolithic Southern Scandinavia. The retreat of the Fenno-Scandian ice sheet $\sim$11.7 kya allowed Eastern and Western hunter-gatherers to quickly colonize southern Scandinavia, forming the Scandinavian hunter-gatherer population (Günther et al., 2018). Neolithic material culture spread to Scandinavia relatively late, only arriving with arrival in southern Scandinavia of the farming Funnel Beaker culture (TRB) $\sim$5.5 kya. Archaeological evidence shows the emergence several hundred years later of another group of people along the coasts and islands of the region, the hunter-gathering Pitted Ware culture (PWC). In-

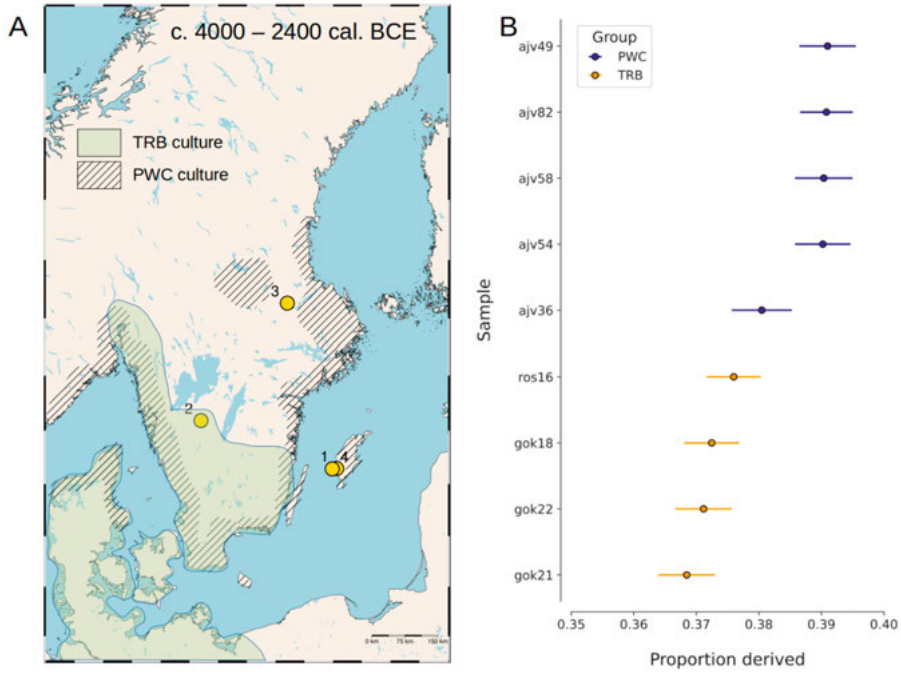


Figure 4.3. (A) Map showing distributions of the Funnel Beaker culture (TRB), and the sub-Neolithic Pitted Ware culture (PWC) during the Nordic Middle Neolithic period. All PWC individuals from 1. Ajvide burial, Gotland. TRB burial sites sampled include 2. Gökhem passage grave, and 3. Rössberga passage grave. Mesolithic individual sf12 from 4) Stora Förvar on Stora Karlsö. (B) Results of anchor test of population continuity, with Scandinavian Mesolithic hunter-gatherer sf12 used as anchor. Although ajv36 deviates from the general pattern, the results are consistent with the PWC hunter-gatherers being continuous with sf12 while TRB farmers show clear discontinuity with both sf12 and the PWC individuals. Error bars show two weighted-block jackknife standard errors.

triguingly, while the TRB show genetic affiliation to other farming cultures of Central and Eastern Europe, the PWC have been shown to exhibit genetic relatedness to earlier Mesolithic hunter-gatherers. Both archaeological and previous genetic studies suggest limited admixture between these two groups for several hundred years. These findings are recapitulated here, with the anchor statistic showing strong population continuity among the PWC and the contemporaneous TRB farmers showing significant discontinuity. Hence we provide further evidence to support the surprising picture of two cultures, overlapping both geographically and temporally, yet with limited gene-flow between them for several hundred years.

4.4 Paper IV

The Khoe-San have been shown to possess among the highest genetic diversities of all present-day human groups. Despite this fact, and the fact that they represent one branch of the deepest population divergence among modern humans (all other populations representing the other branch), the ancestry of Khoe-San people is still relatively unknown. Recent genetic studies (Schlebusch et al., 2012; Coutinho et al., 2021; Vicente et al., 2019; Schlebusch et al., 2017; Schlebusch et al., 2020) have shown considerable genetic diversity among this group, with deep divergences between the traditionally hunter gathering Northern San and Southern San, and the traditionally pastoralist Khoe people. The Khoe have been shown to have been formed from admixture between local San groups and people of East African ancestry $\sim$1.3 kya, when pastoralist herders arrived in southern Africa. A previous study (Schlebusch et al., 2017) sequenced the genome of an ancient ($\sim$2 kya) hunter-gatherer boy from Ballito Bay and found the surprising result that all modern Khoe-San populations possess some significant degree of non-Khoe-San ancestry. In this paper we sequenced 28 ancient Khoe-San genomes from South Africa with a broad geographic and temporal distribution, stretching from recent times to the Later Stone Age ($\sim$10.2 kya). A subsample of individuals at one site alone, Matjes River Rock Shelter, span a stratigraphy of $\sim$8,000 years. This enabled us to study Khoe-San ancestry in new depth; investigating their relationships to each other, to other ancient hunter-gatherers in East and Central Africa, and to modern Khoe-San groups.

