Review

A genetic history of migration, diversification, and admixture in Asia

Melinda A. Yang

Department of Biology, University of Richmond, Richmond, VA, USA;
E-Mail: myang@richmond.edu

Abstract

L.L. Cavalli-Sforza spearheaded early efforts to study the genetic history of humans, recognizing the importance of sampling diverse populations worldwide. He supported research on human evolutionary genetics in Asia, with research on human dispersal into Asia and genetic distances between present-day East Asians in the late 20th century. Since then, great strides have been made in understanding the genetic history of humans in Asia, through large-scale genomic sequencing of present-day humans and targeted sequencing of DNA from ancient humans. In this review, I survey the genetic prehistory of humans in Asia, based on research using sequence data from humans who lived in Asia as early as 45,000 years ago. Genetic studies comparing present-day Australasians and Asians show that they likely derived from a single dispersal out of Africa, rapidly differentiating into three main lineages: one that persists partially in South Asia, one that is primarily found today in Australasia, and one that is widely represented across Siberia, East Asia, and Southeast Asia. Studies of ancient DNA from human remains in Asia dating from as far back as 45,000 years has greatly increased our understanding of the population dynamics leading to the current Asian populations.

Based on "Jin L, Underhill PA, Doctor V, Davis RW, Shen P, Cavalli-Sforza LL, Oefner PJ. Distribution of haplotypes from a chromosome 21 region distinguishes multiple prehistoric human migrations. Proc Natl Acad Sci U S A. 1999;96(7):3796-3800".

Keywords: human migration; admixture; genetic structure; Australasians; Asians

1. Introduction

In the mid to late 20th century, L.L. Cavalli-Sforza helped to initiate the field of 'genetical demography' of humans [1], which sought to use genetic patterns to elucidate our past population history. He recognized that quantitative analysis of genetic data could add a unique dimension to human history and was a major advocate for developing a diverse set of human genetic samples for public research, as exemplified by the Human Genome Diversity Project [2]. His collaborations with researchers as varied as linguists and archaeologists proved that he was also a strong proponent for interdisciplinary research [3,4]. Cavalli-Sforza recognized the importance of studying the genetic history of humans in Asia to understand the dispersal of modern humans. One early example of this comes from a study conducted by him and his student Li Jin, among others, in 1999 [5]. They examined the distribution of a contiguous segment (i.e., a haplotype) of chromosome 21 across present-day populations worldwide. They found that after African populations, Oceanians showed the next highest level of haplotype diversity, and that the haplotype patterns of Oceanians were different from those found in East Asians. Based on these results, they concluded that there had been at least three distinct migrations, one to Oceania, one to Asia (and the Americas), and one to Europe. This early study demonstrated the importance of Asia in understanding human genetic history.

Asia is the largest inhabited continent in the world, home to nearly 60% of all humans with high ethnic diversity. While most of Cavalli-Sforza's studies focused on the genetic history of humans in Europe [3,4], he also contributed to and supported research on human genetics in Asia. He trained and worked with East Asian scholars to examine the association of genetic distances and surnames [6,7] in the 1980s-1990s, and he spoke highly of human genetic research in Asia, as demonstrated by a commentary he published on the earliest large-scale human genetic project from the region: the Chinese Human Genome Diversity Project [8]. Major contributions from Asian researchers in the growing field of human evolutionary genetics led to efforts such as the Human Genome Organization (HUGO) Pan-Asian SNP (single nucleotide polymorphism) Consortium [9,10], increasing our understanding of modern human genetic diversity in Asia. Such programs were the precursors of the broad genomic sequencing efforts seen today, such as the GenomeAsia 100K (GA100K) project, which has sequenced DNA from 1,739 individuals from 64 Asian countries [11]. In many ways, the GA100K Project is a direct product of the vision Cavalli-Sforza had of cataloging worldwide human genetic variations.

In the three decades since Cavalli-Sforza's initial call for the HGDP, the great improvements in sequencing techniques have allowed the creation of vast repositories of genome-wide data from diverse human populations, and thousands of present-day human genomes [9,11–16]

have been sequenced and analyzed. In addition, improved techniques for sequencing small amounts of DNA, preventing contamination, and compensating for post-mortem DNA damage [17–20] have increased access to DNA data from ancient humans, both archaic (Neanderthals and Denisovans) and modern (*H. sapiens*), allowing an unprecedented exploration of human genetic history [21–23]. The aim of this review is to understand the modern human dispersal in Asia over the past 45,000 years by examining genomic data from ancient and present-day modern humans. Our understanding of modern human expansion into Asia has greatly increased in scale and resolution because of the successful sequencing of ancient human DNA (Figure 1A). I will focus on findings that have increased our understanding of (1) the initial migration of modern humans into Asia, (2) the movement and interaction of humans in Upper Paleolithic Asia, and (3) the rapid dispersal of humans within Asia in the past 10,000 years.
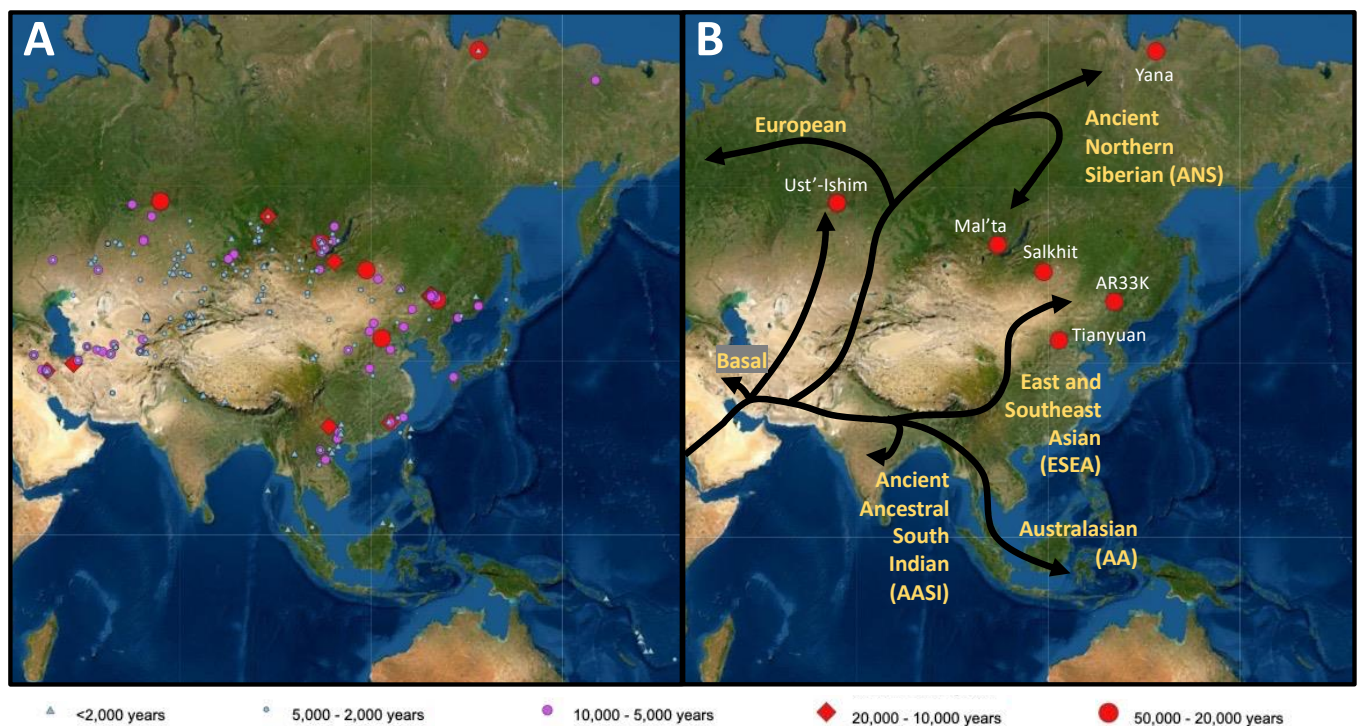

Figure 1 Maps indicating the location of ancient DNA samples from Asia and the patterns of genetic differentiation in Asia in the Early Upper Paleolithic. (A) Location of ancient individuals from Asia and Australasia who have been sequenced to date. Symbols refer to the age of individuals sequenced from that archaeological site, with the key found at the bottom. Red symbols date to 50,000–10,000 years ago (Upper Paleolithic), purple and blue symbols are younger than 10,000 years. (B) Differentiation after dispersal out of Africa in the Early Upper Paleolithic (45,000–20,000 years ago), with labeled lineages or ancestries in yellow next to the associated branch. The tree diagram shows divergence patterns and is not meant to depict migration routes from the branches or geographic origins of ancestral populations from the internal nodes. The branches predominantly associated with present-day Asian populations include the Ancient Ancestral South Indian (AASI) lineage, Australasian (AA) lineage, and East and Southeast Asian (ESEA) lineage. White labels refer to specific archaeological sites dating to the Early Upper Paleolithic.

2. Initial migration of modern humans into Asia and Australasia

An out-of-Africa model of human dispersal has been well supported by genetic studies, as discussed in [21,24,25], though later sequencing of archaic humans has demonstrated a small but notable contribution from archaic humans into early modern humans [26–28]. Later genetic studies also established that separation from African populations likely occurred 65,000–45,000 years ago [15,29]. Beginning at the turn of the 21st century, genetic studies began to address human dispersal out of Africa, such as the number of dispersals and the shape of human expansion in Eurasia and beyond.

Two models of dispersal have been proposed regarding migration into Asia and Australasia (the region consisting of Australia, New Zealand, and neighboring South Pacific islands). The single dispersal model describes a single migration out of Africa into Eurasia, with present-day Australasians deriving from an early offshoot of the Asian lineage. The multiple dispersal model posits several dispersals out of Africa by modern humans, with Australasians deriving from an earlier dispersal separate from the one contributing to mainland Asians and Europeans. Jin, Cavalli-Sforza, and others found high levels of genetic diversity in present-day populations from Oceania, and they used this finding to argue for a distinct out of Africa migration into Australasia [5], consistent with a multiple dispersal model. A phylogenetic analysis including 55,000 genome-wide single nucleotide polymorphisms (SNPs) across Asia by the HUGO Pan-Asian SNP Consortium found that East Asians and Australasians shared a closer genetic relationship to each other than to present-day Europeans [9], which they argued did not support a multiple dispersal model, but rather a single dispersal model.

