

# Southeast Asian islands and Oceania: human genetics

---

Murray P. Cox

*This chapter discusses the genetic record of human migration into island Southeast Asia and Oceania, firstly during the Pleistocene, and later during the Neolithic with its associated populations of Malayo-Polynesian speakers. The picture is one of intergradation between Asian and indigenous (western Pacific) population components.*

The Neolithic period was a time of great change in island Southeast Asia and Oceania. Many characteristics of the peoples living in this region today have their foundation in the Neolithic, including their languages, modes of subsistence, and many aspects of modern culture. The same is true of their biology. The Neolithic period saw major population movements from mainland Asia, substantially changing the biological makeup of communities in what are today the island nations of Taiwan, the Philippines, Indonesia, Papua New Guinea, and Solomon Islands. Descendants of these Neolithic voyagers ultimately developed the advanced seafaring skills necessary to explore and settle the remote islands of the vast Pacific Ocean.

## Genetics as history

History can be reconstructed in many different ways. It can be inferred from the distribution and relationships of languages, an approach first applied in the Pacific during the 1770s by the natural historian Johann Reinhold Forster during Captain James Cook's second voyage. History can be reconstructed from the distribution and relationships of archaeological artifacts. Importantly, however, history is also carried in the DNA of living people, as well as being preserved in the bones of ancient individuals.

Such genetic evidence is often considered to be of recent provenance. In fact, the earliest studies of genetic diversity in the Indo-Pacific region were carried out in the

1920s by Bais and Verhoef (1924) and Heydon and Murphy (1924). These first papers reported ABO blood group frequencies in Java and New Guinea, but new studies were quickly initiated across island Southeast Asia and Oceania. Contrary to popular opinion, geneticists have been helping to reconstruct Pacific prehistory for the last 90 years.

These early studies examined so-called “classical” genetic markers, mainly blood proteins like the well-known ABO blood groups. Markers like this dominated human population genetics for the next sixty years, and proved to be remarkably informative about broad regional trends. One clear pattern was the distinction between populations in Melanesia on the one hand, and island Southeast Asia and Polynesia on the other. Several classical genetic markers exhibit frequency differences between these groups (Cavalli-Sforza et al. 1994; Cox 2008; Mourant et al. 1976). Nevertheless, these early genetic markers lacked resolution over small geographical areas, and reconstructing fine-scaled, directional population movements remained out of reach.

By the mid-1980s, newly developed techniques, particularly the Polymerase Chain Reaction (PCR), were allowing geneticists to study DNA sequences directly for the first time. This was a landmark era for human population genetics and the field has grown rapidly over the past twenty-five years. As with earlier technological developments, these new approaches were quickly applied to questions of Indo-Pacific prehistory. In 1989, Hertzberg and colleagues identified a DNA variant that is largely restricted to Polynesians and connects them firmly, and quite recently, back to the Asian mainland. This study was the first to reconstruct directional movements across the Pacific, a key goal of human population genetics that is still pursued today.

## Genetic prehistory of the Indo-Pacific region

Hertzberg and his colleagues studied just one genetic region – mitochondrial DNA (mtDNA), but their work was followed later by studies of the Y chromosome. Mitochondrial DNA is inherited maternally – it is passed down only from a mother to her offspring, while the Y chromosome is inherited paternally – it is passed down only from a father to his sons. These simple modes of transmission provide a tractable model system to study the human past, and most genetic research in the Pacific has revolved around these two important genetic regions (summarized in Hill et al. 2007; Karafet et al. 2010; Tabbada et al. 2010).