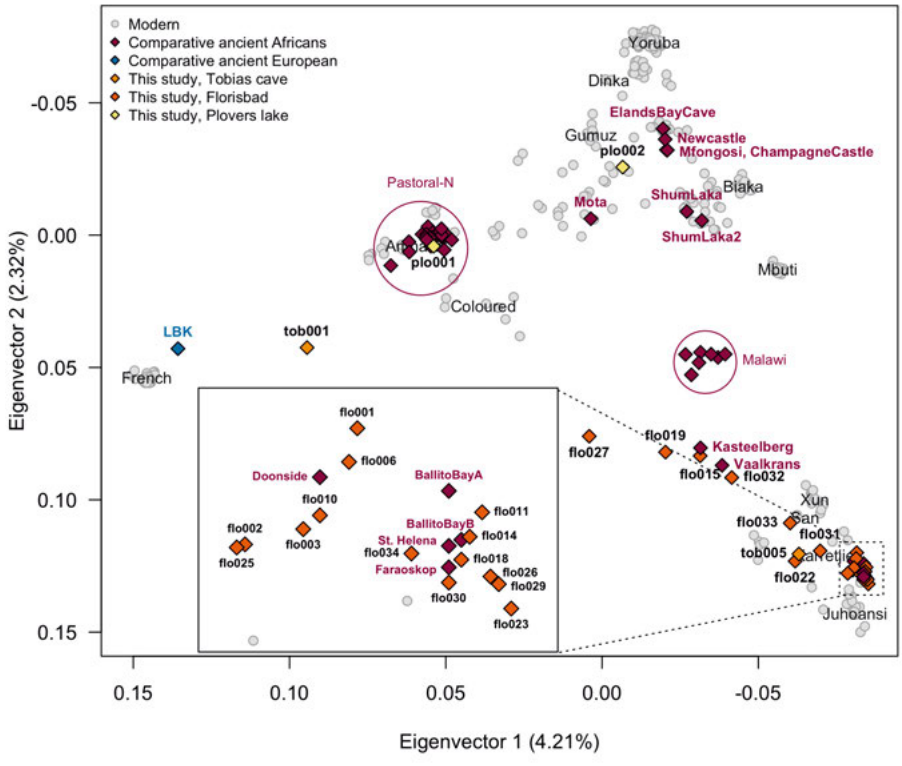


Figure 4.4. Projected PCA of ancient southern Africans (bold black font) and a comparative panel of ancient (purple font) and modern African populations (grey points). Components of variation have been built using modern high coverage comparative samples.

Individuals flo022, flo031, flo032, flo015, flo019 and flo027 fall outside the core ancient south African cluster, between modern Khoe-San and modern East African populations in figure 4.4. All these individuals, with the exception of flo019, possess the Khoe-San associated mtDNA haplogroup L0d, and it is interesting to note that all were dated to 0-720 years ago. All samples older than $\sim$720 years ago demonstrate a remarkable level of genetic homogeneity lasting from at least the Late Stone Age to relatively recent times. Strong clustering in PCA, ADMIXTURE and outgroup-$f3$ statistics evidence a long isolated and genetically continuous population with no evidence of admixture from hunter-gatherers in other parts of Africa. A previous study (Skoglund et al., 2017) used ancient genomes from Malawi to evidence a clinal IBD-relationship between ancient populations in southern and eastern Africa, however we can detect no evidence of East African hunter-gatherer ancestry among our ancient samples $>$2 kya. This indicates that gene-flow from hunter-gatherers of eastern African ancestry did not extend as far as the region studied here between 2 kya and at least $\sim$10 kya.

MSMC and ROH results suggest smaller effective population sizes in ancient southern African hunter-gatherers than are inferred using present-day Khoe-San genomes. This corroborates the findings of Schlebusch et al. (2020), who propose a common bottleneck in the history of all human populations, with the greater past effective population sizes of modern Khoe-San groups resulting from recent admixture with non-Khoe-San groups (also demonstrated by our results). Taken together, our results suggest that for a long time the southern African foragers appear to have been isolated from other populations, and it is only in the last millennia that migrants have reached southern Africa and mixed with the indigenous groups. These ancient hunter-gatherers capture the deepest splits in the tree of modern humans, and provide evidence of a shared bottleneck $>$300 kya in the ancestral population of all modern humans.

5. Conclusions and future prospects

In this thesis I have used novel and established population genetic methods to investigate the demographic past of human populations. The major demographic forces impacting the patterns we observe in the genetic variation of genomes can include population divergences, admixture events and long-term population continuity. Here I have developed and evaluated new methods for estimating population divergence times under a simple split model, applicable to ancient and modern DNA, and requiring only a single diploid genome per population. I investigate the sensitivities of such approaches to violations of model assumptions and non-tree-like demography, and apply them to estimate divergence times among pairs of modern and archaic human populations. Non-tree-like demography is common in the evolutionary history of human populations. In particular, the existence of admixture between archaic hominins and the ancestors of modern humans is well known. Here I investigate evidence of admixture from Denisovans into the ancestors of present-day human populations in Island South-east Asia. Using SNP-array data collected from 118 ethnic groups in the Philippines, I show that, globally, the highest levels of Denisovan ancestry are found among the Ayta Magbukon. I show evidence of possible multiple admixture events, further highlighting the complex history between modern human groups and the archaic hominins that occupied Eurasia before our arrival. Population genetics has made major contributions to the study of the human past. Among the most important findings is the demonstration that cultural transitions and the spread of technologies have often been associated with migrations of people, rather than cultural diffusion. These movements can result in the admixture, displacement or replacement of the previous inhabitants. Ancient DNA provides us with a powerful tool for assessing these trends through time. I developed a statistical tool to detect levels of population continuity through time, and evaluated its performance using simulation. Applied to a dataset of ancient hunter-gatherers and farmers from Early Neolithic Scandinavia, I demonstrate the presence of population continuity among Pitted Ware culture hunter-gatherers on the Baltic Island of Gotland, despite overlapping both geographically and temporally with Funnel Beaker farmers. Long-term population continuity and isolation was also detected among another population of hunter-gatherers, this time in a study of the ancestors of the San people of southern Africa. Until surprisingly recent times, the ancestors of the San people were unaffected by admixture from outside southern Africa. However I also provide evidence of the strong impact

that migrations into South Africa in the past $\sim$2,000 years have had, by showing that all modern Khoe-San groups exhibit significant levels of admixture with people of non-Khoe-San ancestry.

The future of population genetics will be characterized by genomic datasets of increasing size, more efficient and flexible simulation software, and powerful methods of demographic inference capable of taking advantage of both. Ancient genomes are often said to provide windows into our past, but in order to make effective use of the information contained in ancient genomes, it is vital that population genetic methods are developed that account for temporal, as well as geographic sampling of DNA sequences. The results of this thesis are a step in this direction, combining ancient and modern genomes to gain a deeper understanding of the human past.

6. Svensk Sammanfattning

Den genetiska variation vi observerar idag är ett resultat av demografiska krafter såsom populationsdivergens, migration och populationskontinuitet. Nya sekvenseringsteknologier tillsammans med utvecklingen av molekylära och bioinformatiska metoder, har resulterat i en kraftig ökning av sekvenserat DNA från forntida humana stickprov som kan användas för demografisk härledning. För att kunna utnyttja sådan antika stickprov av genom på ett så effektivt sätt som möjligt krävs dock att nya härledningsverktyg utvecklas. Specifikt behöver sådana verktyg inte bara tar hänsyn till den geografiska aspekten av var genomen kommer ifrån utan även tidsaspekten av hur gammla de är.

