

Chapter 5: Future directions

5.1 Future directions for the population history of Sahul

The broad outlines of the genetic history of Sahul are starting to become clear, and that history appears to be relatively simple at the coarsest level of resolution, with genetic isolation from the rest of the world from the initial settlement ~ 50 kya until the period of Southeast Asian contact and European colonization. However, much is still unknown about population history within Sahul, especially Australia, and a simple relationship to worldwide populations does not mean that the internal history was also simple. Indeed, using the well-studied history of Europe as an example, much of the complex Holocene population changes and admixture there occurred between populations that were all different branches of the broad western Eurasian ancestry. It is entirely possible that similarly dramatic processes have occurred within Sahul, involving differentiated branches of the broad Sahul ancestry, while not altering the relationships to worldwide outgroups. The first in-depth look at population structure in PNG, presented in chapter 3, hints at such complex processes, at least here. Sahul represents an opportunity to study a part of the world in isolation, providing independent insight into patterns of human evolution and the behaviour of populations, especially as many of the major cultural changes in Eurasia and Africa appear to be connected to some extent.

One of the major events that we do currently have knowledge of is the seemingly deep split time between Aboriginal Australians and Papuans. This establishes Sahul as a continent with a present-day population structure almost as old as that between western and eastern Eurasia. A better understanding of this split is therefore of great interest. Firstly, more confidence is needed in the actual split time estimates, with e.g. current MSMC estimates being associated with technical uncertainty surrounding haplotype phasing quality. Experimental phasing, e.g. with 10x Genomics as described in Chapter 3 here, might be a promising approach to resolve this. Secondly, more comprehensive sampling of populations in northern Australia, as well the Torres Strait Islands between Australia and New Guinea, is needed to better characterize the seemingly very sharp genetic divide between the two. Current data, as described in Chapter 2, already gives some hints of a closer relationship to New Guineans in some northern Australian individuals. Further work is needed to determine if this just reflects admixture in relatively recent times, or if there have been more long-standing contacts. The separation between Aboriginal Australian and New Guinean populations, long before rising sea levels, has no obvious geographical, climatic or cultural explanation. It clearly predates the development of agriculture in PNG, excluding this as a cause of cultural differentiation. In terms of geography and climate, northern Australia with its rain forests might in some respects be more similar to New Guinea than to the rest of Australia. Further understanding of the causes for this early split might need to come from fields other than population genetics.

Another question in the early history of Sahul relates to the peopling of the Melanesian islands,

and the relationship of the populations here to those in the rest of continent. The more or less implicit assumption in the literature so far, seemingly quite natural on geographical grounds, is that Melanesian islanders are a sister group to Papua New Guineans, to the exclusion of Aboriginal Australians. The analyses here of Bismarck Archipelago groups (New Ireland and New Britain) are consistent with this; however, another study reported an increased Aboriginal Australian affinity in certain groups further east in the Pacific (Skoglund, et al. 2016). In some explorative analyses presented here, the Bismarck groups also displayed difficult-to-interpret behaviour in this direction. Further analyses, and more comprehensive sampling across near and far Melanesia, are needed to better understand these relationships. Another question is the time-scale of separation between Melanesian islanders and New Guinea mainlanders. The array genotypes indicate strong allele frequency differentiation, but whole-genome sequencing data will likely be needed to get at the question of actual split time. Such analyses will also need to overcome the confounding effects of the Southeast Asian admixture that is present in all island populations.

A question related to this, and one that will increase in relevance with an increasingly detailed understanding of Sahul and Melanesian population structure, is that of the origin of the Sahul-related ancestry of Polynesians. Seemingly all Polynesian groups, as far as Hawaii and Rapa Nui, carry on the order of 20% Sahul-related ancestry. Little is known about where they picked this up: whether it was from the mainland of New Guinea, the large Bismarck islands, smaller islands further out, or some combination of these.

