School of Doctoral Studies in Biological Sciences
University of South Bohemia in České Budějovice
Faculty of Science

Human population history and evolution of culture: A phylogenetic approach

Ph.D. Thesis

Mgr. Pavel Duda

Supervisor: prof. RNDr. Jan Zrzavý, CSc.

Department of Zoology, Faculty of Science, University of South Bohemia

České Budějovice 2017

This thesis should be cited as:

Duda, P. 2017: Human population history and evolution of culture: A phylogenetic approach, Ph.D. Thesis. University of South Bohemia, Faculty of Science, School of Doctoral Studies in Biological Sciences, České Budějovice, Czech Republic, 263 pp.

- **Annotation**

This thesis uses phylogenetic and phylogenetic comparative methods to investigate human population history and evolution of cross-cultural variation. Such research requires a robust phylogenetic hypothesis of human populations as a framework. This thesis uses a supertree approach to infer a composite phylogeny of human populations based on published phylogenetic trees based on genetic and linguistic data. It assesses the stability of the inferred supertree topology and identifies individual populations whose phylogenetic position is particularly unstable. It assesses the congruence between genetic and linguistic data and tests hypothesis about language relationships and coevolution between genes and languages on a global scale. The supertree is used to reconstruct the origin and evolution of religious beliefs and behaviors using a global sample of hunter-gatherer populations and a set of phylogenetic comparative methods. This thesis also describes evolution of Central African pygmies, a group of human populations that represents an interesting case of morphological and cultural adaptation in human species, and reviews the history and current developments of phylogenetic (tree building) approaches to cross-cultural variation.

- **Declaration [in Czech]**

Prohlašuji, že svoji disertační práci jsem vypracoval samostatně pouze s použitím pramenů a literatury uvedených v seznamu citované literatury.

Prohlašuji, že v souladu s § 47b zákona č. 111/1998 Sb. v platném znění souhlasím se zveřejněním své disertační práce, a to v nezkrácené podobě elektronickou cestou ve veřejně přístupné části databáze STAG provozované Jihočeskou univerzitou v Českých Budějovicích na jejích internetových stránkách, a to se zachováním mého autorského práva k odevzdanému textu této kvalifikační práce. Souhlasím dále s tím, aby toutéž elektronickou cestou byly v souladu s uvedeným ustanovením zákona č. 111/1998 Sb. zveřejněny posudky školitele a oponentů práce i záznam o průběhu a výsledku obhajoby kvalifikační práce. Rovněž souhlasím s porovnáním textu mé kvalifikační práce s databází kvalifikačních prací Theses.cz provozovanou Národním registrem vysokoškolských kvalifikačních prací a systémem na odhalování plagiátů.

V Českých Budějovicích, 27.2.2017

…………………

Mgr. Pavel Duda

This thesis originated at the Faculty of Science, University of South Bohemia, in the study programme Zoology.

Přírodovědecká
fakulta
Faculty
of Science

- **Financial support**

This thesis was supported by Grant Agency of the University of South Bohemia (GA JU 042/2013/P; GA JU 140/2013/P) and The Czech Science Foundation (GA ČR 16-26369S; GA ČR 14-36098G).

Acknowledgements

A special thank you goes to my supervisor Jan Zrzavý, who have been supervising me since I started working on my bachelor’s thesis. I am grateful for his interesting ideas and keen insights, his relentless enthusiasm and sense of humor. I especially appreciate that he enabled me to pursue my own research interests. He provided indispensable assistance and showed lots of patience while letting me do what I want, a decision he had to defend on a national radio. Thanks to David Storch for taking interest in my research and for inviting me to the Center for Theoretical Study (CTS). The time spent at the CTS was worthwhile and inspiring. Thanks to Vojtěch Novotný for his interest in my research and support. I am grateful to Olaf Bininda-Emonds for allowing me to visit him at the department Systematics and Evolutionary Biology at The Carl von Ossietzky University of Oldenburg. Our discussions were invaluable and I like to remember the time spent in Oldenburg. A big thank you goes to Hervey Peoples, who approached me at the conference of Human Behavior and Evolution Society in Montpellier and persuaded me to work on “The origins of religion” with her. The production of the paper was complicated to say the least but through countless of extensive emails, multiple re-analyses, rewritings, and resubmissions, we were able to produce a worthwhile piece of research. Thanks to Frank Marlowe for his contribution to the paper. Thanks to Jan Havlíček for inviting me to co-write books *Pygmejové* and *Biologické a evoluční teorie kultury* and for introducing me to University Research Center (UNCE), Nature and culture working group at Charles University of Prague. Thanks to Linda Hroníková, Zuzana Schierová, and Lenka Ovčáčková for their editorial assistance. I am grateful to all my colleagues and friends at the Department of Zoology. František Sedláček was always helpful and benevolent as chair of the specialist board of the Ph.D. study programme Zoology. I was lucky to have shared the office with Honza Robovský, Míša Másílková, Terka Holicová, and others, to whom I am grateful for pleasant working environment, engaging discussions, and an occasional welcome distraction. I would like to thank my family, especially to my stepfather František, who enabled me to study. I am immensely grateful to my wife Pavla, who was with me during the whole Ph.D. study, and who I give most credit for the fact that the past five years was, overall, the happiest, most fulfilling time of my life.