In 2011, one of the first ancient DNA studies on a modern human was published, generating genome-wide data from a 100-year-old aboriginal Australian [30]. Comparison of the aboriginal Australian DNA to sequences from present-day Europeans (French) and Asians (Han Chinese) using a test of relative genetic similarity known as the **D-statistic** (bold text indicates methods or software described in Box 1) showed an excess of shared alleles between Europeans and Asians relative to Aboriginal Australians, supporting a multiple dispersal model. A complicating factor, however, was that Denisovans, early archaic hominins from Siberia [27,31,32] and the Tibetan Plateau [33,34], contributed up to 5% of their ancestry to populations living in Southeast Asia and Australasia island nations [27,32,35]. Because much smaller proportions of Denisovan ancestry were found in mainland Asian populations (0.05%-2%), the deep divergences observed genetically were potentially explained by archaic admixture rather than an earlier dispersal of modern humans [11,35–38].

Several studies have examined the relationship of Australasian populations to other modern humans after accounting for archaic ancestry from Neanderthals and Denisovans. After sequencing Denisovan DNA to 30-fold coverage, Meyer *et al*. estimated a maximum likelihood tree allowing admixture (**Treemix**, Box 1) and found that Papuans and East Asians were grouped together relative to Europeans [32]. Mallick *et al*. found that a well-fitting admixture graph (**qpGraph**, Box 1) grouped Papuans, Australians, and the Andamanese Onge with East Asians, with additional Denisovan admixture into Papuans and Australians [15]. Andamanese islanders such as the Onge do not show high Denisovan admixture, so Mondal *et al*. [39] compared whole-genome sequences from ten Andamanese individuals to other present-day populations. Using both Treemix and D-statistic analyses, they found that the Andamanese shared a close genetic relationship to mainland Indians, East Asians, and other Australasians [39]. Malaspinas *et al*. [40] examined the likelihood of a single dispersal model versus a multiple dispersal model by comparing the observed joint frequency spectrum to the expected joint frequency spectrum using **fastsimcoal2** (Box 1). They found that accounting for Denisovan admixture led to better support for the single dispersal model, while excluding Denisovan admixture led to better support for the multiple dispersal model [40]. Collectively, these studies showed that after accounting for archaic admixture from Neanderthals and Denisovans, Australasians consistently grouped closely with mainland Asian populations, supporting a single dispersal model.

Box 1 Commonly used tools and software for analyzing genetic relationships among ancient and present-day humans.

ADMIXTURE [41]—A statistical tool that assumes individual genomes can be modeled as a mixture of shared components and estimates the proportion of each genome that can be associated with a component. The number of components, *K*, is provided by the user, and this tool is usually used over multiple *K*.

Ancient DNA capture [42,43]—A sequencing technique that enriches for a panel of single nucleotide polymorphisms (SNPs) rather than sequencing all DNA found in a sample. This allows targeting of DNA specific to the species of interest, in this case human, rather than sequencing of environmental DNA. With low endogenous DNA in ancient samples, this lab method is popular as it allows efficient retrieval of DNA from the targeted species. Several SNP panels have been developed for humans, and studies using ancient DNA capture techniques referenced in this review use one of the following SNP panels developed in two studies [42,43], where the smaller ones are subsets of the larger ones: a 3.8M SNP panel, a 2.2M SNP panel, and a 1.2M SNP panel (most common).

D- or f4-statistics [44,45]—A four population test of relative genetic similarity, typically of the form *D(A, B; C, Outgroup)*, which measures the number of shared alleles between B and C relative to A and C. A higher number of shared alleles indicates higher genetic similarity between those two populations relative to the third.

diCal2 [46]—A population genetic inference tool that estimates parameters such as population size changes over time and split time estimates using a conditional sampling distribution. This method allows inclusion of fully parametric demographic models, allowing parameters related to migration.

f4-ratio statistic [44,45]—A tool used to estimate mixture proportions from an admixture event using f4-statistics. This method assumes availability of a closely related sample to one source population that did not also mix with the target or other source populations.

fastsimcoal2 [47]—A population genetic inference tool that uses simulation and the site frequency spectrum for sampled populations and a composite-likelihood approach to infer parameters for a given model.

fineSTRUCTURE [48]—A statistical tool that assigns blocks of contiguous sequence, or haplotypes, with high similarity amongst a set of individuals to the same component. This method, referred to as 'chromosome painting', allows examination of similarity (and thus, shared descent) at a finer scale, using the physical location of each base pair.

momi2 [49]—A population genetic inference tool that compares the computed site frequency spectrum for a given model to the observed site frequency spectrum for a set of sampled populations to infer parameters for a given model.

Multiple sequentially Markovian coalescent (MSMC) [50]—A population genetic inference tool relying on a small set of genomes to estimate parameters such as population sizes over time and coalescence rates within and across different subgroups of individuals to infer split times.

Outgroup f3-statistics [45,51]—A three population test of the form f3(Outgroup; A, B), where a higher f3 value indicates higher genetic similarity between populations A and B.

Principal component analysis (PCA) [52]—A statistical tool to simplify data by dimensionality reduction, where data are regrouped into a set of new variables based on the amount of variance explained by the data (called principal components). Cavalli-Sforza and others first introduced the use of PCA to human population genetic studies, by making the connection that the first few principal components are often informative about evolutionary genetic relationships [53]. In ancient DNA studies, present-day populations are often used for a PCA, upon which ancient individuals are then projected [54].

qpAdm [42,45]—A tool that allows estimate of mixture proportions in a specified target from *n* specified sources. It measures mixture

proportions by comparing the target and source to a set of reference populations that are differentially related to the specified sources using f4-statistics, where the target's f4-statistics are assumed to be a linear combination of the sources' f4-statistics if the target is truly a mixture of ancestries related to the sources.

qpGraph [45]—Estimation of an admixture graph allowing branching and admixture relationships between a set of specified populations. For a given admixture graph, the expected f2-, f3-, and f4-values are compared to the observed f2-, f3-, and f4-values for the specified populations to determine the feasibility of the model.

Treemix [55]—A tool that estimates a maximum likelihood tree using allele frequency correlations, allowing *m* migration events specified by the user.

While a closer relationship between Australasians and Asians than Europeans was broadly supported, some scholars still argued for a small but notable contribution to Australasians from an earlier dispersal of modern humans out of Africa. Pagani *et al*. used **MSMC split time estimates** (Box 1) to argue that Papuans separated from African populations earlier than mainland Eurasian populations [56]. Assuming 30 years per generation and a mutation rate of 1.25×10^{-8} per generation per site, they estimated a 90,000-year split time between Papuans and Yorubans and a 75,000-year split time between mainland Eurasians and Yorubans. Using **fineSTRUCTURE** (Box 1), they examined haplotypes in Papuans associated with a deeper divergence and found that 2% of the haplotypes in Papuan genomes could not be explained by Denisovan admixture or shared origins with mainland Eurasians, and thus were best explained as originating from an earlier dispersal of modern humans out of Africa [56]. Though an earlier dispersal may be partially represented in the genomes of Australasians, the main pattern observed in their genomes indicates a shared evolutionary history with populations widespread today in much of the eastern regions of Asia.

3. Defining Asian Ancestries

In the rest of this review, I examine genetic findings related to past humans from Asia, often using the term 'ancestry' to communicate key interpretations of genetic patterns. Ancestry is a useful concept for conveying relationships in evolutionary genetics, but imprecise application of the term can result in racial categorizations that unintentionally bolster misleading claims of the biological reality of race [57]. Mathieson and Scally [58] recently published a nuanced exploration of how the term was used and misconstrued in human evolutionary genetics, a portion of which is summarized here to clearly define how the term ancestry is used in this review.

As described by Mathieson and Scally [58], a technical definition of genetic ancestry refers to the set of paths across an individual's ancestors through which a DNA segment was inherited. For a population, the set of genetic ancestries across all individuals is highly complex, so for practical reasons, researchers focus on summarizing the general demographic relationships observed between populations. This concept of 'population ancestry' then assumes that a population can be represented as a mixture of different source populations, denoted here as *ancestries*, dependent on the pattern of variation across the genome.

My objective is to articulate key human ancestries of Asia that have been described in the literature to date, inferred from findings of high genetic similarity among ancient and present-day populations. When a described ancestry has only been associated with current populations or is only known through partial representation in an ancient individual or present-day population, I use the term *lineage* rather than ancestry to indicate descent from a hypothetical ancestor that has yet to be sampled. As Mathieson and Scally note, all source populations, or ancestries, are constructs that (1) may be only distantly related to the associated sampled individuals; (2) may not actually be represented in the available samples ('ghost populations'); and (3) create discrete sources for a population when discrete categories are not actually applicable [58]. The ancestries and lineages referred to throughout this review (first appearance **_bold italic_**, with description in Box 2) are theoretical constructs that have been associated with samples from different past or present populations, and they cannot be described as physical ancestors of any present-day individuals.

Box 2 Major lineages and ancestries described in the text.

Lineages found in Asia and Australasia that contributed to ancestries that would shape present-day Asians and Australasians
Ancient Ancestral South Indian (AASI) lineage—this lineage refers to an ancestral population that primarily contributed to humans living in South Asia, particularly southern India. Partially represented in 5,000–1,500-year-old individuals from in or near the Indus Periphery and present-day Indians [59,60].
Australasian (AA) lineage—this lineage refers to the ancestral population that primarily contributed to human populations in Australasia, or the region consisting of Australia, New Zealand, and neighboring islands in the South Pacific Ocean. Represented primarily by present-day Australasians, e.g. Papuans and Aboriginal Australians.
East and Southeast Asian (ESEA) lineage—this lineage refers to an ancestral population that primarily contributed to humans living in mainland East and Southeast Asia. Represented primarily by present-day East and Southeast Asians, e.g. Han and Kinh.

Ancestries in the ESEA lineage

Amur ancestry—ancestry associated with populations in the Amur River region, Mongolia, and Siberia, with the oldest individual sampled to date represented by a 14,000-year-old individual from the Amur River region, i.e. Amur14K [61]. Populations associated with this ancestry likely contributed to the ancestors of Native Americans and populations associated with Paleosiberian ancestry.

Fujian ancestry—ancestry associated with populations in the Fujian region of coastal, southern China from 12,000–4,000 years ago. The oldest individual sampled to date is an individual from Fujian, i.e. Qihe3 [62]. Populations associated with this ancestry contributed to Austronesian speakers today.

Guangxi ancestry—ancestry associated with a 10,500-year-old individual from Guangxi, i.e. Longlin [62]. Populations associated with this ancestry are partially observed in 8,000–6,000-year-old hunter-gatherers from Guangxi and are not observed in historical Guangxi individuals and present-day East and Southeast Asians.