I den här avhandlingen har jag utvecklat, utvärderat och tillämpat nya metoder för att uppskatta divergenstiden mellan forntida och moderna populationer. Jag använde mig av datorsimuleringar för att studera hur känsliga dessa metoder är för olika typer av antaganden som gjorts för att härleda dessa metoder. När detta var gjort tillämpade jag metoderna för att studera populationsdivergens mellan par av populationer som var samplade över hela världen. Ett antagande som dessa metoder bygger på är att populationer som delas upp i två populationer inte delar genetiskt material vid senare tillfällen. Detta antagande om “trädliknande" demografi stämmer inte alltid och icke-trädliknande demografi är vanligt förekommande i det mänskliga förflutna. Till exempel kan inblandning från arkaiska mäniskor i genomen hos alla icke-afrikanska moderna människor påvisas. Med hjälp av SNP-arraydata som samlats in från 118 etniska grupper i Filippinerna visar jag att de högsta nivåerna av inblandning från de arkaiska Denisovamänniskorna återfinns bland Ayta Magbukon. Detta belyser den komplexa historien mellan moderna mänskliga grupper och de arkaiska homininerna som ockuperade Eurasien innan vår ankomst. Bland de viktigaste insikterna som populationsgenetik bidragit med till studiet av den moderna människans förflutna är att nya kulturer och spridning av teknologier ofta associerades med migrerande grupper av människor. Detta kan resultera i uppblandning eller till och med att en population ersätts av en annan. Fossilt DNA ger oss möjlighet att studera sådana förlopp direkt över tid. Ett annat statistiskt verktyg ämnat för att studera fossilt DNA som jag utvecklade handlade om populationskontinuitet. Detta verktyg kan, tillsammans med fossilt DNA, användas för att studera populationskontinuitet över tiden Jag utvärderdera dess prestanda med hjälp av datorsimuleringar och tillämpade det på en uppsättning av antika genom från det tidiga neolitiska Skandinavien. Jag kunde påvisa populationskontinuitet i den jägar-samlande Pitted Ware-kulturen, trots

att dessa människor överlappar både geografiskt och temporärt med den jordburkande Trattbägarkulturen. I en annan studie av förfäderna till det jägarsamlande San-folket i södra Afrika hittade jag bevis på att denna grupp varit isolerad under lång tid. De verkar ha varit opåverkad av inblandning från populationer utanför södra Afrika tills förvånansvärt nyligen. Dessa forntida genom belyser att alla moderna Khoe-San-populationer uppvisar betydande nivåer av blandning med människor av icke-Khoe-San härkomst, vilket ytterligare påvisar den starka inverkan befolkningsmigrationer i denna region har haft under de senaste $\sim$2,000 åren.

7. Acknowledgements

First and foremost I would like to thank my supervisors Per Sjödin and Mattias Jakobsson. Per it has been my pleasure to work alongside you and learn from you. Throughout this PhD I have found your constant curiosity an inspiration, your perspective always enlightening, and your stout never weak! Thanks for your patience, putting up with all the questions and safely guiding me through the choppy waters of population genetics. Mattias thanks for your unending support and remarkable enthusiasm throughout this PhD, and for creating such a stimulating and supportive environment at the Human Evolution program. Your passion and curiosity for these various projects were a constant source of motivation.

I would like to express my thanks to all coauthors and collaborators for their contributions to the work underlying this thesis.

A special thanks to Carolina Bernhardsson, for her hard work and assistance particularly in the final stretch, and whose plotting skills know no bounds (even if it is in R). I'd also like to thank Cesar Fortes Lima for his thoughtful suggestions and assistance in improving Paper IV. Thanks to Magdalena Fraser, Helena Malmström and Olaf Thalmann for their generous assistance and support in the work contributing to Paper III.

To everyone in the Human Evolution program and in particular all PhDs past and present - you have made this an enjoyable and memorable experience. Thanks for all the great conversations, presentations, the endless knitting and the countless fikas!

I would additionally like to thank everyone at IEG and IOB for creating such a fantastic scientific and social environment.

Last, and certainly not least, thanks to my family, friends and those adventurous souls who have shared this journey with me and supported me every step of the way.

Tusen tack!

Bibliography

Adrion, Jeffrey R et al. (2020). "A community-maintained standard library of population genetic models". In: *Elife* 9.

Albers PK. McVean, G. (2020). "Dating genomic variants and shared ancestry in population-scale sequencing data". In: *PLoS Biol* 18.1. DOI: 10.1371/journal.pbio.3000586.

Albrechtsen, Anders, Finn Cilius Nielsen, and Rasmus Nielsen (2010). "Ascertainment biases in SNP chips affect measures of population divergence". In: *Molecular biology and evolution* 27.11, pp. 2534–2547.

Alexander, DH., J. Novembre, and K. Lange (2009). "Fast model-based estimation of ancestry in unrelated individuals". In: *Genome Res* 19.9, pp. 1655–1664. DOI: 10.1101/gr.094052.109.

Allentoft, ME. et al. (June 2015). "Population genomics of Bronze Age Eurasia". In: *Nature* 522.7555, p. 167. DOI: 10.1038/nature14507.

Ao, Hong et al. (2017). "An updated age for the Xujiayao hominin from the Nihewan Basin, North China: implications for Middle Pleistocene human evolution in East Asia". In: *Journal of human evolution* 106, pp. 54–65.

Atkinson, Quentin D (2011). "Phonemic diversity supports a serial founder effect model of language expansion from Africa". In: *Science* 332.6027, pp. 346–349.

Avery, OT, CM MacLeod, and M McCarty (1944). "Studies on the chemical nature of the substance inducing transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III". In: *J. Exp. Med* 79, pp. 137–158.

Baumdicker, Franz et al. (2022). "Efficient ancestry and mutation simulation with msprime 1.0". In: *Genetics* 220.3, iyab229.

Bergström, Anders et al. (2017). "A Neolithic expansion, but strong genetic structure, in the independent history of New Guinea". In: *Science* 357.6356, pp. 1160–1163.

Bergström, Anders et al. (2020). "Insights into human genetic variation and population history from 929 diverse genomes". In: *Science* 367.6484. DOI: 10.1126/science.aay5012.

Briggs, Adrian W and Patricia Heyn (2012). "Preparation of next-generation sequencing libraries from damaged DNA". In: *Ancient DNA*. Springer, pp. 143–154.

Briggs, Adrian W et al. (2010). "Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA". In: *Nucleic acids research* 38.6, e87–e87.