- **List of publications and author's contribution**

This thesis is based on the following publications (listed chronologically):

I. **Duda, P.**, & Zrzavý, J. (2016). Human population history revealed by a supertree approach. *Scientific Reports, 6*, 10. doi:10.1038/srep29890
(IF $_{2015}$ = 5.228)

Pavel Duda is the first and corresponding author of the study. He conceived and designed the study with J. Zrzavý, analyzed the data, created the artworks and co-wrote the paper with J. Zrzavý.

II. Peoples, H. C., **Duda, P.**, & Marlowe, F. W. (2016). Hunter-gatherers and the origins of religion. *Human Nature-an Interdisciplinary Biosocial Perspective, 27*(3), 261-282. doi:10.1007/s12110-016-9260-0
(IF $_{2015}$ = 1.895)

Pavel Duda designed the study with H. Peoples, analyzed the data, created the artworks and co-wrote the paper with H. Peoples.

III. **Duda, P.** (2015). Pygmejové pohledem evoluční biologie. In: Hroníková, L., Schierová, Z. (Eds.). *Pygmejové: nejmenší lidé pohledem antropologie & Šebestova sbírka v Hrdličkově muzeu člověka PřF UK* (pp. 18-48). Praha: Academia.

Pavel Duda searched the literature, created the artworks and wrote the manuscript.

IV. **Duda, P.** (*in press*). Kulturní fylogenetika - Využití fylogenetických metod ke studiu Evoluční historie jazyka a kultury. In: Ovčáčková, L. (Ed.), *Biologické a evoluční teorie kultury*. Praha: Academia.

Pavel Duda searched the literature, created the artworks and wrote the manuscript.

- **Co-author agreement**

Jan Zrzavý, the supervisor of this Ph.D. thesis and co-author of the paper “Human population history revealed by a supertree approach”, fully acknowledges the major contribution of Pavel Duda to the presented paper.

…………………………………………

prof. RNDr. Jan Zrzavý, CSc., co-author

Hervey Peoples, the first and corresponding author of the paper “Hunter-gatherers and the origins of religion”, fully acknowledges the major contribution of Pavel Duda to the presented paper.

Hervey C. Peoples

…………………………………………

Hervey C. Peoples, first author

• Contents

Introduction

1. Evolution of culture

Skimming through the hundreds of variables listed in George P. Murdock's and Douglas R. White's *Standard Cross-Cultural Sample* (Murdock & White, 1969), one must wonder whether there is any rhyme or reason behind the bewildering array of subsistence modes, life-history strategies, behavior, and culture in different human populations around the globe.

The diversity of human cultures is truly stunning. Human populations that live in an extraordinary range of environments differ in their social organization, patterns of reproduction, and parental behavior, and have diverse and elaborate beliefs, social institutions, traditions, and norms. Humans practice monogamy, polygyny, or polyandry, pay dowry or bride price, and trace their ancestry through the paternal or the maternal line of descent. They follow various rules of inheritance of social status and resources, operate under different modes of subsistence as hunters-gatherers, herders, or farmers, eat different foods, make and use different tools for foraging, farming and other purposes. There are over 7,000 languages spoken worldwide (Lewis et al., 2016). Human intraspecific diversity is, based on the number of ethno-linguistic groups, comparable to the species-level diversity of tetrapod classes (i.e., amphibians, reptiles, birds, and mammals). This diversity of ethno-linguistic groups goes hand in hand with a wide range of life-history, behavioral and cultural adaptations that have enabled humans to inhabit virtually every environment on Earth (Brown et al., 2011).