Hòabìnhian ancestry—ancestry on the ESEA lineage associated with 8,000–4,000-year-old hunter-gatherers [63] associated with the Hòabìnhian culture in Laos and Malaysia. This ancestry is deeply diverged from the common ancestor of present-day East and Southeast Asians and Tianyuan ancestry.

Jōmon ancestry—ancestry associated with 8,000–3,000-year-old individuals in the Japanese archipelago. The oldest individual sampled to date is Higashimyo, from Kyushu, Japan [64]. Populations associated with this ancestry contributed partially to present-day Japanese populations.

Tianyuan ancestry—ancestry on the ESEA lineage associated with Upper Paleolithic individuals dating to 40,000–33,000 years ago in northern China and Mongolia, i.e. Tianyuan, Salkhit, and AR33K [61,65,66]. This ancestry is deeply diverged from the common ancestor of present-day East and Southeast Asians and Tianyuan ancestry.

Tibetan ancestry—ancestry associated with 3,000–1,000-year-old individuals in the Himalayan region of the Tibetan Plateau [67]. Populations associated with this ancestry contributed to present-day Tibetan and Sherpa populations.

Yellow River ancestry—ancestry associated with populations in the Yellow River region, with the oldest individual sampled to date represented by a 9,500-year-old individual from the lower reaches of the Yellow River in Shandong, i.e. Bianbian [68]. Populations associated with this ancestry greatly impacted most present-day East and Southeast Asians.

Ancestries not deriving completely from the AA, AASI, or ESEA lineages

Ancient Northern Siberian (ANS) ancestry—ancestry associated with 33,000-year-old individuals from near the Yana River in northern Siberia [69]. 24,000 and 17,000-year-old individuals from the Lake Baikal region of Siberia [51], associated with Ancient North Eurasian ancestry, derive from a lineage associated with ANS ancestry. This ancestry is more closely related to ancestry found in present-day Europeans rather than ancestry found in present-day East and Southeast Asians.

Basal ancestries—ancestries associated with populations that diverged very early in the dispersal out of Africa. At least three distinct ancestries have been reported, including one related to a 45,000-year-old individual from Ust'-Ishim Cave in western Siberia [70], and two related 12,000–7,000-year-old individuals from the Levant and Iran, who are some of the earliest agriculturalists sampled to date [71]. Populations associated with Basal Iranian ancestry contributed to populations in Central and South Asia.

Indus Periphery (IP) ancestry—ancestry associated with 5,000–4,000-year-old individuals from near the Indus Valley [59,60]. This ancestry is a mixture of ancestry related to the AASI lineage and Basal Iranian ancestry, and populations associated with this ancestry contributed to populations in South Asia.

Native American ancestry—ancestry associated with present-day Native Americans. This ancestry is a mixture of ancestries associated with Amur ancestry and ANS ancestry, and it is closely related to Paleosiberian ancestry. The oldest sequenced individual associated with this ancestry is an 11,500-year-old individual from Upward Sun River, Alaska [72].

Paleosiberian ancestry—ancestry associated with 14,000–10,000-year-old individuals from the Lake Baikal region of Siberia (Ust-Kyakhta-3) and Far East Siberia (Kolyma1) [69,73]. This ancestry is a mixture of ancestry related to ANS and Amur ancestries, and it is closely related to ancestry associated with Native Americans.

Steppe ancestry—ancestry associated with 5,000–3,500-year-old individuals from the central and western steppe, related to the Yamnaya and Afanasievo cultures [42,74]. This ancestry is a mixture of ancestries related to populations associated with ANS ancestry, basal ancestry from the Levant and Iran, and ancestry related to hunter-gatherers from the Caucasus regions of eastern Europe.

3.1 Early lineages of Asia

Though present-day Asians and Australasians are more closely related to each other than to present-day Europeans, genetic comparisons highlight deep separations between mainland East and Southeast

Asians, island Southeast Asians, and Australasians. Philippine Negritos, Papuans, and aboriginal Australians share a close genetic relationship to each other relative to mainland East and Southeast Asians [11,15,59,75] and are collectively referred to here as belonging to an ***Australasian (AA) lineage*** (Box 2). In contrast, mainland East and Southeast Asians and other Pacific islanders (e.g., Austronesian speakers) are closely related to each other [9,15,16] and here denoted as belonging to an ***East and Southeast Asian (ESEA) lineage*** (Box 2). In South Asia, present-day populations are highly admixed, but recent sequencing of ancient DNA indicated the presence of a deeply diverged Asian ancestry only distantly related to populations associated with the AA or ESEA lineages [59]. This distinct South Asian ancestry, denoted as the ***Ancient Ancestral South Indian (AASI) lineage*** (Box 2), was only found in a small percentage of ancient and present-day South Asians. Present-day Onge from the Andamanese Islands are the best reference population to date, but Narasimhan *et al*. used qpGraph to show that the divergence between the AASI lineage and the ancestry found in present-day Onge was very deep [59]. Ancestry associated with the AASI lineage was found at low levels in almost all present-day Indian populations, particularly southern Indians [59,60], emphasizing the high impact of the AASI lineage in South Asia.

The AA, ESEA, and AASI lineages showed a closer genetic relationship to each other than lineages observed in present-day Europeans [59] and together represent the main branches of Asian-related ancestry sampled to date (Figure 1B). The relationship of the AA, ESEA, and AASI lineages to each other is not well resolved, in part due to high levels of admixture in present-day Asian and Australasian populations, either with archaic humans [26,27,32,37], populations carrying other non-African ancestries [44,51], or with each other [9]. Recently, Hajdinjak *et al*. used **ancient DNA capture targeting a 3.8M SNP panel** (Box 1) to sequence DNA for three 45,000-year-old individuals from Bacho Kiro Cave in Bulgaria [76]. They found that the Bacho Kiro individuals were genetically more similar to present-day Asians than to present-day Europeans in an **outgroup f3-analysis** (Box 1), extending the range in which ancestry associated with present-day Asian populations has been found as far west as Bulgaria.

4. Differentiation and admixture in Asia 45,000–10,000 years ago

To date, no Upper Paleolithic individuals sampled have been associated with the AASI and AA lineages, though **ADMIXTURE analyses** (Box 1) on present-day populations associated with these lineages in South Asia and Australasia indicate the presence of complex structure and admixture [11]. Genetic data from ancient individuals have been informative about the ESEA lineage. All humans sampled to date from

the Upper Paleolithic (45,000–10,000 years ago) in Asia were excavated in China, Mongolia, and Russia, in part because preservation of genetic material in human specimens is more likely in colder and drier regions [77]. In the next sections, I review how ancient DNA has helped to clarify genetic structure and admixture in populations associated with the ESEA lineage in the Upper Paleolithic.

4.1 Early Upper Paleolithic: 45,000 to 20,000 years ago

Three individuals from northern China [61,66,78] and northeastern Mongolia [65] dating to 40,000 to 33,000 years ago (Figure 1B) were sequenced using ancient DNA capture targeting a 1.2M (AR33K) or 2.2M (Tianyuan, Salkhit) SNP panel. Comparison with present-day populations using outgroup f3- and D-statistics showed that they were genetically most similar to present-day East Asians and Southeast Asians, thus showing that Tianyuan, Salkhit and AR33K are associated with the ESEA lineage, rather than the AA or AASI lineages [61,65,66,78]. In a test of genetic continuity, Tianyuan was shown to derive from a population distinct from the one that contributed to present-day East and Southeast Asians, indicating differentiation of the ESEA lineage in the Early Upper Paleolithic [66]. D-statistic analyses showed that AR33K and Salkhit were more closely related to Tianyuan than to present-day East and Southeast Asians [61,65]. Thus, by 40,000–33,000 years ago, one or more populations in northern China and Mongolia were genetically differentiated from the source population that contributed to present-day East and Southeast Asians. I denote the shared ancestry between Tianyuan, AR33K, and Salkhit individuals as ***Tianyuan ancestry*** (Box 2), which diverged prior to the shared common ancestor of present-day East and Southeast Asians.

No DNA has been retrieved from 40,000–20,000-year-old humans in southern regions of Asia, but sampling of 8,000–4,000-year-old hunter-gatherers [63] associated with the Hòabìnhian culture in Laos and Malaysia reveals another ancestry that must have existed in the Early Upper Paleolithic. Ancient DNA capture on the 1.2M SNP panel and assessment of relative genetic similarity using D-statistics showed that these Southeast Asian hunter-gatherers were most closely related to Tianyuan and present-day East and Southeast Asians [63]. However, D-statistic comparisons also showed that they were as genetically different from Tianyuan as they were from present-day East and Southeast Asians [63,68] and are denoted here as associated with ***Hòabìnhian ancestry*** (Box 2). Together, the genetic patterns described above show that the ESEA lineage differentiated into at least three distinct ancestries: Tianyuan ancestry which can be found 40,000-33,000 years ago in northern East Asia, ancestry found today across present-day populations of East Asia, Southeast Asia, and Siberia, but whose origins are unknown, and Hòabìnhian ancestry found 8,000-4,000 years ago in Southeast Asia, but whose origins in the Upper Paleolithic are unknown.

DNA sampling of ancient humans in western Siberia, the Near East, and Central Asia showed that populations existed in the past in the western regions of Asia that are not directly represented today. For instance, Fu *et al*. [70] retrieved and sequenced 1.86 gigabases of the autosomal genome from a femur found at Ust'-Ishim, a settlement in western Siberia (Figure 1B). Using D-statistics, they showed that the Ust'-Ishim individual was similarly related to Upper Paleolithic European hunter-gatherers and present-day East Asians [66,70,79]. After targeting 1.2M SNPs [42,43] in 12,000–7,000-year-old individuals from the Levant and Iran, Lazaridis *et al*. found that Ust'-Ishim and present-day East Asians shared more alleles with each other than with ancient individuals from the Levant and Iran [71], which indicates that they diverged basally from other non-Africans. In a principal component analysis (**PCA**, Box 1), ancient individuals from the Levant and Iran clustered at opposite ends of a Near Eastern cline, indicating high genetic differentiation between these two populations [59,71]. Ust'-Ishim and the 12,000–7,000-year-old individuals from the Levant and Iran represent deeply diverged non-African ancestries (denoted here as ***Basal ancestries***, Box 2) that are partially or not represented in present-day populations.