Brown, Peter et al. (2004). "A new small-bodied hominin from the Late Pleistocene of Flores, Indonesia". In: *Nature* 431.7012, pp. 1055–1061.
Browning, Sharon R et al. (2018). "Analysis of human sequence data reveals two pulses of archaic Denisovan admixture". In: *Cell* 173.1, pp. 53–61.
Bycroft, Clare et al. (2018). "The UK Biobank resource with deep phenotyping and genomic data". In: *Nature* 562.7726, pp. 203–209.
Cavalli-Sforza, Luigi Luca et al. (1994). *The history and geography of human genes*. Princeton university press.
Ceballos, Francisco C, Scott Hazelhurst, and Michele Ramsay (2019). "Runs of homozygosity in sub-Saharan African populations provide insights into complex demographic histories". In: *Human Genetics* 138.10, pp. 1123–1142.
Ceballos, Francisco C et al. (2018). "Runs of homozygosity: windows into population history and trait architecture". In: *Nature Reviews Genetics* 19.4, pp. 220–234.
Chen, Fahu et al. (2019). "A late middle Pleistocene Denisovan mandible from the Tibetan Plateau". In: *nature* 569.7756, pp. 409–412.
Chen, Lu et al. (2020). "Identifying and interpreting apparent Neanderthal ancestry in African individuals". In: *Cell* 180.4, pp. 677–687.
Choin, Jeremy et al. (2021). "Genomic insights into population history and biological adaptation in Oceania". In: *Nature* 592.7855, pp. 583–589.
Clark, Andrew G et al. (2005). "Ascertainment bias in studies of human genome-wide polymorphism". In: *Genome research* 15.11, pp. 1496–1502.
Cole, Christopher B et al. (2020). "Ancient Admixture into Africa from the ancestors of non-Africans". In: *bioRxiv*.
Conroy, Glenn C and Herman Pontzer (2012). *Reconstructing human origins: a modern synthesis*. WW Norton & Company.
Coutinho, Alexandra et al. (2020). "The Neolithic Pitted Ware culture foragers were culturally but not genetically influenced by the Battle Axe culture herders". In: *American Journal of Physical Anthropology* 172.4, pp. 638–649. DOI: https://doi.org/10.1002/ajpa.24079. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/ajpa.24079. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/ajpa.24079.
Coutinho, Alexandra et al. (2021). "Later Stone Age human hair from Vaalkrans Shelter, Cape Floristic Region of South Africa, reveals genetic affinity to Khoe groups". In: *American Journal of Physical Anthropology* 174.4, pp. 701–713.
Dabney, Jesse et al. (2013). "Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments". In: *Proceedings of the National Academy of Sciences* 110.39, pp. 15758–15763.
Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection*. Murray, London.

Dennell, Robin et al. (2020). “A demographic history of Late Pleistocene China”. In: *Quaternary International* 559, pp. 4–13.
Détroit, Florent et al. (2019). “A new species of Homo from the Late Pleistocene of the Philippines”. In: *Nature* 568.7751, pp. 181–186.
Douka, Katerina et al. (2019). “Age estimates for hominin fossils and the onset of the Upper Palaeolithic at Denisova Cave”. In: *Nature* 565.7741, pp. 640–644.
Duarte, Cidália et al. (1999). “The early Upper Paleolithic human skeleton from the Abrigo do Lagar Velho (Portugal) and modern human emergence in Iberia”. In: *Proceedings of the National Academy of Sciences* 96.13, pp. 7604–7609.
Durand, Eric Y et al. (2011). “Testing for ancient admixture between closely related populations”. In: *Molecular biology and evolution* 28.8, pp. 2239–2252.
Durvasula, A. and S. Sankararaman (2020). “Recovering signals of ghost archaic introgression in African populations”. In: *Science Advances* 6.7. DOI: 10.1126/sciadv.aax5097.
Engelhardt, Barbara E and Matthew Stephens (2010). “Analysis of population structure: a unifying framework and novel methods based on sparse factor analysis”. In: *PLoS genetics* 6.9, e1001117.
Fan, Shaohua et al. (2019). “African evolutionary history inferred from whole genome sequence data of 44 indigenous African populations”. In: *Genome Biology* 20.1, pp. 1–14.
Fisher, RA. (1918). “XV.âThe Correlation between Relatives on the Supposition of Mendelian Inheritance”. In: *Transactions of the Royal Society of Edinburgh* 52.2, pp. 399–433.
– (1922). “On the dominance ratio”. In: *Proc. R. Soc. Edinb* 42, pp. 321–341.
Fortes-Lima, Cesar et al. (2022). “Demographic and selection histories of populations across the Sahel/Savannah belt”. In: *Molecular biology and evolution* 39.10, msac209.
François, O. et al. (2019). “Inference of population genetic structure from temporal samples of DNA”. In: *bioRxiv*. DOI: 10.1101/801324.
François, Olivier and Flora Jay (2020). “Factor analysis of ancient population genomic samples”. In: *Nature communications* 11.1, pp. 1–11.
Fraser, Magdalena et al. (2018). “New insights on cultural dualism and population structure in the Middle Neolithic Funnel Beaker culture on the island of Gotland”. In: *Journal of Archaeological Science: Reports* 17, pp. 325–334.
Fu, Q. et al. (2013). “A revised timescale for human evolution based on ancient mitochondrial genomes”. In: *Nature Rev. Genet.* 23, pp. 553–559.
Gibbs, Richard A et al. (2003). “The international HapMap project”. In.
Gittelman, Rachel M et al. (2016). “Archaic hominin admixture facilitated adaptation to out-of-Africa environments”. In: *Current Biology* 26.24, pp. 3375–3382.

Gower, Graham et al. (2022). “Demes: a standard format for demographic models”. In: *Genetics* 222.3, iyac131.

Gravel, Simon et al. (2011). “Demographic history and rare allele sharing among human populations”. In: *Proceedings of the National Academy of Sciences* 108.29, pp. 11983–11988.

Green, RE. et al. (2010). “A Draft Sequence of the Neandertal Genome”. In: *Science* 328.5979, pp. 710–722. DOI: 10.1126/science.1188021.

Griffith, Fred (1928). “The significance of pneumococcal types”. In: *Epidemiology & Infection* 27.2, pp. 113–159.

Gronau, I. et al. (2011). “Bayesian inference of ancient human demography from individual genome sequences.” In: *Nature Genetics* 43, pp. 1031–1035. DOI: 10.1038/ng.937.

Guldemann, Tom (2008). “A linguist’s view: Khoe-Kwadi speakers as the earliest food-producers of southern Africa”. In: *Southern African Humanities* 20.1, pp. 93–132.