However, within the last few years, a third generation of genetic data has appeared on the scene. Multiple ancestry-informative markers drawn from across the human genome (i.e. the autosomal chromosomes) are increasingly becoming the new gold standard in molecular anthropology. Although only a few studies have examined Indo-Pacific populations, they are already providing a broad outline of the region’s prehistory, particularly the statistical distribution of Asian-Melanesian ancestry across large swathes of island Southeast Asia and Oceania (Cox et al. 2010). In the west, Taiwanese aboriginals have almost 100 percent Asian ancestry. In island Melanesia, Asian ancestry

drops to around 20 percent, but then increases again further east, reaching around 80 percent Asian ancestry in the Polynesian gene pool (Friedlaender et al. 2008; Kayser et al. 2008).

This uneven distribution of Asian-Melanesian ancestry is a direct outcome of a complex history stretching back nearly 50,000 years (O'Connell & Allen 2004). The islands of Southeast Asia and Oceania essentially form a cul-de-sac, characterized by recurrent, unidirectional population incursions from mainland Asia. These movements occurred in at least three phases, broadly aligned with the Pleistocene, Neolithic, and historic eras.

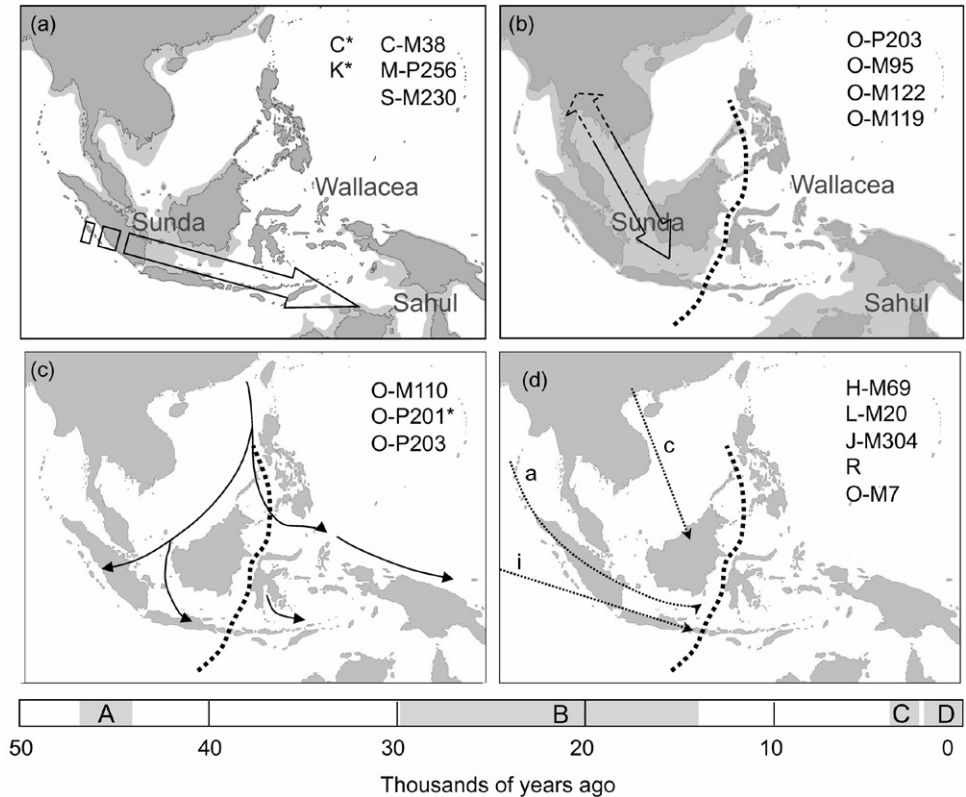
In addition, the genetic history of women appears to have differed slightly from that of men. This disparity is caused by contrasting social histories, particularly marriage practices that affect men and women in different ways. Pulses of immigration during the Neolithic were dominated by matrilineal communities (Jordan et al. 2009), where wealth and status were dispersed through maternal lines. These communities apparently also practiced matrilocal residence, whereby married women remained in their home community, while their husbands moved. This process favored the loss of incursive male diversity (e.g. Asian Y-chromosome lineages) while retaining incursive female diversity (e.g. Asian mtDNA lineages), a process that was recently detected at a genome-wide scale across island Southeast Asia (Cox et al. 2010).