In northeastern Siberia, Upper Paleolithic populations were not associated with the ESEA, AA, or AASI lineages. Two individuals (Yana 1 and 2) were sampled in northeastern Siberia at the Yana Rhinoceros Horn site off the Yana River dating to 31,000 years ago [69], and another three (Mal'ta 1 and Afontova Gora 2/3), dating to 24,000–17,000 years ago, were sampled in south central Siberia [51] (Figure 1B). Comparison to ancient and present-day populations in Europe and Asia using **f4-statistics** (non-normalized version of the D-statistic, Box 1) showed that these individuals were genetically more similar to European hunter-gatherers than to Tianyuan or present-day East and Southeast Asians [51,69], which suggests that Upper Paleolithic Siberians were split at an early point from a population that contributed to ancient European hunter-gatherers [79]. The ancestry represented by the Yana individuals is denoted as ***Ancient Northern Siberian (ANS) ancestry*** (Box 2, Figure 1B), based on Sikora *et al*. [69]. Mal'ta1 and Afontova Gora 2/3 [51], often referred to as representing Ancestral North Eurasian ancestry, is closely related to ANS ancestry found in Siberia 31,000 years ago and is suggested to have originated from ANS ancestry [69]. The persistence of ANS ancestry in Siberia from 31,000-17,000 years ago shows that ANS ancestry played an essential role in the early history of human migration across northern Asia. These patterns in Siberia and East and Southeast Asia show that 40,000-20,000 years ago in the Early Upper Paleolithic, eastern Eurasia was replete with populations that were genetically distinct from one another. Both ANS and ESEA ancestries were instrumental in shaping human genetic history.

Interaction between populations associated with Tianyuan ancestry and ANS ancestry played a key role in shaping the distribution of humans in

East Asia and Siberia. For instance, a symmetry test using D-statistics showed that Salkhit shares more genetic similarity with the ANS-related Yana individuals than either Tianyuan or AR33K [61,65], which suggests that populations in Mongolia associated with Tianyuan ancestry likely interacted more with populations associated with ANS ancestry than those in the Amur River region and the North China Plain. D-statistic analyses additionally show that Salkhit and Tianyuan share more alleles with a 35,000-year-old European hunter-gatherer from Goyet Cave, Belgium (GoyetQ116-1) than with present-day East and Southeast Asians [61,65,66]. Yang *et al*. [66] could not fully explain the population dynamics leading to a genetic connection between western European and northern Asian individuals, but a similar affinity to GoyetQ116-1 recently observed in the 45,000-year-old Bacho Kiro individuals from Bulgaria using qpGraph showed that the extended connections were potentially facilitated by Asian-related populations that once lived in Europe [76]. The interconnections among Early Upper Paleolithic European and Asian populations show that these Early Upper Paleolithic hunter-gatherers were not totally isolated--admixture and migration played a fundamental role in their ancestral makeup.

4.2 Late Upper Paleolithic: 20,000 to 10,000 years ago

Remains of humans dating from 20,000 to 10,000 years ago have begun to provide insights into the ancestries that shaped present-day Siberians, East Asians, and Southeast Asians. A 1.2M SNP panel was used to analyze ancient DNA from a 19,000-year-old individual from the Amur River region (AR19K) and a 12,000-year-old individual from the Qihe Cave in Fujian, China (Qihe3), as well as younger coastal northern and southern Chinese populations dating from 9,500 to 4,000 years ago. They all share greater genetic similarity to present-day East and Southeast Asians than Tianyuan or the 8,000–4,000-year-old Hòabìnhians [68,80,81]. Across several studies, phylogenetic and D-statistic analyses were used to show that northern populations from the Yellow River region to the Amur River region shared a closer genetic relationship to each other than to southern populations from coastal southern East Asia, or the Fujian region [68,80,81].

As mentioned earlier, ancient individuals from southern central Siberia dating to 24,000–17,000 years ago (Mal'ta 1 and Afontova Gora 2/3) [51] are descended from a population associated with ANS ancestry like that found in the Yana individuals of northeastern Siberia [69]. Raghavan *et al*. [51] used outgroup f3-statistics to show that Mal'ta1 was most closely related to present-day Native Americans. This pattern led to a model in which the dispersal of humans to the Americas originated in a population in Siberia derived from a mixture of two distinct ancestries, one associated with East Asians today and another associated with Mal'ta1. Recent studies of DNA from ancient Siberians [69,73] provided direct evidence of admixed populations in Siberia that are closely related to the Native American lineage. Using qpGraph,

Yu *et al*. showed that a 14,000-year-old human (UKY) from the Ust-Kyakhta-3 site south of Lake Baikal [73] and a 10,000-year-old human (Kolyma1) [69] from near the Kolyma River in Far East Siberia both shared a close genetic relationship to Native Americans. Using simulation-based demographic modeling with fastsimcoal2, Sikora *et al*. found that Kolyma1 can be described as a mixture of ancestry related to present-day East Asians (e.g., Han) and ANS ancestry. This admixed ancestry found across the Siberian landscape [69] is denoted by Sikora *et al*. as ***Paleosiberian ancestry*** (Box 2, Figure 2).

A question persists regarding which populations in East Asia mixed with populations associated with ANS ancestry, resulting in the signature of genetic admixture found in Paleosiberians and Native Americans. Using shotgun sequencing, Sikora *et al*. [69] generated whole genomes from six 8,000-year-old individuals from Devil's Gate Cave in the Primorye region of Siberia (DevilsCave_N) who were previously partially sequenced [82]. They showed that these individuals grouped with present-day East Asians, and represented the best Asian source population for Kolyma1 [69]. In a later study, Mao *et al*. [61] targeted the 1.2M SNP panel for a 14,000-year-old individual from the Amur River region (AR14K) and found using a **qpAdm mixture analysis** (Box 1] and qpGraph modeling analysis that AR14K was a better source population for Asian ancestry in Kolyma1 than the DevilsCave_N individuals. Using f4-statistics, both DevilsCave_N and AR14K share a close genetic relationship to each other and group phylogenetically with other ancient northern East Asian individuals rather than ancient southern East Asian individuals [61,68]. Mao *et al*. argue that populations in the Amur River region likely played an important role in interactions with populations associated with ANS ancestry [61]. Denoted here as ***Amur ancestry*** (Box 2, Figure 2), the ancestry associated with AR14K and DevilsCave_N found in the Amur River and Primorye region greatly affected the human landscape in Siberia and the Americas.

Sikora *et al*. also estimated split times among Siberian and East Asian populations using fastsimcoal2. With a mutation rate of 1.25 x 10^{-8} per generation per site and a generation time of 29 years, they estimated that the Asian source that contributed to Kolyma1 diverged around 24,000 years ago [69]. If that Asian source was Amur ancestry, then 24,000 years ago may also indicate the split time separating Amur ancestry from other Asian ancestries. Moreno-Mayar sequenced a high coverage genome of an 11,500-year-old individual (USR1) at Upward Sun River, Alaska [72] who shares a close relationship with present-day Native Americans. They also estimated split times by modeling the evolutionary relationship between USR1 and present-day East Asians, Siberians, and Native Americans with **diCal2** and **momi2** (Box 1). Using the same mutation rate and generation time as Sikora *et al*., they estimated that Native Americans and present-day Siberians separated 36,000–25,000 years ago, and that USR1 separated from other Native

Americans around 20,000 years ago. They used these estimates to suggest that admixture related to ANS ancestry likely occurred between 25,000 and 20,000 years ago [72]. These results are consistent with qpGraph analyses including Paleosiberians, Native Americans, and East Asians [69,73], where the Asian source for Native Americans separated earlier than the Asian source for Paleosiberians.
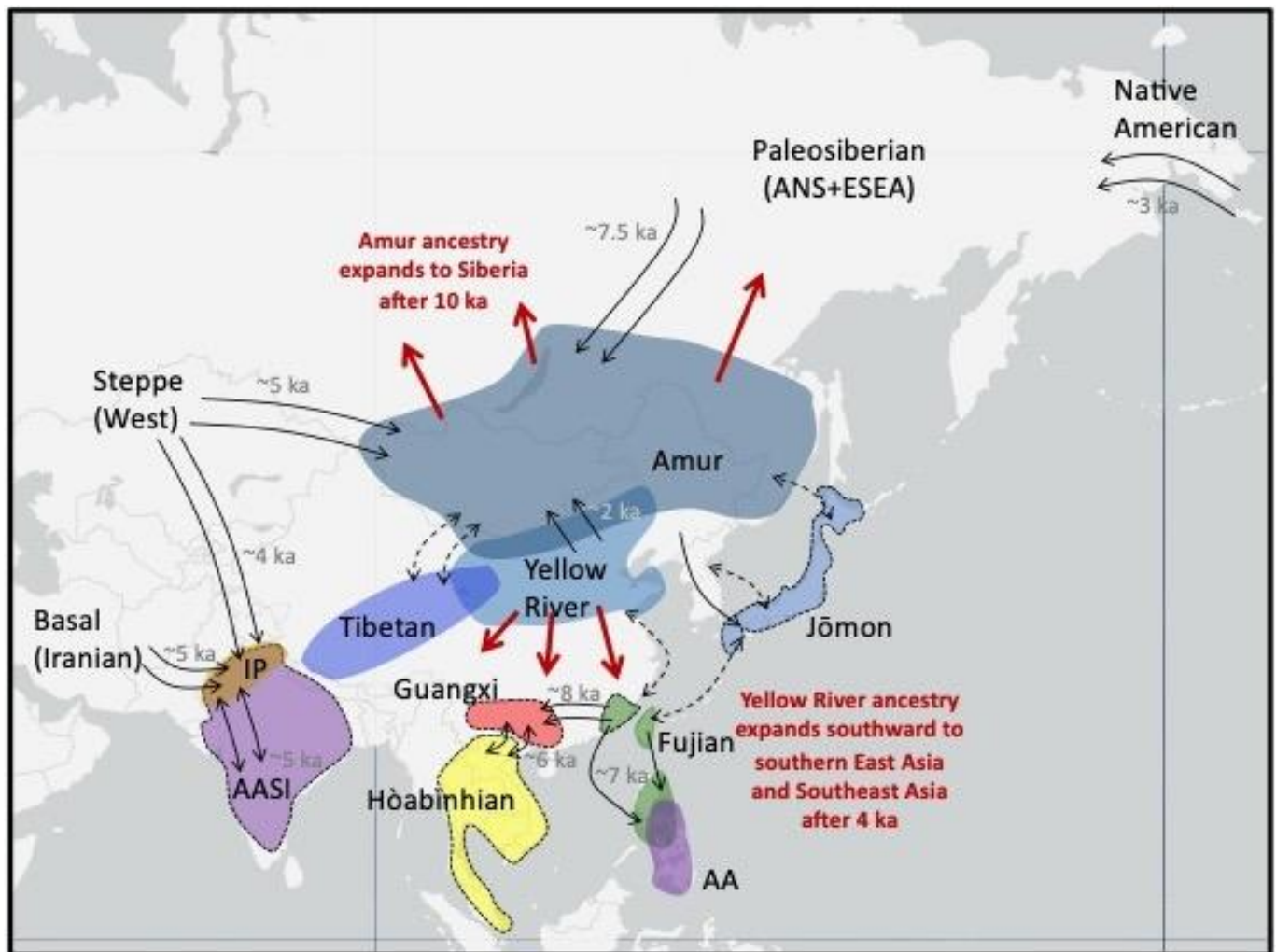