Günther, T. et al. (Jan. 2018). “Population genomics of Mesolithic Scandinavia: Investigating early postglacial migration routes and high-latitude adaptation”. In: *PLOS Biology* 16.1, pp. 1–22. DOI: 10.1371/journal.pbio.2003703. URL: https://doi.org/10.1371/journal.pbio.2003703.

Gunz, P. et al. (2009). “Early modern human diversity suggests subdivided population structure and a complex out-of-Africa scenario”. In: *Proc. Natl. Acad. Sci. USA* 106, pp. 6094–6098.

Haak, W. et al. (Mar. 2015). “Massive migration from the steppe was a source for Indo-European languages in Europe”. In: *Nature* 522.7555, p. 207. DOI: 10.1038/nature14317.

Haller, Benjamin C and Philipp W Messer (2017). “SLiM 2: flexible, interactive forward genetic simulations”. In: *Molecular biology and evolution* 34.1, pp. 230–240.

– (2019). “SLiM 3: forward genetic simulations beyond the Wright–Fisher model”. In: *Molecular biology and evolution* 36.3, pp. 632–637.

Harpending, Henry and Alan Rogers (2000). “Genetic perspectives on human origins and differentiation”. In: *Annual review of genomics and human genetics* 1.1, pp. 361–385.

Hellenthal, Garrett and Matthew Stephens (2007). “msHOT: modifying Hudson’s ms simulator to incorporate crossover and gene conversion hotspots.” In: *Bioinformatics* 23.4.

Henn, BM., LL. Cavalli-Sforza, and MW. Feldman (2012). “The great human expansion”. In: *Proceedings of the National Academy of Sciences* 109, pp. 17758–17764.

Hublin, Jean-Jacques et al. (2017). “New fossils from Jebel Irhoud, Morocco and the pan-African origin of Homo sapiens”. In: *Nature* 546.7657, pp. 289–292.

Hudson, R. (1983). “Properties of a neutral allele model with intragenic recombination”. In: *Theor Popul Biol* 23, pp. 183–201. DOI: 10.1016/0040-5809(83)90013-8.
– (2002). “Generating samples under a WrightâFisher neutral model of genetic variation.” In: *Bioinformatics* 18, pp. 337–338. DOI: 10.1093/bioinformatics/18.2.337.
Huerta-Sánchez, Emilia et al. (2014). “Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA”. In: *Nature* 512.7513, pp. 194–197.
Jacobs, Guy S et al. (2019). “Multiple deeply divergent Denisovan ancestries in Papuans”. In: *Cell* 177.4, pp. 1010–1021.
Jakobsson, M. et al. (2008). “Genotype, haplotype and copy-number variation in worldwide human populations”. In: *Nature* 451, pp. 998–1003.
Jinam, Timothy A et al. (2017). “Discerning the origins of the Negritos, first Sundaland people: deep divergence and archaic admixture”. In: *Genome Biology and Evolution* 9.8, pp. 2013–2022.
Jobling, Mark A, Matthew Hurles, and Chris Tyler-Smith (2019). *Human evolutionary genetics: origins, peoples and disease*. Garland Science.
Johri, Parul et al. (2021a). “Statistical inference in population genomics”. In: *demography* 27.28, p. 29.
Johri, Parul et al. (2021b). “The impact of purifying and background selection on the inference of population history: problems and prospects”. In: *Molecular biology and evolution* 38.7, pp. 2986–3003.
Kamm, Jack et al. (2020). “Efficiently inferring the demographic history of many populations with allele count data”. In: *Journal of the American Statistical Association* 115.531, pp. 1472–1487.
Kelleher, J., AM. Etheridge, and G. McVean (2016). “Efficient Coalescent Simulation and Genealogical Analysis for Large Sample Sizes”. In: *PLoS Comput Biol* 12, e1004842. DOI: 10.1371/journal.pcbi.1004842.
Kelleher, J. et al. (2019). “Inferring whole-genome histories in large population datasets”. In: *Nature Genetics* 51, pp. 1330–1338. DOI: 10.1038/s41588-019-0483-y.
Kelleher, Jerome and Konrad Lohse (2020). “Coalescent simulation with msprime”. In: *Statistical Population Genomics*. Humana, New York, NY, pp. 191–230.
Kimura, M. and T. Ohta (1969). “The average number of generations until fixation of a mutant gene in a population”. In: *Genetics* 61, pp. 763–771.
Kingman, J. (1982a). “On the Genealogy of Large Populations”. In: *journal of Applied Probability* 19, pp. 27–43.
– (1982b). “The coalescent”. In: *Stochastic Processes and their Applications* 13, pp. 235–248.
Krause, J. et al. (2010). “The complete mitochondrial DNA genome of an unknown hominin from southern Siberia”. In: *Nature* 464, pp. 894–897.
Kunsch, Hans R (1989). “The jackknife and the bootstrap for general stationary observations”. In: *The annals of Statistics*, pp. 1217–1241.

Lachance, Joseph and Sarah A Tishkoff (2013). "SNP ascertainment bias in population genetic analyses: why it is important, and how to correct it". In: *Bioessays* 35.9, pp. 780–786.

Lam, Yin M, Osbjorn M Pearson, and Cameron M Smith (1996). "Chin morphology and sexual dimorphism in the fossil hominid mandible sample from Klasies River Mouth". In: *American Journal of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists* 100.4, pp. 545–557.

Lander, Eric S et al. (2001). "Initial sequencing and analysis of the human genome". In: *Nature* 409.6822, pp. 860–921.

Larena, Maximilian et al. (2021). "Multiple migrations to the Philippines during the last 50,000 years". In: *Proceedings of the National Academy of Sciences* 118.13, e2026132118.

Laval, Guillaume and Laurent Excoffier (2004). "SIMCOAL 2.0: a program to simulate genomic diversity over large recombining regions in a subdivided population with a complex history". In: *Bioinformatics* 20.15, pp. 2485–2487.

Lawson, DJ., L. van Dorp, and D. Falush (2018). "A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots". In: *Nat Commun* 9.3258. DOI: 10.1038/s41467-018-05257-7.

Lazaridis, I. et al. (2016). "Genomic insights into the origin of farming in the ancient Near East". In: *Nature*, pp. 1476–4687. DOI: 10.1038/nature19310.

Le, Megan K et al. (2022). "1,000 ancient genomes uncover 10,000 years of natural selection in Europe". In: *bioRxiv*.

Li, H. and R. Durbin (2011). "Inference of human population history from individual whole-genome sequences". In: *Nature* 475, pp. 493–496. DOI: 10.1038/nature10231.