In combination, these processes produced large-scale patterns of genetic diversity that persist in island Southeast Asia and Oceania even today. Broadly speaking, around 60 percent of island Southeast Asian and Oceanic individuals across the region, including both Austronesian and non-Austronesian speakers, carry mtDNA lineages that have existed in the Pacific since the Pleistocene. Conversely, approximately 40 percent of such individuals carry mtDNA lineages that can be traced back to incursive population movements into the Pacific during the Neolithic, often ultimately from China and mainland Southeast Asia. Of course, the exact proportions vary widely across island Southeast Asia and Oceania; for instance, Polynesians largely carry mtDNA lineages that trace back to population movements from mainland Asia during the Neolithic, while New Guinea highlanders predominantly carry mtDNA lineages with local antecedents stretching far back into the Pleistocene. Between these two extremes lies a wide range of regional variation. This is explored in greater detail below.

In the following sections, we will explore how this wealth of mtDNA and Y-chromosome evidence is filling in the regional details of this broad historical outline (also see Kayser 2010).

### **Pleistocene continuity**

Population movements during the Neolithic were inscribed on a palimpsest of genetic diversity laid down during the Pleistocene. Much of the Indo-Pacific region was settled from around 50 kya. These early settlers were part of the first waves of modern humans who left Africa to settle in Eurasia and the Americas, and the descendants of these pioneers continue to dominate parts of the region today.



**Figure 37.1** A multistage settlement model for island Southeast Asia: (a) Initial wave of colonization 40–50 kya, (b) Paleolithic contribution from mainland Asia, (c) Austronesian expansion, and (d) migration in historic times. Y-chromosome haplogroups and paragroups listed in each panel are postulated to have arrived during each stage. Arrows do not denote precisely defined geographic routes. The dotted arrowhead in panel (b) indicates possible bidirectional gene flow. The dotted line in panels (b), (c), and (d) represents Wallace’s biogeographic boundary. Small letters associated with arrows in panel (d) refer to migrations from India (i), Arabia (a) and China (c). Shaded areas with capital letters (A–D) under panels refer to approximate time frames for each stage of colonization. Coastlines in panels (a) and (b) are drawn with sea levels 50 and 120 m below current levels. Figure reproduced unmodified from Karafet et al. (2010) by permission of Oxford University Press.

On the tree of mtDNA diversity (Van Oven & Kayser 2009), these first settlers are represented by the major regional haplogroups P and Q, which today are largely restricted to the Papuan strongholds of eastern Indonesia, New Guinea, and island Melanesia. Populations in this region are also characterized by many unique, low-frequency lineages that branch off early from the mtDNA tree (i.e. shortly after the African diaspora). These lineages reflect the extraordinary isolation and population substructure of this region, a key characteristic of Melanesia to this day.

The same general pattern can be seen on the tree of Y-chromosome diversity (Karafet et al. 2008). Haplogroup C, the first non-African branch, is frequent in eastern Indonesia, New Guinea, and island Melanesia, particularly in its derived form, C-M38\*. However, the Out of Africa expansion also caused a rapid radiation of clade K, which includes haplogroups S and M. As with mtDNA lineages P and Q, these Y-chromosome haplogroups are essentially restricted to eastern Indonesia, New Guinea, and island Melanesia, reflecting their long history and relative isolation in this region.

Soon afterwards, movements from the Asian mainland may have introduced the first haplogroup O lineages into western parts of island Southeast Asia.