While the big picture of Sahul history is that of isolation from the rest of the world, it has not been completely isolated. In New Guinea there has of course been Southeast Asian admixture in the Holocene. While the dominant view is that of a relatively rapid expansion in the last few thousand years originating in Taiwan, there are also suggestions of earlier migrations (Soares, et al. 2016). There is also still uncertainty about the timing of arrival from Southeast Asia of certain domesticated crops, as well as domesticated pigs. More genome-wide data from many different Southeast Asian groups might be needed to better characterize the history of migration into New Guinea. In Australia, the dingo is evidence of external contacts, and while no Southeast Asian or other pre-European colonization admixture has been found in Aboriginal Australian genomes, more geographically comprehensive studies are needed before it can be ruled out completely. Further, in particular genome-wide, studies of dingos, as well as the related New Guinean Singing dogs, and Southeast Asian and perhaps Polynesian dogs might also provide answers as to who brought these animals to Sahul.

Major questions remain on the population history within New Guinea and within Australia, histories we are now only beginning to scratch the surface of. In Australia, very little is known about population structure, and studies are complicated by the widespread and extensive European admixture as well as the fact that many Aboriginal Australians are unaware of their deeper geographical origins because of forced movements during European colonial rule. Efforts to make use of

well-documented museum collections of hair samples, many probably also predating much of the European admixture, might be a promising way to overcome this issue (Tobler, et al. 2017). From current data, population structure at least between western and eastern groups appears to be fairly old, at least pre-Holocene. A study of mitochondrial genomes suggested that separation between present-day Aboriginal Australian groups even dates back all the way to the initial peopling ~50 kya (Tobler, et al. 2017). However, the limited and stochastic information provided by uniparental lineages, especially in the presence of potentially strong drift in small sub-populations, arguably makes such a conclusion difficult to draw, and furthermore it is not compatible with current estimates of the split time from Papua New Guineans. In any case, a major challenge is to explain how potentially fairly deep genetic divergences within Australia relate to the fact that a single language family, the Pama-Nyungan family, is spoken across 90% of the continent. The shared languages imply contact on the timescale of the last few, perhaps at most 10, thousand years, and lessons from the rest of the world suggest that languages typically spread with genes. There was no agricultural or similar transition in Australia that could have driven this, but there was a change in stone tool technology, and population expansions among hunter-gatherers cannot be ruled out. Indeed, such expansions have been demonstrated from ancient DNA in Pleistocene Europe (Fu, et al. 2016). Understanding population dynamics within Australia, including studying the non-Pama-Nyungan speaking groups in the northern parts of the continent, should be a key focus of future studies. Lastly, the genetic history of Tasmania remains virtually completely unknown, except for mitochondrial studies which at least have linked them to mainland Aboriginal Australians (Presser, et al. 2002; McAllister, et al. 2013). No unadmixed Aboriginal Tasmanians remain today, meaning that any insights would need to come from analyses of admixed genomes or ancient DNA.

In New Guinea, further work is needed to understand the population genetic consequences of the agricultural transition, and the extent to which current population structure was shaped by this as opposed to earlier, or later, processes. Whole-genome sequencing data from larger number of lowland groups (only one, from the Sepik region, was available for the analyses presented here) is needed to obtain a better understanding of the timing of separation between highlanders and lowlanders, and the demographic histories of the latter. This might illuminate whether or not the separation definitely occurred prior to the development of agriculture in the highlands, and thereby how agriculture and Trans-New Guinea languages reached the lowlands. Additionally, data from the Indonesian half of New Guinea island, so far very much underrepresented in genetic studies relative to the PNG half, is needed to complete the picture of New Guinean population structure.