Li, JZ. et al. (2008). "Worldwide human relationships inferred from genome-wide patterns of variation". In: *Science* 319, pp. 1100–1104.

Lillie, Malcolm et al. (2000). *Europe's first farmers*. Cambridge University Press.

Lipson, Mark et al. (2020). "Ancient West African foragers in the context of African population history". In: *Nature* 577.7792, pp. 665–670.

Lipson, Mark et al. (2022). "Ancient DNA and deep population structure in sub-Saharan African foragers". In: *Nature* 603.7900, pp. 290–296.

Llamas, Bastien, Eske Willerslev, and Ludovic Orlando (2017). "Human evolution: a tale from ancient genomes". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 372.1713, p. 20150484.

Llorente, M. et al. (2015). "Ancient Ethiopian genome reveals extensive Eurasian admixture in Eastern Africa". In: *Science* 350.6262, pp. 820–822. DOI: 10.1126/science.aad2879.

Lordkipanidze, David et al. (2007). "Postcranial evidence from early Homo from Dmanisi, Georgia". In: *Nature* 449.7160, pp. 305–310.

Malaspinas, Anna-Sapfo et al. (2016). “A genomic history of Aboriginal Australia”. In: *Nature* 538.7624, pp. 207–214.

Mallick, Swapan et al. (2016). “The Simons genome diversity project: 300 genomes from 142 diverse populations”. In: *Nature* 538.7624, pp. 201–206.

Malmström, H. et al. (2009). “Ancient DNA Reveals Lack of Continuity between Neolithic Hunter-Gatherers and Contemporary Scandinavians”. In: *Curr. Biol.* 19.20, pp. 1758–1762.

Malmström, H. et al. (2019). “The genomic ancestry of the Scandinavian Battle Axe Culture people and their relation to the broader Corded Ware horizon”. In: *Proceedings of the Royal Society B* 286. DOI: 10.1098/rspb.2019.1528.

Malmström, Helena et al. (2015). “Ancient mitochondrial DNA from the northern fringe of the Neolithic farming expansion in Europe sheds light on the dispersion process”. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 370.1660, p. 20130373.

Martin, Alicia R et al. (2018). “The critical needs and challenges for genetic architecture studies in Africa”. In: *Current Opinion in Genetics & Development* 53, pp. 113–120.

Mathieson, Iain and Gil McVean (2014). “Demography and the age of rare variants”. In: *PLoS genetics* 10.8, e1004528.

Mazet, O. et al. (2016). “On the importance of being structured: instantaneous coalescence rates and human evolution–lessons for ancestral population size inference?” In: *Heredity* 116, pp. 362–371. DOI: 10.1038/hdy.2015.104.

McEvedy, C. and R. Jones (1978). *Atlas of world population history.* Penguin Books Ltd.

McQuillan, Ruth et al. (2012). “Evidence of inbreeding depression on human height”. In: *PLoS genetics* 8.7, e1002655.

McVean, G. (2009). “A genealogical interpretation of principal components analysis”. In: *Nature* 5.10, e1000686. DOI: 10.1371/journal.pgen.1000686.

McVean, G. and NJ. Cardin (2005). “Approximating the coalescent with recombination”. In: *Philosophical transactions of the Royal Society of London* 360.1459, pp. 1387–1393. DOI: 10.1098/rstb.2005.1673.

Mendel, GJ. (1866). “Versuche über Pflanzen-Hybriden (Experiments Concerning Plant Hybrids)”. In: *Verhandlungen des naturforschenden Vereines in Brünn (Proceedings of the Natural History Society of Brünn)* 103, pp. 3–47.

Menozzi, Paolo, Alberto Piazza, and Luigi Cavalli-Sforza (1978). “Synthetic Maps of Human Gene Frequencies in Europeans: These maps indicate that early farmers of the Near East spread to all of Europe in the Neolithic.” In: *Science* 201.4358, pp. 786–792.

Messer, Philipp W (2013). “SLiM: simulating evolution with selection and linkage”. In: *Genetics* 194.4, pp. 1037–1039.

Meyer, M. et al. (2012). “A High-Coverage Genome Sequence from an Archaic Denisovan Individual”. In: *Science* 338.6104, pp. 222–226. DOI: 10.1126/science.1224344.
Meyer, Matthias et al. (2016). “Nuclear DNA sequences from the Middle Pleistocene Sima de los Huesos hominins”. In: *Nature* 531.7595, pp. 504–507.
Midgley, Magdalena (1992). “TRB culture. The first farmers of the north European plain.” In.
Mijares, Armand Salvador et al. (2010). “New evidence for a 67,000-year-old human presence at Callao Cave, Luzon, Philippines”. In: *Journal of human evolution* 59.1, pp. 123–132.
Moreno-Mayar, J Víctor et al. (2020). “A likelihood method for estimating present-day human contamination in ancient male samples using low-depth X-chromosome data”. In: *Bioinformatics* 36.3, pp. 828–841.
Morgan, TH. et al. (1915). *The Mechanism of Mendelian Heredity*. Henry Holt and Company, New York.
Murray, Gemma GR et al. (2017). “Natural selection shaped the rise and fall of passenger pigeon genomic diversity”. In: *Science* 358.6365, pp. 951–954.
Nakatsuka, Nathan et al. (2020). “ContamLD: estimation of ancient nuclear DNA contamination using breakdown of linkage disequilibrium”. In: *Genome biology* 21.1, pp. 1–22.
Need, Anna C and David B Goldstein (2009). “Next generation disparities in human genomics: concerns and remedies”. In: *Trends in Genetics* 25.11, pp. 489–494.
Nordborg, Magnus (2019). “Coalescent theory”. In: *Handbook of Statistical Genomics: Two Volume Set*, pp. 145–30.
Novembre, J. et al. (2008). “Genes mirror geography within Europe”. In: *Nature* 456, 98–U5.
Novembre, John and Benjamin M Peter (2016). “Recent advances in the study of fine-scale population structure in humans”. In: *Current opinion in genetics & development* 41, pp. 98–105.
Novembre, John and Sohini Ramachandran (2011). “Perspectives on human population structure at the cusp of the sequencing era”. In: *Annual review of genomics and human genetics* 12.1, pp. 245–274.
Orlando, Ludovic et al. (2021). “Ancient DNA analysis”. In: *Nature Reviews Methods Primers* 1.1, pp. 1–26.
Pääbo, Svante (1989). “Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification.” In: *Proceedings of the National Academy of Sciences* 86.6, pp. 1939–1943.
Pagani, Luca et al. (2015). “Tracing the route of modern humans out of Africa by using 225 human genome sequences from Ethiopians and Egyptians”. In: *The American Journal of Human Genetics* 96.6, pp. 986–991.