These genetic patterns dominated the Indo-Pacific for the first four-fifths of its history (from roughly 50 to 10kya). Indeed, in New Guinea and on neighboring islands, Pleistocene lineages essentially reach a combined frequency of 100 percent even today. However, shortly after 10kya, the development of agriculture in China, coupled with its rapid spread across island Southeast Asia after four thousand years ago, radically changed the population dynamics in most parts of the Pacific (Bellwood 2005; and see chapters 35, 36, 41). Immigration from mainland Asia, in part flowing through Taiwan and likely associated with the spread of Austronesian languages, substantially altered the genetic constitution of many Indo-Pacific populations. These patterns suggest that related, but slightly different, processes occurred in western island Southeast Asia, eastern island Southeast Asia, and Oceania. These regional variations are discussed in greater detail below.

## The Neolithic period

### Western island Southeast Asia

The weight of evidence suggests that Neolithic populations followed a two-pronged expansion into island Southeast Asia. One of these movements coursed down into western island Southeast Asia, most likely from Taiwan, particularly affecting the western Indonesian islands of Sumatra, Borneo, Java, and Bali. These Neolithic settlers are represented by a wide range of lineages on the tree of mtDNA diversity. Haplogroup E occurs across island Southeast Asia, and although it may in part have Pleistocene connections, subgroup E1a appears to link island Southeast Asia with indigenous populations on Taiwan. In comparison, the origin of haplogroup F is probably centered on southern China and mainland Southeast Asia; the mainland subgroup F1a1a is today spread widely across both western and eastern parts of island Southeast Asia. Subgroup N9a6 may provide additional evidence of mainland Asian connections, again with southern China and mainland Southeast Asia.

On the Y chromosome, the Neolithic saw a massive influx of haplogroup O lineages into island Southeast Asia, and this haplogroup remains characteristic of western island Southeast Asians today. Subgroup O-P203, which in turn derives from O-M119, probably arrived in western island Southeast Asia from China, possibly via Taiwan, during this time period.

### Eastern Island Southeast Asia

A slightly different process seems to have occurred in the east, again likely starting in Taiwan and spreading through parts of the Philippines to the small island groups of eastern Indonesia. A range of mtDNA lineages dominated this expansion, including D5, F1a\*, F3b, F4, M7b3, M7c1c, and Y2, all of which show some evidence of connections to indigenous Taiwanese. D5, M7c1c, and Y2 are also common in Sumatra and its offshore islands, Nias and Mentawai, sometimes reaching higher frequencies in western Indonesia than further east. A strong Taiwanese connection is therefore observed across island Southeast Asia, caused either by a geographically broad expansion of individuals with indigenous Taiwanese ancestry, or perhaps by later gene flow from east to west. The absence of eastern haplogroups P and Q in western island Southeast Asia tends to favor the hypothesis of a widespread Taiwanese dispersal.

Undoubtedly the most widely studied lineage in eastern island Southeast Asia is B4a1a1, which incorporates the so-called “Polynesian motif.” This is one of the few lineages to link Taiwan and Polynesia, via a relatively low frequency of B4a1a1 in eastern island Southeast Asia (Melton et al. 1995). The full Polynesian motif likely developed in this region, although its antecedents can be traced to Taiwan, as well as further west to the Asian mainland (Trejaut et al. 2005).

On the Y chromosome, this expansion seems to be reflected most clearly in lineages O-M110 and O-P201\*. The lineage O-P201\*, which in turn derives from O-M122, also connects Taiwan and Polynesia, although O-P201\* occurs at lower frequency in Polynesians than mtDNA haplogroup B4a1a1 (Cox et al. 2007).

Today, Pleistocene and Neolithic lineages are found side by side in eastern Indonesia, which partially accounts for the extreme genetic diversity of this region (Karafet et al. 2010). In places, Pleistocene/Neolithic genetic diversity is statistically correlated with Papuan/Austronesian language diversity, even over extremely small geographical areas (Lansing et al. 2007). Indeed the Austronesian expansion appears to be an ongoing process in this region, although the actual cultural underpinnings of these population movements still remain poorly understood (Lansing et al. 2011).