Very little is known about positive selection in the history of the populations of Sahul. An analysis in (Malaspinas, et al. 2016) identified a number of candidate loci under selection in Aboriginal Australians, with possible links to adaptation to desert cold and dehydration; however, these results were tentative. Positive selection studies in other populations have uncovered several interesting instances of adaptation to local environments and lifestyles, and given the long, independent history of Sahul there might be good opportunities to make more here. Such discoveries could provide

general insights into human biology and physiology, as well as insights specifically into the population and lifestyle histories of humans in Sahul. A question of particular interest is if any genetic variants were exposed to positive selection during the agricultural transition in New Guinea, and if so, how that process compares to agricultural adaptations in other parts of the world.

Ultimately, ancient DNA will be necessary to gain a truly in-depth understanding of Sahul population history. Data from ancient individuals would illuminate every aspect of this history, including the timing and nature of the Aboriginal Australian and Papuan split, positive selection, population structure and dynamics within Australia and the agricultural transition in New Guinea. While e.g. the first of these questions might require DNA from samples that are very old, perhaps on the order of 30 ky or more, the latter two would be greatly informed by samples that are 10 ky old or less. While successful DNA extraction has traditionally been mostly limited to samples from colder environments, recent technological improvements have enabled analyses of samples from e.g. eastern Africa (Gallego Llorente, et al. 2015), the Near East (Lazaridis, et al. 2016; Haber, et al. 2017) and Pacific islands (Skoglund, et al. 2016). There are thus real prospects for ancient DNA from Sahul, especially from the drier environments of Australia. The humid climate of PNG will likely prove more challenging, but on the other hand, the highlands are relatively cool and even feature snow-covered mountaintops. Ancient DNA is currently transforming our understanding of human evolutionary history in many parts of the world, and will be essential in Sahul as well.

5.2 Future directions for human population history generally

While the genetic history of western Eurasian and Native American populations is now becoming understood in some detail, much is still unclear about the history of eastern Eurasians. This includes the early events in the diversification between the ancestors of present-day East Asians, the populations of Sahul and those of Island Southeast Asia, as well as the admixture with Denisovans.

A number of so-called Negrito populations in Southeast Asia appear to share some parts of their early history with Aboriginal Australian and Papuans, after the separation from the ancestors of East Asians. This includes certain populations in the Philippines and Malaysia, as well as the Andaman islanders. However, the shape of the relationships of these various groups to each other is not understood with high confidence. Many of these Southeast Asian groups display a higher affinity to East Asians than Aboriginal Australian and Papuans do, likely due to East Asian-related admixture. This admixture could be relatively recent, i.e. occurring in the last 10 ky following the expansions of agriculturalists across Southeast Asia, and/or potentially more ancient, occurring during the early diversification of these lineages ~50 kya. The observation that certain Philippine groups harbour more Denisovan admixture than expected given their levels of East Asian versus Sahul related ancestry suggests a more complex history than one of simple East Asian dilution (Reich, et al. 2011). Some models of the population split and admixture topologies relating these various lineages have been put forth (Reich, et al. 2011; Lipson and Reich 2017), but further work is needed. Furthermore, it has arguably not been conclusively determined if the small amounts

(on the order of 0.1%) of Denisovan ancestry that is present in South and East Asian populations derives from the same admixture event as the material that is present in large amounts (~4%) in Aboriginal Australians and Papuans. Analyses of predicted Denisovan haplotypes across these groups, including the Philippine Negritos, might provide further insights into the details of Denisovan gene flow into modern humans, as well as into the relationships between the modern human groups themselves.

The relationships between these populations also need to be placed onto a temporal axis – little is known about e.g. the genetic split times between the various Negrito groups, Sahul populations and East Asians. Current split time estimates between Sahul populations and East Asians are associated with technical uncertainty and need to be made more confident, e.g. through application of MSMC to experimentally phased genomes as well as independent dating methods. The determination of split times between Southeast Asian groups will likely be complicated by complex histories of admixture, such that simple pairwise split estimates might not be very meaningful. Decomposition of such estimates given knowledge of overall ancestry proportions (Pagani, et al. 2016), or more explicit temporal model fitting based on e.g. the site frequency spectrum (Excoffier and Foll 2011) or rare variant sharing (Schiffels, et al. 2016), might be fruitful.