Palkopoulou, Eleftheria et al. (2015). “Complete genomes reveal signatures of demographic and genetic declines in the woolly mammoth”. In: *Current Biology* 25.10, pp. 1395–1400.

Patin, E. et al. (2017). “Dispersals and genetic adaptation of Bantu-speaking populations in Africa and North America”. In: *Science* 356.6337, pp. 543–546. DOI: 10.1126/science.aal1988.

Patterson, N. et al. (2012). “Ancient admixture in human history”. In: *Genetics* 192, pp. 1065–1093. DOI: 10.1534/genetics.112.145037.

Pearson, Karl (1901). “LIII. On lines and planes of closest fit to systems of points in space”. In: *The London, Edinburgh, and Dublin philosophical magazine and journal of science* 2.11, pp. 559–572.

Peng, Bo and Marek Kimmel (2005). “simuPOP: a forward-time population genetics simulation environment”. In: *Bioinformatics* 21.18, pp. 3686–3687.

Peter, BM. (2016). “Admixture, Population Structure, and F-Statistics”. In: *Genetics* 202, pp. 1485–1501. DOI: 10.1534/genetics.115.183913.

Pickrell, Joseph and Jonathan Pritchard (2012). “Inference of population splits and mixtures from genome-wide allele frequency data”. In: *Nature Precedings*, pp. 1–1.

Pleurdeau, David et al. (2012). “âOf sheep and menâ: earliest direct evidence of caprine domestication in southern Africa at Leopard Cave (Erongo, Namibia)”. In: *PloS one* 7.7, e40340.

Prado-Martinez, Javier et al. (2013). “Great ape genetic diversity and population history”. In: *Nature* 499.7459, pp. 471–475.

Pritchard, JK., M. Stephens, and P. Donnelly (2000). “Inference of Population Structure Using Multilocus Genotype Data”. In: *Genetics* 155.2, pp. 945–959.

Prufer, K. et al. (2014). “The complete genome sequence of a Neanderthal from the Altai Mountains”. In: *Nature* 505, pp. 43–49. DOI: 10.1038/nature12886.

Prugnolle, Franck, Andrea Manica, and François Balloux (2005). “Geography predicts neutral genetic diversity of human populations”. In: *Current Biology* 15.5, R159–R160.

Qin, Pengfei and Mark Stoneking (2015). “Denisovan ancestry in East Eurasian and native American populations”. In: *Molecular biology and evolution* 32.10, pp. 2665–2674.

Racimo, F., G. Renaud, and M. Slatkin (2016). “Joint Estimation of Contamination, Error and Demography for Nuclear DNA from Ancient Humans”. In: *PLOS Genetics* 12.4, pp. 1–27. DOI: 10.1371/journal.pgen.1005972.

Raghavan, Maanasa et al. (2014). “Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans”. In: *Nature* 505.7481, pp. 87–91.

Ragsdale, Aaron P et al. (2022). “A weakly structured stem for human origins in Africa”. In: *bioRxiv*.

Ramachandran, Sohini et al. (2005). “Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa”. In: *Proceedings of the National Academy of Sciences* 102.44, pp. 15942–15947.

Rasmussen, M. et al. (2010). “Ancient human genome sequence of an extinct Palaeo-Eskimo”. In: *Nature* 463.7282, pp. 757–762. DOI: 10.1038/nature08835.

Rasmussen, M. et al. (2015). “The ancestry and affiliations of Kennewick Man”. In: *Nature* 523, pp. 455–458. DOI: 10.1038/nature14625.

Rasmussen, MD. et al. (2014). “Genome-wide inference of ancestral recombination graphs”. In: *PLoS Genet* 10, e1004342.

Rasmussen, Morten et al. (2011). “An Aboriginal Australian genome reveals separate human dispersals into Asia”. In: *Science* 334.6052, pp. 94–98.

Reich, D. et al. (2010). “Genetic history of an archaic hominin group from Denisova Cave in Siberia”. In: *Nature* 468, pp. 1053–1060. DOI: 10.1038/nature09710.

Reich, David et al. (2009). “Reconstructing Indian population history”. In: *Nature* 461.7263, pp. 489–494.

Reich, David et al. (2011). “Denisova admixture and the first modern human dispersals into Southeast Asia and Oceania”. In: *The American Journal of Human Genetics* 89.4, pp. 516–528.

Reid, Lawrence A (2013). “Who are the Philippine negritos? Evidence from language”. In: *Human Biology* 85.1/3, pp. 329–358.

Renaud, G. et al. (2015). “Schmutzi: Estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA”. In: *Genome Biol* 16, p. 224.

Renaud, Gabriel et al. (2017). “gargammel: a sequence simulator for ancient DNA”. In: *Bioinformatics* 33.4, pp. 577–579.

Rosenberg, N. et al. (2002). “Genetic structure of human populations”. In: *Science* 298, pp. 2381–2385. DOI: 10.1126/science.1078311.

Sadr, Karim (1998). “The first herders at the Cape of Good Hope”. In: *African Archaeological Review* 15.2, pp. 101–132.

Sanger, Frederick, Steven Nicklen, and Alan R Coulson (1977). “DNA sequencing with chain-terminating inhibitors”. In: *Proceedings of the national academy of sciences* 74.12, pp. 5463–5467.

Sanger, Frederick et al. (1977). “Nucleotide sequence of bacteriophage φX174 DNA”. In: *nature* 265.5596, pp. 687–695.

Sankararaman, Sriram et al. (2016). “The combined landscape of Denisovan and Neanderthal ancestry in present-day humans”. In: *Current Biology* 26.9, pp. 1241–1247.

Sawyer, S. et al. (2012). “Temporal Patterns of Nucleotide Misincorporations and DNA Fragmentation in Ancient DNA”. In: *PLoS ONE* 7.3, e34131.

Scally, A. and R. Durbin (2012). “Revising the human mutation rate: implications for understanding human evolution”. In: *Nature Reviews Genetics* 13, pp. 745–753.

Scerri, EML. et al. (2018). “Did Our Species Evolve in Subdivided Populations across Africa, and Why Does It Matter?” In: *Trends in Ecology & Evolution* 33.8, pp. 582 –594. ISSN: 0169-5347. DOI: 10.1016/j.tree.2018.05.005.