### Oceania

Eastern island Southeast Asia provided a staging post for the settlement of the vast Pacific Ocean. Islands as far east as the Solomons were settled by c.40-55 kya, but colonization of the greater Pacific region was only accomplished within the last three to four thousand years. Although their ancestors passed through Melanesia, Polynesians trace relatively little of their biological ancestry to Melanesians (20%), but instead exhibit substantial Asian ancestry (80%). In terms of mtDNA and the Y chromosome, most of the Asian lineages that occur in Polynesians find their immediate antecedents in eastern Indonesia.

For mtDNA, this east–west connection is almost entirely represented by the Polynesian motif. This lineage, which – with the exception of Madagascar – is observed no

further west than central Indonesia, reaches frequencies of nearly 100 percent across Polynesia. In some island groups, the Melanesian lineages P and Q also occur, but at frequencies no higher than a few percent. The Y-chromosome story is slightly different. Lineage O-P201\* occurs at moderate frequency, illustrating clear connections back to the Asian mainland. However, it is the Melanesian lineage C-P33 that reaches highest frequencies in the Pacific, even attaining a frequency of 100 percent on the most remote Polynesian island, Easter Island/Rapanui (Cox et al. 2007). This lineage stems from C-M208, in turn derived from C-M38, which is characteristic of populations in eastern Indonesia.

The histories of mtDNA and the Y chromosome are therefore similar, but not identical. Both show evidence of founder effects and genetic drift, whereby genetic variants initially present at low frequencies reach high frequencies just by chance. In the case of mtDNA, the lineage that won this battle of chance was of Asian derivation. On the Y chromosome, a Melanesian lineage was instead picked up and carried to high frequency in Polynesians. Of course, when a large number of genomic markers are examined, it quickly becomes clear that Polynesians have predominantly Asian ancestry (about 80%). This provides an important warning that individual genetic markers like mtDNA and the Y chromosome can only tell us so much about our genetic prehistory; a range of markers from across the human genome will be necessary for us to reconstruct a fuller account of the human past.

### **Transition to the historic period**

As the Neolithic period gave way to the historic era, an increasingly large number of cultures began leaving their impact on the Indo-Pacific region. Relatively few traces of these societies are found in the mtDNA (i.e. female) record, mainly lineages – such as I and U7 – that reflect recent European contact. A more cosmopolitan history is recorded on the male Y chromosome, including traces of Chinese (O-M7, O-M134), Arab (J, L) and Indian (H, R, Q) contributions. Importantly, however, none of these lineages reach more than a few percent in frequency today.

### **Summary**

The Neolithic period was a defining era for island Southeast Asia and Oceania. It consolidated genetic patterns that were first laid down during the Pleistocene, but tempered them with new genetic variants emerging from mainland Asia. Following the region's initial settlement, the scale of these processes changed – from relatively local movements within island Southeast Asia during the Late Pleistocene, to regional mobility during the Neolithic, to immigration on a global scale during the historic era (Cox & Hammer 2010). However, on the borders of Melanesia, lineages that trace back to the Pleistocene are still in constant flux with lineages that first appeared in this region during the Neolithic. Many of these population dynamics continue into the present. In a very real sense, the history of Pacific peoples is still being played out.

SEE ALSO: 35 Southeast Asian islands and Oceania: Austronesian linguistic history; 36 Southeast Asian islands: archaeology; 41 Polynesia, East and South, including transpacific migration