There are observations that hint at the possibility that populations with a Negrito-like ancestry were once more widespread, not just across Island Southeast Asia, but perhaps even mainland Asia. First, the genetic split time between Andamanese islanders, in the past sometimes informally hypothesized to have been isolated in the Indian Ocean since shortly after the migration of humans out of Africa, and other Eurasian populations appears to have occurred closer to ~20 kya (Mondal, et al. 2016). Second, a slightly higher affinity towards Negrito and Sahul populations among some South American groups relative to other Native Americans (Skoglund, et al. 2015) suggests structure in the Siberian source populations ~20 kya that was correlated to Negrito affinity. Ancient DNA from East Asia might reveal populations with this kind of ancestry, that perhaps have been replaced or absorbed by expanding groups harbouring the ancestry that today dominates across mainland East Asia.

In mainland East Asia, very little is known about the processes underlying current population structure. One of the key aspects that needs an explanation is the genetic homogeneity of the region, which overall is comparable to Europe, and strikingly different from Papua New Guinea. For example, F_{ST} between Japanese and Vietnamese is approximately 1%. It is tempting to speculate that, similarly to Europe, this is the result of one or more recent expansions after the development of agriculture. Studying the factors behind the genetic homogenization of East Asia, and comparing them to what did or did not happen in Europe, Papua New Guinea and elsewhere, will provide further fuel to this “comparative population genetics” approach to understanding the driving forces in human history. Further work is needed in East Asia, and, as elsewhere, ancient DNA will be required to achieve an in-depth understanding.

In a global context, the biggest gaps in our understanding of human population history are arguably in Africa. A few basic features seem to be firmly established, including the early divergence of the Khoe-San people of southern Africa, the divergence of central African rainforest hunter-gatherers after that, the diversification of what is largely the present-day western and eastern African ancestries after that, large-scale migration and admixture following the Bantu expansion in the last few thousand years, and gene flow from non-African sources in the Middle East also in the last few thousand years. Some of the earlier population structure has likely been obscured by the recent expansions and admixture. There is still great uncertainty about the shape and timing of the early diversification events within Africa, which likely occurred somewhere between 100-200 kya, but potentially even earlier. Understanding the time depth of human population structure is of key interest not least because it puts a lower bound on the emergence of anatomically modern human traits.

An increased understanding of African population history might also aid the understanding of the dispersal out of Africa. In principle, one kind of observation that would provide evidence that different non-African populations derive ancestry from different migrations out of Africa would be differences in their relationships to African population structure. For example, if Aboriginal Australians and Papuan showed higher affinity to western Africans while Eurasian showed higher affinity to eastern Africans, this would constitute such evidence. A major obstacle, however, is the extensive backflow that has occurred into Africa from sources more closely related to Middle Easterners and Europeans than to other worldwide populations, affecting especially eastern, but also southern, Africa. Further to this, a key question is where in the African population history topology the group that became non-Africans fits in, and the related question of which present-day African populations are most closely related to non-Africans. Because of population dynamics within Africa in the last 50 ky, it is, however, very unlikely that the direct sister group to non-Africans will still exist in un-admixed form. Modelling approaches that incorporate past admixture events will be required to elucidate the complex population history within Africa.

In recent years there have been a number of archaeological findings of human presence outside of Africa at times earlier than predicted by current population genetic models, e.g. in China 100 kya (Liu, et al. 2015), North America 120 kya (Holen, et al. 2017), Australia 65 kya (Clarkson, et al. 2017), as discussed in Chapter 2, and Sumatra (Westaway, et al. 2017) 63-73 kya, and a similar suggestion from analyses of Neanderthal genomes (Kuhlwilm, et al. 2016). An important question is if these are traces of people that are the ancestors of present-day non-Africans, and that we are currently underestimating the time depth of human dispersal, or if they reflect earlier migrating groups that died out and did not contribute detectable ancestry to present-day people. There are observations that arguably provide reasonably firm upper bounds on the timing of certain events, e.g. the timing of Neanderthal admixture (Fu, et al. 2014) or the divergence of non-African uniparental chromosomes from African ones (Fu, Mittnik, et al. 2013; Poznik, et al. 2016), and which therefore are incompatible with at least some of the earliest archaeological findings being