Figure 2 Schematic depicting major ancestries in Asia from the last 15,000 years. Black labels indicate ancestries and lineages described in the text and in Box 2. Highlighted regions show where ancient individuals associated with the labeled ancestry have been sampled. Excepting Indus Periphery (IP) ancestry, ancestries not associated with a highlighted region do not derive complete from the AA, AASI, or ESEA lineages. Purple regions indicate non-ESEA lineages (AASI and AA lineages). All other labeled ancestries are subsets of ESEA lineage, where Hòabìnhian ancestry (yellow region) is deeply diverged from all other labeled ancestries. A dashed border indicates ancestries that are only partially (AASI, Jōmon, Fujian in mainland Asia) or not (Guangxi, Hòabìnhian) represented in present-day populations. Black arrows with a date in gray (ka = thousand years ago) indicate documented gene flow related to those ancestries, while black arrows with a dashed line indicate that connections were observed but the underlying demographic history is not well-known. Red arrows and associated red text describe expansion of ancestry associated with northern East Asians associated with Amur or Yellow River ancestry.

Further south, the 1.2M SNP panel was used to analyze ancient DNA from a 10,500-year-old individual (Longlin) from Longlin Cave in Guangxi, a region in southern China [62] between the Fujian region and Southeast Asia. Both phylogenetic analyses using Treemix and tests of relative genetic similarity using f4-statistics showed that Longlin was more closely related to 9,000–4,000-year-old East Asians from northern and southern coastal China than with Tianyuan or Hòabìnhians [62]. However, these same analyses showed that Longlin was an outgroup of the 9,000–4,000-year-old northern and southern East Asians. The ancestry associated with Longlin, denoted as *Guangxi ancestry* (Box 2, Figure 2), was not observed in either historical samples from Guangxi or present-day East and Southeast Asians [62], which suggests that Guangxi ancestry did not persist to the present-day.

Inferences using ancient DNA from 8,000–3,000-year-old humans in Japan indicated that a population in the Japanese archipelago was also differentiating during the end of the Upper Paleolithic. These ancient individuals were associated with the Jōmon period of Japan, the first cultural period found in the archaeological record dating from 16,000 to 2,800 years ago. They were originally from the northern reaches of Hokkaido [83], the central and northern regions of Honshu [63,81,84,85], and the southern island of Kyushu [64]. These Jōmon individuals consistently cluster together in a PCA and show high genetic similarity to each other distinct from that found in other Asian populations; their associated ancestry is denoted here as *Jōmon ancestry* (Box 2, Figure 2). Like Longlin, they are more closely related to 9,000–4,000-year-old East Asians from coastal China than to Tianyuan or Hòabìnhians, but are an outgroup of these northern and southern East Asians. Some have argued for the presence of excess connections to Hòabìnhians by fitting the data to a graph that includes admixture with a Hòabìnhian-related population and finding different f4 patterns for Hòabìnhians compared to younger Southeast Asians in comparisons to a Jōmon individual [63]; however, alternative admixture graphs and f4-statistic comparisons do not show evidence for this connection [68,85,86].

A lack of Upper Paleolithic samples representing South Asia or Southeast Asia makes it difficult to assess the ancestral patterns within these regions. While sampling of ancient humans in the Upper Paleolithic is still sparse, the findings to date reveal a diverse set of ancestries across Asia with high genetic structure across East Asia and interactions between distinct populations in northern Asia.

5. Rapid Human Dispersals Within Asia in the Last 10,000 Years

Ancient DNA capture on the 1.2M SNP panel has been performed on multiple individuals across Asia dating to the last 10,000 years, providing insight on the population dynamics leading to the genetic composition

observed today. In the following section, I review demographic patterns determined from analysis of humans dating to the last 10,000 years.

5.1 Population dynamics in northern East Asia, Siberia, and the Tibetan Plateau

A PCA including present-day populations of East Asian ancestry in northern East Asia, Siberia, and the Tibetan Plateau shows two major clines [61,62,67]. One cline includes present-day populations in the Amur River region (e.g., the Ulchi) and inland across the eastern steppe (e.g., the Xibo), whereas another cline represents populations on or near the Tibetan Plateau. Analysis of ancient individuals from these regions revealed a more complex history than observed with present-day East Asians.

In the Late Upper Paleolithic of the Amur River region and Siberia, ancient individuals sampled were associated with Amur ancestry (e.g., AR14K) or an admixed Paleosiberian ancestry (e.g., UKY and Kolyma1). In the Amur River region, 14,000–3,000-year-old individuals shared greater genetic similarity with each other, and in a PCA, they and the 8,000–7,000-year-old individuals at Devil's Gate Cave and Boisman-2 cemetery from the coastal Primorye region adjacent to the Amur River region [69,81,82] clustered together and with present-day populations from the region, particularly the Tungusic and Mongolic speakers [61,69,80–82]. Thus, shared ancestry has persisted within this region from at least 14,000 years ago to the present-day.

Further west in central Siberia, DNA sampling of ancient humans from 8,500 years ago to historical times from the Baikal region, Mongolia, and Inner Mongolia has consistently revealed high genetic diversity and interaction in central Siberian populations. In mixture models using qpAdm [61,69,74,81,87], most 8,500–3,000-year-old populations from central Siberia show a mixture of Amur ancestry (e.g., DevilsCave_N) and ANS ancestry (e.g., from Mal'ta1 or Kolyma1). Several ancient individuals dating from 8,000-5,000 years ago in eastern Mongolia and Inner Mongolia [68,87] and Early Iron Age individuals dating to 3,000 years ago in Central Mongolia (i.e., Slab Grave) [87] show little to no evidence of admixture with populations carrying ANS ancestry, which suggests that Mongolia was an interaction zone for diverse populations. From 5,000 to 3,500 years ago, central and western steppe populations associated with the Yamnaya and Afanasievo cultures, followed by the Sintashta culture, migrated to Mongolia and contributed to Mongolian and southern central Siberian populations (Figure 2). They brought an admixed ancestry related to populations associated with ANS ancestry, basal ancestry from the Levant and Iran, and ancestry related to hunter-gatherers from the Caucasus region of eastern Europe [42,59,81,87], denoted here as **_Steppe ancestry_** (Box 2). Beginning around 2,000 years ago in Mongolia, however, historical individuals were more closely

related to the present-day Han than individuals associated with Amur ancestry [87], indicating interactions with populations south of Mongolia (Figure 2).

Based on PCA, several 8,000–5,000-year-old individuals in the Baikal region, Mongolia, and Inner Mongolia shifted towards the Tibetan cline rather than following the northern Asian cline like most Amur and Primorye region samples [68,81]. In D-statistic and Treemix analyses, 3,000–1,000-year-old individuals from Nepal who shared a close genetic relationship with present-day Tibetans [67] tended to group more with ancient individuals from the Shandong region associated with Yellow River ancestry than with more southern Fujian individuals [68]. Present-day Tibetans show a similar genetic pattern [81]. Thus, the currently sampled Tibetan populations fall within the genetic diversity observed in northern East Asians, although sampling of ancient individuals is needed to determine if Tibetans possess deeply diverging ancestry local to the plateau, as has been suggested previously [88]. Thus far, sparse sampling on the Tibetan Plateau have made it difficult to clarify which demographic processes could explain the shift in the PCA towards the Tibetan cline in populations associated predominantly with Amur ancestry (Figure 2).

Present-day northern and northeastern Siberians (e.g., Even) share a close genetic relationship to northern East Asians, particularly those associated with Amur ancestry. With low levels of Paleosiberian ancestry in present-day Siberians as estimated using qpAdm mixture analysis, Sikora *et al*. concluded that population movement from northern East Asia into Siberia must have occurred [69] (Figure 2). On the Siberian coast by the Beringian Sea, 3,000–2,000-year-old individuals are associated with Native American and Paleosiberian ancestries, suggesting back migration of populations from the Americas that re-entered and mixed with the remaining populations associated with Paleosiberian ancestry still found in the far reaches of northeastern Siberia [69] (Figure 2).

Along the middle to lower reaches of the Yellow River, ancient DNA was retrieved from 9,500–2,000-year-old individuals [68,80,81]. D-statistic, outgroup f3-analysis, and phylogenetic analyses show that these ancient individuals shared a close genetic relationship to each other and ancient individuals from the Amur River region [68,80,81]. In a PCA, they fall at the base of the northern East Asian cline [61,68,80,81]. They are closely related to each other to the exclusion of Amur River populations, indicating a shared ancestry distinct from that found in the Amur River region, denoted here as ***Yellow River ancestry*** (Box 2, Figure 2). Ancient individuals living between the Amur and Yellow River regions showed varying affinities to populations from these two regions, indicating high levels of interaction between populations associated with these two ancestries [80]. In tests of genetic similarity, Yang *et al*. found that

ancient individuals from the lower reaches of the Yellow River in the Shandong region, who were associated with Yellow River ancestry, shared more alleles with present-day East and Southeast Asians [68] than ancient individuals from further south. These results suggested that populations along the Yellow River may have played a major role in the formation of present-day populations of East and Southeast Asia.