## References

- Bais, W. J. & Verhoef, A. W. (1924) On the biochemical index of various races in the East Indian archipelago. *Journal of Immunology* 9, 383.
- Bellwood, P. (2005) *First Farmers: The Origins of Agricultural Societies*. Oxford: Blackwell.
- Cavalli-Sforza, L. L., Menozzi, P., & Piazza, A. (1994) *The History and Geography of Human Genes*. Princeton: Princeton University Press.
- Cox, M. P. (2008) The genetic environment of Melanesia: clines, clusters and contact. In V. T. Koven, (ed), *Population Genetics Research Progress*. New York: Nova Science, pp 45–83.
- Cox, M. P. & Hammer, M. F. (2010) A question of scale: Human migrations writ large and small. *BMC Biology* 8, 98.
- Cox, M. P., Karafet, T. M., Lansing, J. S. et al. (2010) Autosomal and X-linked single nucleotide polymorphisms reveal a steep Asian-Melanesian ancestry cline in eastern Indonesia and a sex bias in admixture rates. *Proceedings of the Royal Society B* 277, 1589–1596.
- Cox, M. P., Redd, A. J., Karafet, T. M. et al. (2007) A Polynesian motif on the Y chromosome: Population structure in Remote Oceania. *Human Biology* 79, 525–535.
- Friedlaender, J. S., Friedlaender, F. R., Reed, F. A., et al. (2008) The genetic structure of Pacific Islanders. *PLoS Genetics* 4, e19.
- Hertzberg, M., Mickleson, K. N. P., Serjeantson, S. W., et al. (1989) An Asian specific 9-bp deletion of mitochondrial DNA is frequently found in Polynesians. *American Journal of Human Genetics* 44, 504–510.
- Heydon, G. & Murphy, T. (1924) The biochemical index in the natives of the territory of New Guinea. *Medical Journal of Australia*, Sup. 1, 235–237.
- Hill, C., Soares, P., Mormina, M. et al. (2007) A mitochondrial stratigraphy for island southeast Asia. *American Journal of Human Genetics* 80, 29–43.
- Jordan, F. M., Gray, R. D., Greenhill, S. J., et al. (2009) Matrilineal residence is ancestral in Austronesian societies. *Proceedings of the Royal Society B* 276, 1957–1964.
- Karafet, T. M., Mendez, F. L., Meilerman, M. B. et al. (2008) New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Research* 18, 830–838.
- Karafet, T. M., Hallmark, B., Cox, M. P., et al. (2010) Major east-west division underlies Y chromosome stratification across Indonesia. *Molecular Biology and Evolution* 27, 1833–1844.
- Kayser, M. (2010) The human genetic history of Oceania: Near and Remote views of dispersal. *Current Biology* 20, R194–R201.
- Kayser, M., Lao, O., Saar, K. et al. (2008) Genome-wide analysis indicates more Asian than Melanesian ancestry of Polynesians. *American Journal of Human Genetics* 82, 194–198.
- Lansing, J. S., Cox, M.P., Downey, S. S. et al. (2007) Coevolution of languages and genes on the island of Sumba, eastern Indonesia. *Proceedings of the National Academy of Sciences USA* 104, 16022–16026.
- Lansing, J. S., Cox, M. P., de Vet, T. A., et al. (2011) An ongoing Austronesian expansion in Island Southeast Asia. *Anthropological Archaeology* 30, 262–272.
- Melton, T., Peterson, R., Redd, A. J., et al. (1995) Polynesian genetic affinities with Southeast Asian populations as identified by mtDNA analysis. *American Journal of Human Genetics* 57, 403–414.



- Mourant, A. E., Kopeć, A. C., & Domaniewska-Sobczak, K. (1976) *The Distribution of the Human Blood Groups and Other Polymorphisms*. Oxford: Oxford University Press.
- O'Connell, J. F. & Allen, J. (2004) Dating the colonization of Sahul (Pleistocene Australia–New Guinea): A review of recent research. *Journal of Archaeological Science* 31, 835–853.
- Tabbada, K. A., Trejaut, J., Loo, J.-H., et al. (2010) Philippine mitochondrial DNA diversity: A populated viaduct between Taiwan and Indonesia? *Molecular Biology and Evolution* 27, 21–31.
- Trejaut, J. A., Kivisild, T., Loo, J. H., et al. (2005) Traces of archaic mitochondrial lineages persist in Austronesian-speaking Formosan populations. *PLoS Biology* 3, e247.
- Van Oven, M. & Kayser, M. (2009) Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Human Mutation* 30, E386–394.