associated with ancestors of present-day populations. However, recent findings from ancient DNA are showing that humans have often been highly mobile in the past, and individuals from ancient populations that do not seem to have contributed to present-day people have been discovered (Fu, et al. 2014; Fu, et al. 2015), so the idea of small groups of humans moving large distances and living in a new area for a while before dying out might now appear more plausible.

Another question that is still unresolved is that of whether or not there was admixture into modern humans from any archaic human groups within Africa. In the absence of ancient DNA from any such groups, studies analysing particularly divergent haplotypes have suggested evidence for such “ghost admixture” in certain present-day African populations (Hammer, et al. 2011; Hsieh, Woerner, et al. 2016). Such analyses are necessarily highly statistical and indirect, but perhaps larger amounts of data from diverse African populations will allow these analyses to be more conclusive. However, while conditions for ancient DNA preservation are poor in most of Africa, ideally direct tests will be made possible at some point, with DNA isolated from human remains or possibly even recovered from the environment (Willerslev, et al. 2003; Slon, et al. 2017).

5.3 Concluding remarks

It is truly an exciting time for the field of human population genetics. Many important findings have been made even just while the work described in this thesis was carried out, and more will undoubtedly come in the near future. It is becoming increasingly apparent that ancient DNA is extremely valuable for understanding population histories. While inference from modern genomes often requires highly sophisticated analyses to infer properties of past populations, the power of ancient DNA is that it allows us to observe genomes from such populations directly. However, there is still arguably value in generating data from modern genomes. The higher quality and larger amounts of data typically obtained serve as useful reference points for ancient data, and also more readily enable certain analyses that dependent on high data quality. For example, ancient DNA studies have not provided much insight into changes in effective population sizes over time, while analyses of modern genomes have. There might also be practical limits on how much ancient DNA will be generated in certain parts of the world and from certain time periods. In particular, to address the deep history of humanity in Africa on the order of ~ 200 kya, or maybe even earlier, we might still have to rely on indirect inferences from combinations of modern and ancient genomes, rather than just observations of ancient DNA directly from those periods. Nonetheless, ancient DNA will likely drive most progress in the field from now onwards.

Over the next few years and onwards, very large numbers of human genomes will be sequenced, mainly for biomedical purposes. There are potentially exciting prospects for gaining population-genetic insights from such data. While simply having a very large number of genomes from a given population is unlikely to provide any dramatic gains to the understanding of the history of that population and its relationships to others, such datasets might greatly inform the more

functional and mechanistic aspects of human genetics. This could include questions related to the mechanisms and consequences of mutation and recombination, gene essentiality and disease penetrance, positive and purifying selection and classical genetics concepts such as epistasis, dominance and heritability. With a greater functional characterization of human biology, including an ever-expanding list of variants that are associated with trait variation, will also come challenges and opportunities to understand this biology in a population history framework.

Lastly, efforts need to be made to ensure that not only people in the more economically privileged parts of the world benefit from the progress in human genomics. This argument is often made in the context of medical genetics, where results from studies carried out in populations of European ancestry might not translate well to other populations, but a version of the argument could also be applied to the study of population history. Currently, the part of the world with by far the best understood genetic history is Europe. There are practical and logistical reasons for this, including good conditions for ancient DNA preservation in many parts of Europe and comprehensive previous archaeological work, but greater efforts need nonetheless be made to study other parts of the world in greater detail as well. Such studies will be needed to gain a truly comprehensive picture of human evolutionary history and diversity.