5.2 Ancestry and admixture in southern East Asia, Southeast Asia, and the Japanese archipelago

In Fujian, a second 9,000-year-old individual from the Qihe cave site and 8,000–7,000-year-old individuals from Liang Island [68] off the coast of Fujian shared a close genetic relationship with Qihe3, suggesting that they are all associated with the same ancestry [62], denoted here as ***Fujian ancestry*** (Box 2, Figure 2). These ancient individuals also share a close genetic relationship with individuals from other Fujian sites dating to 4,000 years ago, showing that Fujian ancestry persisted at high levels from 12,000-4,000 years. Yang *et al*. [68], used tests of genetic differentiation to show that ancient individuals associated with Yellow River and Fujian ancestries were more genetically differentiated from each other than from present-day northern and southern East Asians. This was largely due to present-day East and Southeast Asians sharing more alleles with individuals associated with Yellow River ancestry than individuals associated with Fujian ancestry. Estimates of mixture proportions using qpAdm showed that most present-day East and Southeast Asians are a mixture of Yellow River, Fujian, and Paleosiberian ancestries [68]. The ancestry of Southeast Asian farmers postdating 4,000 years [63,89], unlike those associated with the deeply diverged Hòabìnhian ancestry, is primarily related to present-day East Asians, indicating that migration southwards from East Asia from 4,000 years ago profoundly influenced the genetic makeup of Southeast Asians (Figure 2). Taken together, these patterns support a hypothesis that migration and admixture between ancient populations associated with Yellow River and Fujian ancestries greatly influenced the current human landscape across East and Southeast Asia.

In the islands of Taiwan and Southeast Asia, sampling of ancient and present-day individuals associated with an Austronesian expansion that eventually occupied far reaching islands of Polynesia are associated with Fujian ancestry. Among present-day populations, Austronesians share the most genetic similarity with 8,000–4,000-year-old individuals from Fujian, showing that unlike mainland Asia, Fujian ancestry persisted at high levels across island Austronesian populations [68]. Some studies have suggested that the Austronesian dispersal moved from Taiwan into the islands of Polynesia [90], but others claim more than one migratory route, including through mainland Southeast Asia [91,92]. Larena *et al*. [75] recently sampled present-day Cordillerans in the Philippines. Using qpAdm mixture analysis, they observed that the

ancient individuals from Fujian and present-day Austronesian speakers in Taiwan showed admixture with individuals associated with Yellow River ancestry, a pattern not observed for the Cordillerans. They used this pattern to argue that Cordillerans must have derived from another migration, perhaps through mainland Southeast Asia [75]. In the islands of Southeast Asia and Oceania, ancient and present-day populations have shown varying levels of ancestry associated with the AA and CSEA lineages [75,93,94], revealing complex layers of migration and interaction between deeply divergent populations in the islands of Southeast Asia and Southwest Pacific (Figure 2).

In the Guangxi region, the Late Upper Paleolithic Longlin was found to be an outgroup to a clade containing ancient individuals associated with Yellow River and Fujian ancestries and associated with Guangxi ancestry. Guangxi ancestry is not associated with any present-day East and Southeast Asians, and 8,000–6,000-year-old individuals from Guangxi (Dushan, Baojianshan) hint at large genetic shifts as the explanation. Dushan and Baojianshan have a closer genetic relationship to individuals associated with Fujian ancestry (Qihe3) than individuals associated with Guangxi ancestry (Longlin), and estimates of mixture proportions using qpAdm showed that they were best described as a mixture of populations associated with Guangxi and Fujian ancestries [62] (Figure 2). In Treemix and qpAdm analyses, the 6,000-year-old Baojianshan additionally showed evidence of admixture with a Southeast Asian hunter-gatherer group, the first detection of Hòabìnhian ancestry in southern China [62] (Figure 2). Historical individuals in Guangxi, dating to 1,500 and 500 years ago, are genetically most similar to present-day southern East Asians, and thus show high levels of admixture from populations associated with Yellow River ancestry.

In the Japanese archipelago, 7,000–3,000-year-old individuals sampled to date associated with the Jōmon culture are all closely related to each other and are associated with Jōmon ancestry. Using Treemix and **f4-ratio analyses** (Box 1), Adachi *et al*. [64] found that present-day Japanese populations had 10% Jōmon ancestry. This finding broadly supports a core tenet of the dual structure hypothesis for the population history of Japan, wherein migrants from mainland Asia, likely through the Korean peninsula, moved to the Japanese archipelago starting 3,000 years ago, and mixed with indigenous Jōmon populations [86,95]. Adachi *et al*. also estimated that present-day Korean and Ulchi populations in northeast Asia show 5%–8% Jōmon ancestry [64]. Furthermore, in f4-statistics, Jōmon individuals show connections to present-day Austronesians and 8,000–7,000-year-old individuals from coastal southern East Asia and Siberia [85,86]. These ties to coastal and island populations suggest that the Jōmon may not have been completely isolated after their migration into the Japanese archipelago (Figure 2).

5.3 Ancestry and admixture in South Asia

In South Asia, sampling has largely been limited to present-day populations. However, recent sampling of ancient individuals from around the Indus Valley up to 5,000 years ago [59,60] revealed two patterns. First, from PCA analyses, 5,000–4,000-year-old individuals were distributed on a cline where individuals at one end were associated with the AASI lineage and individuals at the other end were associated with basal ancestry links to ancient Iranians [59]. Narasimhan *et al*. [59] described the cline between individuals associated with AASI and basal Iranian ancestries the Indus Periphery cline. In a qpAdm analysis these individuals fit best as a source population for younger individuals from the Indus Periphery region and present-day South Asians [59], and thus, I denote the ancestry associated with these individuals as ***Indus Periphery ancestry*** (Box 2, 'IP' in Figure 2). Estimates of mixture proportions using qpAdm in 4,000–3,000-year-old individuals from northern regions of South Asia showed admixture from populations of increased Steppe ancestry associated with the Yamnaya and other Central Asian and Steppe populations [59,74]. This pattern suggests populations associated with Steppe ancestry migrated southward into the Indus Valley from 4,000 years ago, contributing to a widespread signature of Steppe ancestry across Inner Asia [59].

Reich *et al*. performed genome-wide sequencing of present-day Indian populations and found a north to south cline [44]. Comparison with ancient individuals from South Asia showed that all present-day Indians have a mixture of ancestry related to the AASI lineage, basal Iranian ancestry, and Steppe ancestry [59]. Northern and southern Indians are both associated with Indus Periphery ancestry observed in populations near and in the Indus Valley older than 4,000 years [59]. Southern Indian populations possess additional ancestry related to the AASI lineage beyond that found in the ancient Indus Valley individuals, which suggests that ancient individuals representing the AASI lineage, who have yet to be sampled, likely lived in southern India [59]. Northern Indians show genetic patterns similar to those found in ancient populations near the Indus Valley younger than 4,000 years; all show admixture with populations associated with Steppe ancestry [59]. These patterns illustrate that in South Asia, the formation of ancestries associated with northern and southern Indians likely post-dated 4,000 years ago, where northern Indian populations associated with the Indus Periphery cline mixed with populations of Steppe ancestry and southern Indian populations in the Indus Periphery cline primarily mixed with populations of the AASI lineage [59] (Figure 2). Just as in other regions of Asia, admixture played a key role in the formation of present-day Indian populations.

6. Conclusions

At the turn of the century, Cavalli-Sforza and his colleagues pushed for cataloging of genetic variation across multiple present-day human populations worldwide, including Asia, to allow deep insight into human migration patterns [2], culminating in the Human Genome Diversity Project. Prior to the availability of these datasets, Cavalli-Sforza showed how DNA sequences from humans worldwide were key to confirming modern human origins in Africa and revealing the number of dispersals into Eurasia [5]. Denser sampling in the 21st century of present-day Asians and Australasians helped to clarify that a single major dispersal was the primary contributor to all Asians and Australasians today.

While Cavalli-Sforza's primary effort at a regional level was to characterize human genetic history in Europe [3,4], he supported similar initiatives in Asia and recognized the importance of understanding human evolutionary history in that region. Research on Asian human genetics has lagged behind that in European human genetics [96], but that has changed greatly in the last decade, with recent efforts such as the GA100K project for large-scale DNA sequencing and analysis of Asian populations [11]. In 2017, only a handful of individuals were sequenced in Asia [22], but ancient DNA studies in this region have rapidly become available since then. Sampling of ancient humans from Asia in the last five years has fundamentally changed our understanding of human movement and interaction in Asia over the last 45,000 years.

Rapid diversification of an ancestral Asian population led to at least three Asian lineages, associated with Australasians and Negritos (AA), South Asians and Andamanese Islanders (AASI), and East and Southeast Asians (ESEA). Sampling across time and space in eastern Asia indicated that the ESEA lineage was highly sub-structured in the Upper Paleolithic, and populations across Siberia, East Asia, and Southeast Asia are associated with many distinct ancestries. Today, most populations from Siberia, East Asia, and Southeast Asia are associated with a mix of ancestries sampled predominantly in ancient individuals of East Asia. Notably, deeply diverged ancestries not associated with the AA, AASI, and ESEA lineages also had a profound influence in Asia; in Upper Paleolithic Siberia, populations associated with a thriving Ancient Northern Siberian (ANS) ancestry were widespread and closely related to European hunter-gatherers (Figure 2). ANS ancestry had a lasting impact on the ancestral population contributing to the First Americans, as well as steppe populations who would later influence populations in Mongolia, Siberia, and central and southern Asia. Population movement and admixture fundamentally changed the genetic landscape for humans in the last 10,000 years (Figure 2), both because populations that once lived during the Upper Paleolithic did not continue to the present-day and because gene flow between previously isolated populations masked past genetic diversity.

With revolutionary advances in sequencing, computing, and statistics, a rich tapestry of human genetic diversity in Asia is being revealed. However, large temporal and spatial gaps still remain among the individuals sequenced. This hints at the many questions still remaining about human population history in Asia, which make it difficult to determine ancestral ranges and source populations. For instance, human movement and interaction on the Tibetan Plateau and southwest China is still unknown, and the Yangtze River region, a major contributor to rice agriculture whose populations likely influenced many regions of Asia, has yet to be sampled. Many ancestries are represented by only one or a few individuals, rendering it difficult to determine whether to attribute excess allele sharing to population structure or genetic admixture. Many demographic inference tools, like MSMC and momi2, require higher coverage and phased genomes – currently still rare for most ancient DNA studies. Greater advances in ancient DNA sequencing techniques and population genetic inference methods incorporating low coverage genomic data, along with greater systematic sampling across Asia are needed to address these limits. Despite these drawbacks, the research to date has clearly demonstrated the power of both spatial genetic sampling, as advocated by Cavalli-Sforza [1,2], and temporal genetic sampling in revealing the genetic history of humans in Asia.

Ethics Statement

Not applicable.

Consent for Publication

Not applicable.