Schiffels, S. and R. Durbin (2014). “Inferring human population size and separation history from multiple genome sequences”. In: *Nature Genetics* 46, pp. 919–925. DOI: 10.1038/ng.3015.

Schlebusch, Carina M et al. (2020). “Khoe-San genomes reveal unique variation and confirm the deepest population divergence in Homo sapiens”. In: *Molecular biology and evolution* 37.10, pp. 2944–2954.

Schlebusch, CM. et al. (2012). “Genomic Variation in Seven Khoe-San Groups Reveals Adaptation and Complex African History”. In: *Science* 338.6105, pp. 374–379. DOI: 10.1126/science.1227721.

Schlebusch, CM. et al. (2017). “Southern African ancient genomes estimate modern human divergence to 350,000 to 260,000 years ago”. In: *Science* 358.6363, pp. 652–655. DOI: 10.1126/science.aao6266.

Schraiber, JG. (2017). “Assessing the Relationship of Ancient and Modern Populations”. In: *Genetics*. DOI: 10.1534/genetics.117.300448.

Ségurel, Laure, Minyoung J Wyman, and Molly Przeworski (2014). “Determinants of mutation rate variation in the human germline”. In: *Annu Rev Genomics Hum Genet* 15.1, pp. 47–70.

Siva, Nayanah (2008). “1000 Genomes project”. In: *Nature biotechnology* 26.3, pp. 256–257.

Sjödin, P., J. McKenna, and M. Jakobsson (2021). “Estimating divergence times from DNA sequences”. In: *Genetics* 217.4, iyab008.

Sjödin, P., P. Skoglund, and M. Jakobsson (2014). “Assessing the Maximum Contribution from Ancient Populations”. In: *Molecular Biology and Evolution* 31.5, pp. 1248–1260. DOI: 10.1093/molbev/msu059.

Skoglund, P., A. Götherström, and M. Jakobsson (2011). “Estimation of Population Divergence Times from Non-Overlapping Genomic Sequences: Examples from Dogs and Wolves”. In: *Molecular Biology and Evolution* 28.4, pp. 1505–1517. DOI: 10.1093/molbev/msq342.

Skoglund, P. et al. (2012). “Origins and genetic legacy of Neolithic farmers and hunter-gatherers in Europe”. In: *Science* 336, pp. 466–469.

Skoglund, P. et al. (2014). “Investigating population history using temporal genetic differentiation”. In: *Mol Biol Evol* 31, pp. 2516–2527. DOI: 10.1093/molbev/msu192.

Skoglund, P. et al. (2017). “Reconstructing Prehistoric African Population Structure”. In: *Cell* 171, pp. 59–71. DOI: 10.1016/j.cell.2017.08.049.

Smith, Andrew B (1992). “Origins and spread of pastoralism in Africa”. In: *Annual Review of Anthropology*, pp. 125–141.

Speidel, L. et al. (2019). “A method for genome-wide genealogy estimation for thousands of samples”. In: *Nature Genetics* 51, pp. 1321–1329.
Stoneking, Mark and Johannes Krause (2011). “Learning about human population history from ancient and modern genomes”. In: *Nature Reviews Genetics* 12.9, pp. 603–614.
Teixeira, João C et al. (2021). “Widespread Denisovan ancestry in Island Southeast Asia but no evidence of substantial super-archaic hominin admixture”. In: *Nature Ecology & Evolution* 5.5, pp. 616–624.
Terhorst, J., JA. Kamm, and YS. Song (2017). “Robust and scalable inference of population history from hundreds of unphased whole-genomes”. In: *Nature genetics* 49, pp. 303–309.
Tishkoff, SA. et al. (2009). “The Genetic Structure and History of Africans and African Americans”. In: *Science* 324, pp. 1035–1044.
Veeramah, Krishna R et al. (2012). “An early divergence of KhoeSan ancestors from those of other modern humans is supported by an ABC-based analysis of autosomal resequencing data”. In: *Molecular biology and evolution* 29.2, pp. 617–630.
Venter, J Craig et al. (2001). “The sequence of the human genome”. In: *science* 291.5507, pp. 1304–1351.
Vicente, Mário et al. (2019). “Genetic affinities among southern Africa hunter-gatherers and the impact of admixing farmer and herder populations”. In: *Molecular biology and evolution* 36.9, pp. 1849–1861.
Wakeley, J. (2009). *Coalescent Theory: An Introduction*. 1st. Greenswood Village, Colorado: Roberts & Company Publishers.
Wang, K. et al. (Mar. 2019). “Tracking human population structure through time from whole genome sequences”. In: *PLOS Genetics* 16.3, pp. 1–24. DOI: 10.1371/journal.pgen.1008552.
Wang, K. et al. (2020). “Ancient genomes reveal complex patterns of population movement, interaction, and replacement in sub-Saharan Africa”. In: *Science Advances* 6.24, eaaz0183.
Watson, James D and Francis HC Crick (1953). “Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid”. In: *Nature* 171.4356, pp. 737–738.
Wheeler, David A et al. (2008). “The complete genome of an individual by massively parallel DNA sequencing”. In: *nature* 452.7189, pp. 872–876.
White, Tim D et al. (2003). “Pleistocene *Homo sapiens* from Middle Awash, Ethiopia”. In: *Nature* 423, pp. 742–747.
Willerslev, E. and A. Cooper (2005). “Ancient DNA”. In: *Proceedings of the Royal Society B: Biological Sciences* 272, pp. 3–16.
Wiuf, C. and J. Hein (1999). “Recombination as a point process along sequences.” In: *Theor Popul Biol* 55, pp. 248–259. DOI: 10.1006/tpbi.1998.1403.
Wood, Bernard and Mark Collard (1999). “The human genus”. In: *Science* 284.5411, pp. 65–71.

Wright, S. (1931). “Evolution in Mendelian populations”. In: *Genetics* 16, pp. 97–159.
– (1932). “The roles of mutation, inbreeding, crossbreeding and selection in evolution”. In: vol. 1, pp. 356–366.
Wright, Sewall (1949). “Population structure in evolution”. In: *Proceedings of the American Philosophical Society* 93.6, pp. 471–478.
Zhang, Dongju et al. (2020). “Denisovan DNA in late Pleistocene sediments from Baishiya Karst Cave on the Tibetan plateau”. In: *Science* 370.6516, pp. 584–587.

Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology 2221

Editor: The Dean of the Faculty of Science and Technology

A doctoral dissertation from the Faculty of Science and Technology, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology. (Prior to January, 2005, the series was published under the title "Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology".)

Distribution: publications.uu.se
urn:nbn:se:uu:diva-489300

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2022