Availability of Data and Material

GPS coordinates of samples were primarily collected from an annotation file associated with a compiled set of published ancient and present-day human genomic data through the Reich Lab (v44.3), downloadable at https://reich.hms.harvard.edu/allen-ancient-dna-resource-aadr-downloadable-genotypes-present-day-and-ancient-dna-data. Remaining GPS coordinates not included in this dataset were retrieved from tables of the corresponding publications. Maps throughout this article were created using ArcGIS® software by Esri, through a license for the University of Richmond (see https://arcg.is/1i9yL4).

Funding

Competing Interests

The author has declared that no competing interests exist.

Author Contributions

M.A.Y. developed the organization and fully wrote the manuscript.

Acknowledgments

Abbreviations

The following abbreviations, listed alphabetically, are used in this manuscript.

AA   Australasian
AASI   Ancient Ancestral Southern Indian
ANS   Ancient Northern Siberian
ESEA   East and Southeast Asian
IP   Indus Periphery
PCA   principal component analysis
SNP   single nucleotide polymorphism

References

1. Edwards AWF, Luigi Luca Cavalli-Sforza. 25 January 1922—31 August 2018. Biographical Memoirs of Fellows of the Royal Society. 2021;70:79–105. DOI
2. Cavalli-Sforza LL. The Human Genome Diversity Project: past, present and future. Nat Rev Genet. 2005;6(4):333–340. DOI
3. Piazza A, Rendine S, Minch E, Menozzi P, Mountain J, Cavalli-Sforza LL. Genetics and the origin of European languages. PNAS. 1995;92(13):5836–5840. DOI

4. Ammerman AJ, Cavalli-Sforza LL. Measuring the Rate of Spread of Early Farming in Europe. Man. 1971;6(4):674–688. [DOI](#)

5. Jin L, Underhill PA, Doctor V, Davis RW, Shen P, Cavalli-Sforza LL, et al. Distribution of haplotypes from a chromosome 21 region distinguishes multiple prehistoric human migrations. PNAS. 1999;96(7):3796–3800. [DOI](#)

6. Chen K-H, Cavalli-Sforza LL. Surnames in Taiwan: Interpretations Based on Geography and History. Hum Biol. 1983;55(2):367–374.

7. Du R, Yuan Y, Hwang J, Mountain J, Cavalli-Sforza LL. Chinese Surnames and the Genetic Differences between North and South China. JCLMS. 1992;(5):1–93.

8. Chu JY, Huang W, Kuang SQ, Wang JM, Xu JJ, Chu ZT, et al. Genetic relationship of populations in China. PNAS. 1998;95(20):11763–11768. [DOI](#)

9. Consortium THP-AS. Mapping Human Genetic Diversity in Asia. Science. 2009;326(5959):1541–1545. [DOI](#)

10. Ngamphiw C, Assawamakin A, Xu S, Shaw PJ, Yang JO, Ghang H, et al. PanSNPdb: The Pan-Asian SNP Genotyping Database. PLOS ONE. 2011;6(6):e21451. [DOI](#)

11. Wall JD, Stawiski EW, Ratan A, Kim HL, Kim C, Gupta R, et al. The GenomeAsia 100K Project enables genetic discoveries across Asia. Nature. 2019;576(7785):106–111. [DOI](#)

12. Altshuler D, Donnelly P. The International HapMap Consortium. A haplotype map of the human genome. Nature. 2005;437(7063):1299–1320. [DOI](#)

13. Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, et al. A second generation human haplotype map of over 3.1 million SNPs. Nature. 2007;449(7164):851–861. [DOI](#)

14. Auton A, Abecasis GR, Altshuler DM, Durbin RM, Abecasis GR, Bentley DR, et al. A global reference for human genetic variation. Nature. 2015;526(7571):68–74. [DOI](#)

15. Mallick S, Li H, Lipson M, Mathieson I, Gymrek M, Racimo F, et al. The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. Nature. 2016;538(7624):201–206. [DOI](#)

16. Bergström A, McCarthy SA, Hui R, Almarri MA, Ayub Q, Danecek P, et al. Insights into human genetic variation and population history from 929 diverse genomes. Science. 2020;367(6484):eaay5012. [DOI](#)

17. Gilbert MTP, Bandelt H-J, Hofreiter M, Barnes I. Assessing ancient DNA studies. Trends Ecol Evol. 2005;20(10):541–544. [DOI](#)

18. Llamas B, Valverde G, Fehren-Schmitz L, Weyrich LS, Cooper A, Haak W. From the field to the laboratory: Controlling DNA contamination in human ancient DNA research in the high-throughput sequencing era. Sci. Technol Archaeol Res. 2017;3(1):1–14. [DOI](#)

19. Dabney J, Meyer M, Pääbo S. Ancient DNA Damage. Cold Spring Harb Perspect Biol. 2013;5(7):a012567. [DOI](#)

20. Green RE, Briggs AW, Krause J, Prüfer K, Burbano HA, Siebauer M, et al. The Neandertal genome and ancient DNA authenticity. EMBO J. 2009;28(17):2494–2502. [DOI]

21. Skoglund P, Mathieson I. Ancient Genomics of Modern Humans: The First Decade. Annu Rev Genomics Hum Genet. 2018;19(1):381–404. [DOI]

22. Marciniak S, Perry GH. Harnessing ancient genomes to study the history of human adaptation. Nat Rev Genet. 2017;18(11):659–674. [DOI]

23. Yang MA, Fu Q. Insights into Modern Human Prehistory Using Ancient Genomes. Trends Genet. 2018;34(3):184–196. [DOI]

24. Henn BM, Cavalli-Sforza LL, Feldman MW. The great human expansion. PNAS. 2012;109(44):17758–17764. [DOI]

25. Relethford JH. Genetic evidence and the modern human origins debate. Heredity. 2008;100(6):555–563. [DOI]

26. Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, et al. A Draft Sequence of the Neandertal Genome. Science. 2010;328(5979):710. [DOI]

27. Reich D, Green RE, Kircher M, Krause J, Patterson N, Durand EY, et al. Genetic history of an archaic hominin group from Denisova Cave in Siberia. Nature. 2010;468(7327):1053–1060. [DOI]

28. Stringer C. Why we are not all multiregionalists now. Trends Ecol Evol. 2014;29(5):248–251. [DOI]

29. Fu Q, Mittnik A, Johnson PLF, Bos K, Lari M, Bollongino R, et al. A Revised Timescale for Human Evolution Based on Ancient Mitochondrial Genomes. Curr Biol. 2013;23(7):553–559. [DOI]

30. Rasmussen M, Guo X, Wang Y, Lohmueller KE, Rasmussen S, Albrechtsen A, et al. An Aboriginal Australian Genome Reveals Separate Human Dispersals into Asia. Science. 2011;334(6052):94–98. [DOI]

31. Slon V, Hopfe C, Weiß CL, Mafessoni F, Rasilla M de la, Lalueza-Fox C, et al. Neandertal and Denisovan DNA from Pleistocene sediments. Science. 2017;356(6338):605–608. [DOI]

32. Meyer M, Kircher M, Gansauge M-T, Li H, Racimo F, Mallick S, et al. A High-Coverage Genome Sequence from an Archaic Denisovan Individual. Science. 2012;338(6104):222–226. [DOI]

33. Zhang D, Xia H, Chen F, Li B, Slon V, Cheng T, et al. Denisovan DNA in Late Pleistocene sediments from Baishiya Karst Cave on the Tibetan Plateau. Science. 2020;370(6516):584–587. [DOI]

34. Chen F, Welker F, Shen C-C, Bailey SE, Bergmann I, Davis S, et al. A late Middle Pleistocene Denisovan mandible from the Tibetan Plateau. Nature. 2019;569(7756):409–412. [DOI]

35. Sankararaman S, Mallick S, Patterson N, Reich D. The Combined Landscape of Denisovan and Neanderthal Ancestry in Present-Day Humans. Curr Biol. 2016;26(9):1241–1247. [DOI]

36. Qin P, Stoneking M. Denisovan Ancestry in East Eurasian and Native American Populations. Molecular Biology and Evolution. 2015;32(10):2665–2674. DOI

37. Prüfer K, Racimo F, Patterson N, Jay F, Sankararaman S, Sawyer S, et al. The complete genome sequence of a Neanderthal from the Altai Mountains. Nature. 2014;505(7481):43–49. DOI

38. Jinam TA, Phipps ME, Aghakhanian F, Majumder PP, Datar F, Stoneking M, et al. Discerning the Origins of the Negritos, First Sundaland People: Deep Divergence and Archaic Admixture. Genome Biol Evol. 2017;9(8):2013–2022. DOI

39. Mondal M, Casals F, Xu T, Dall'Olio GM, Pybus M, Netea MG, et al. Genomic analysis of Andamanese provides insights into ancient human migration into Asia and adaptation. Nat Genet. 2016;48(9):1066–1070. DOI

40. Malaspinas A-S, Westaway MC, Muller C, Sousa VC, Lao O, Alves I, et al. A genomic history of Aboriginal Australia. Nature. 2016;538(7624):207–214. DOI

41. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 2009;19(9):1655–1664. DOI

42. Haak W, Lazaridis I, Patterson N, Rohland N, Mallick S, Llamas B, et al. Massive migration from the steppe was a source for Indo-European languages in Europe. Nature. 2015;522(7555):207–211. DOI

43. Fu Q, Hajdinjak M, Moldovan OT, Constantin S, Mallick S, Skoglund P, et al. An early modern human from Romania with a recent Neanderthal ancestor. Nature. 2015;524(7564):216–219. DOI

44. Reich D, Thangaraj K, Patterson N, Price AL, Singh L. Reconstructing Indian population history. Nature. 2009;461(7263):489–494. DOI

45. Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, et al. Ancient Admixture in Human History. Genetics. 2012;192(3):1065–1093. DOI

46. Kamm J, Terhorst J, Durbin R, Song YS. Efficiently Inferring the Demographic History of Many Populations With Allele Count Data. J Am Stat Assoc. 2020;115(531):1472–1487. DOI

47. Excoffier L, Dupanloup I, Huerta-Sánchez E, Sousa VC, Foll M. Robust Demographic Inference from Genomic and SNP Data. PLOS Genet. 2013;9(10):e1003905. DOI

48. Lawson DJ, Hellenthal G, Myers S, Falush D. Inference of Population Structure using Dense Haplotype Data. PLOS Genet. 2012;8(1):e1002453. DOI

49. Steinrücken M, Kamm J, Spence JP, Song YS. Inference of complex population histories using whole-genome sequences from multiple populations. PNAS. 2019;116(34):17115–17120. DOI

50. Schiffels S, Durbin R. Inferring human population size and separation history from multiple genome sequences. Nat Genet. 2014;46(8):919–925. DOI

51. Raghavan M, Skoglund P, Graf KE, Metspalu M, Albrechtsen A, Moltke I, et al. Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. Nature. 2014;505(7481):87–91. DOI

52. Patterson N, Price AL, Reich D. Population Structure and Eigenanalysis. PLOS Genet. 2006;2(12):e190. DOI

53. Menozzi P, Piazza A, Cavalli-Sforza L. Synthetic maps of human gene frequencies in Europeans. Science. 1978;201(4358):786–792. DOI

54. Lazaridis I, Patterson N, Mittnik A, Renaud G, Mallick S, Kirsanow K, et al. Ancient human genomes suggest three ancestral populations for present-day Europeans. Nature. 2014;513(7518):409–413. DOI

55. Pickrell JK, Pritchard JK. Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. PLOS Genet. 2012;8(11):e1002967. DOI

56. Pagani L, Lawson DJ, Jagoda E, Mörseburg A, Eriksson A, Mitt M, et al. Genomic analyses inform on migration events during the peopling of Eurasia. Nature. 2016;538(7624):238–242. DOI

57. Maier C, Craig A, Adams DM. Language use in ancestry research and estimation. J Forensic Sci. 2021;66(1):11–24. DOI

58. Mathieson I, Scally A. What is ancestry? PLoS Genet. 2020;16(3):e1008624. DOI

59. Narasimhan VM, Patterson N, Moorjani P, Rohland N, Bernardos R, Mallick S, et al. The formation of human populations in South and Central Asia. Science. 2019;365(6457):eaat7487. DOI

60. Shinde V, Narasimhan VM, Rohland N, Mallick S, Mah M, Lipson M, et al. An Ancient Harappan Genome Lacks Ancestry from Steppe Pastoralists or Iranian Farmers. Cell. 2019;179(3):729-735.e10. DOI

61. Mao X, Zhang H, Qiao S, Liu Y, Chang F, Xie P, et al. The deep population history of northern East Asia from the Late Pleistocene to the Holocene. Cell. 2021;184(12):3256-3266.e13. DOI

62. Wang T, Wang W, Xie G, Li Z, Fan X, Yang Q, et al. Human population history at the crossroads of East and Southeast Asia since 11,000 years ago. Cell. 2021;184(14):3829-3841.e21. DOI

63. McColl H, Racimo F, Vinner L, Demeter F, Gakuhari T, Moreno-Mayar JV, et al. The prehistoric peopling of Southeast Asia. Science. 2018;361(6397):88–92. DOI

64. Adachi N, Kanzawa-Kiriyama H, Nara T, Kakuda T, Nishida I, Shinoda K-I. Ancient genomes from the initial Jomon period: new insights into the genetic history of the Japanese archipelago. Anthropol Sci. 2021;129(1):13–22. DOI

65. Massilani D, Skov L, Hajdinjak M, Gunchinsuren B, Tseveendorj D, Yi S, et al. Denisovan ancestry and population history of early East Asians. Science. 2020;370(6516):579–583. DOI

66. Yang MA, Gao X, Theunert C, Tong H, Aximu-Petri A, Nickel B, et al. 40,000-Year-Old Individual from Asia Provides Insight into Early Population Structure in Eurasia. Curr Biol. 2017;27(20):3202-3208.e9. DOI

67. Jeong C, Ozga AT, Witonsky DB, Malmström H, Edlund H, Hofman CA, et al. Long-term genetic stability and a high-altitude East Asian origin for the peoples of the high valleys of the Himalayan arc. PNAS. 2016;113(27):7485–7490. [DOI](#)

68. Yang MA, Fan X, Sun B, Chen C, Lang J, Ko Y-C, et al. Ancient DNA indicates human population shifts and admixture in northern and southern China. Science. 2020;369(6501):282–288. [DOI](#)

69. Sikora M, Pitulko VV, Sousa VC, Allentoft ME, Vinner L, Rasmussen S, et al. The population history of northeastern Siberia since the Pleistocene. Nature. 2019;570(7760):182–188. [DOI](#)

70. Fu Q, Li H, Moorjani P, Jay F, Slepchenko SM, Bondarev AA, et al. Genome sequence of a 45,000-year-old modern human from western Siberia. Nature. 2014;514(7523):445–449. [DOI](#)

71. Lazaridis I, Nadel D, Rollefson G, Merrett DC, Rohland N, Mallick S, et al. Genomic insights into the origin of farming in the ancient Near East. Nature. 2016;536(7617):419–424. [DOI](#)

72. Moreno-Mayar JV, Potter BA, Vinner L, Steinrücken M, Rasmussen S, Terhorst J, et al. Terminal Pleistocene Alaskan genome reveals first founding population of Native Americans. Nature. 2018;553(7687):203–207. [DOI](#)

73. Yu H, Spyrou MA, Karapetian M, Shnaider S, Radzevičiūtė R, Nägele K, et al. Paleolithic to Bronze Age Siberians Reveal Connections with First Americans and across Eurasia. Cell. 2020;181(6):1232-1245.e20. [DOI](#)

74. Damgaard P de B, Martiniano R, Kamm J, Moreno-Mayar JV, Kroonen G, Peyrot M, et al. The first horse herders and the impact of early Bronze Age steppe expansions into Asia. Science. 2018;360(6396):eaar7711. [DOI](#)

75. Larena M, Sanchez-Quinto F, Sjödin P, McKenna J, Ebeo C, Reyes R, et al. Multiple migrations to the Philippines during the last 50,000 years. PNAS. 2021;118(13):e2026132118. [DOI](#)

76. Hajdinjak M, Mafessoni F, Skov L, Vernot B, Hübner A, Fu Q, et al. Initial Upper Palaeolithic humans in Europe had recent Neanderthal ancestry. Nature. 2021;592(7853):253–237. [DOI](#)

77. Bollongino R, Tresset A, Vigne J-D. Environment and excavation: Pre-lab impacts on ancient DNA analyses. Comptes Rendus Palevol. 2008;7(2):91–98. [DOI](#)

78. Fu Q, Meyer M, Gao X, Stenzel U, Burbano HA, Kelso J, et al. DNA analysis of an early modern human from Tianyuan Cave, China. PNAS. 2013;110(6):2223–2227. [DOI](#)

79. Fu Q, Posth C, Hajdinjak M, Petr M, Mallick S, Fernandes D, et al. The genetic history of Ice Age Europe. Nature. 2016;534(7606):200–205. [DOI](#)

80. Ning C, Li T, Wang K, Zhang F, Li T, Wu X, et al. Ancient genomes from northern China suggest links between subsistence changes and human migration. Nat Commun. 2020;11(1):2700. [DOI](#)

81. Wang C-C, Yeh H-Y, Popov AN, Zhang H-Q, Matsumura H, Sirak K, et al. Genomic insights into the formation of human populations in East Asia. Nature. 2021;591(7850):413–419. [DOI](#)

82. Siska V, Jones ER, Jeon S, Bhak Y, Kim H-M, Cho YS, et al. Genome-wide data from two early Neolithic East Asian individuals dating to 7700 years ago. Sci Adv. 2017;3(2):e1601877. [DOI](#)

83. Kanzawa-Kiriyama H, Jinam TA, Kawai Y, Sato T, Hosomichi K, Tajima A, et al. Late Jomon male and female genome sequences from the Funadomari site in Hokkaido, Japan. Anthropol Sci. 2019;127(2):83–108. [DOI](#)

84. Kanzawa-Kiriyama H, Kryukov K, Jinam TA, Hosomichi K, Saso A, Suwa G, et al. A partial nuclear genome of the Jomons who lived 3000 years ago in Fukushima, Japan. J Hum Genet. 2017;62(2):213–221. [DOI](#)

85. Gakuhari T, Nakagome S, Rasmussen S, Allentoft ME, Sato T, Korneliussen T, et al. Ancient Jomon genome sequence analysis sheds light on migration patterns of early East Asian populations. Commun Biol. 2020;3(1):1–10. [DOI](#)

86. de Boer E, Yang MA, Kawagoe A, Barnes GL. Japan considered from the hypothesis of farmer/language spread. Evolut Hum Sci. 2020;2:e13. [DOI](#)

87. Jeong C, Wang K, Wilkin S, Taylor WTT, Miller BK, Bemmann JH, et al. A Dynamic 6,000-Year Genetic History of Eurasia's Eastern Steppe. Cell. 2020;183(4):890-904.e29. [DOI](#)

88. Lu D, Lou H, Yuan K, Wang X, Wang Y, Zhang C, et al. Ancestral Origins and Genetic History of Tibetan Highlanders. Am J Hum Genet. 2016;99(3):580–594. [DOI](#)

89. Lipson M, Cheronet O, Mallick S, Rohland N, Oxenham M, Pietrusewsky M, et al. Ancient genomes document multiple waves of migration in Southeast Asian prehistory. Science. 2018;361(6397):92–95. [DOI](#)

90. Ko AM-S, Chen C-Y, Fu Q, Delfin F, Li M, Chiu H-L, et al. Early Austronesians: Into and Out Of Taiwan. Am J Hum Genet. 2014;94(3):426–436. [DOI](#)

91. Soares PA, Trejaut JA, Rito T, Cavadas B, Hill C, Eng KK, et al. Resolving the ancestry of Austronesian-speaking populations. Hum Genet. 2016;135(3):309–326. [DOI](#)

92. Lipson M, Loh P-R, Patterson N, Moorjani P, Ko Y-C, Stoneking M, et al. Reconstructing Austronesian population history in Island Southeast Asia. Nat Commun. 2014;5(1):4689. [DOI](#)

93. Choin J, Mendoza-Revilla J, Arauna LR, Cuadros-Espinoza S, Cassar O, Larena M, et al. Genomic insights into population history and biological adaptation in Oceania. Nature. 2021;592(7855):583–589. [DOI](#)

94. Skoglund P, Posth C, Sirak K, Spriggs M, Valentin F, Bedford S, et al. Genomic insights into the peopling of the Southwest Pacific. Nature. 2016;538(7626):510–513. [DOI](#)

95. Hudson MJ, Nakagome S, Whitman JB. The evolving Japanese: the dual structure hypothesis at 30. Evol Hum Sci. 2020;2:e6. DOI
96. Popejoy AB, Fullerton SM. Genomics is failing on diversity. Nature. 2016;538(7624):161–164. DOI