The aim of this thesis is to investigate whether and how the methods developed by evolutionary biology to study the mechanisms that generate biological diversity can help us understand the patterns of human behavioral and cultural diversity.

Ever since the publication of Darwin's *On the Origin of Species* (Darwin, 1859), there has been an ongoing debate about whether and how evolutionary ideas can be applied to human culture. Many of the fundamental features of biological and cultural evolution are demonstrably analogous. These include the crucial aspects of evolution - variation, inheritance and selection (Mesoudi et al., 2004; Duda, *in press*). There are parallels between biological and cultural evolution both at the gene or cultural trait level and at the species or population level (Mace & Holden, 2005). Darwin saw similarities between evolution of biological species and evolution of languages (Darwin, 1859), and nineteenth-century linguists, anthropologists and archeologists used phylogenetic thinking ("tree thinking") to infer genealogies of languages and cultural artifacts in the same way biologists use it to infer phylogenetic trees of biological species. During the 19th and 20th century there have been intriguing parallel developments in evolutionary biology and social sciences (historical linguistics, cultural anthropology and archeology) (Atkinson & Gray, 2005; Rivero, 2016; Duda, *in press*).

During the second half of the 19th century, evolutionism has become an influential view that has greatly affected anthropology as a scientific discipline. Evolutionism was the earliest school of thought in anthropology, represented by Herbert Spencer, Edward Burnett Tylor, Lewis Henry Morgan, James George Frazer and others. However, these early anthropologists understood the process of cultural evolution as a progress through predefined developmental stages of society from simple to complex, "from savagery through barbarism to civilization"

(Morgan, 1871), and viewed European culture as a pinnacle of the social evolution. Such "ladder-like" view of cultural evolution as a lineal progression towards more complex types of socio-political organization presents a fundamental misunderstanding of the process of evolution as "descent with modification" (Darwin, 1859). One of the most important contributions of phylogenetic reasoning is the emphasis on the sister-group relationships derived from a phylogenetic tree, instead of the traditional ancestor-descendant continuum. No culture can be a priori considered "primitive", or the lower grade of the other as all cultures are equally distant from their common ancestor (at least temporally) and the process of (common) descent with modification imply a tree-like pattern with many parallel branches. In the early 20th century, evolutionism has been replaced by cultural relativism represented by Franz Boas and his disciples. Boas strongly rejected the idea of cultural evolution, arguing that the theories of evolutionist anthropologists were speculative and unscientific. He claimed that culture developed independently of other characteristics of human populations. The idea that culture, "race", and language constitute mutually independent and unrelated determinants of human existence has become a central tenet of modern anthropology.

In the late 19th and early 20th century, another important controversy arose as linguists and cultural anthropologists have realized that elements of culture are not transmitted only "vertically" along the lines of descent, but also "horizontally" between them. The dispute over the relative importance of horizontal transmission (diffusion of culture between contemporaneous populations through intermarriage, trade, exchange, etc.) has virtually paralyzed the science of cultural evolution for about half a century and remains an important issue to date (Collard et al., 2006; Borgerhoff Mulder et al., 2006; Mesoudi et al. 2006).

Beginning with Spencer and others who brought the term "evolution" into anthropology, anthropologists of the 19th and 20th century have used information obtained by comparisons across human populations to test the ideas about cultural evolution. Since the 1950s, many of these studies have used cross-cultural data accumulated by George P. Murdock who published his first cross-cultural data set, the *World Ethnographic Sample*, including 565 populations in 1957 (Murdock, 1957). In 1967, Murdock published *Ethnographic Atlas* including 1,250 populations (Murdock, 1967). Cross-cultural research became a major branch of anthropology. Most cross-cultural researchers studied functional associations between cultural traits by testing for correlation between two or more traits across a group of cultures (Mace & Pagel, 1994).

However, Francis Galton realized as early as 1889 that cultures cannot be treated as independent for purposes of investigating cross-cultural correlations. The patterns of historical relatedness among cultures mean that they cannot be assumed to have evolved or acquired their particular characteristics independently. This realization has later became known as phylogenetic non-independence or "Galton's problem" (Naroll, 1961; 1965).

Murdock attempted to solve Galton's problem by categorizing the 1,250 populations in the *Ethnographic Atlas* into a smaller number of 186 clusters or "sampling provinces". He assumed that these clusters represented independent groupings while the cultures within a cluster would display many similarities owing to both geographical proximity and common descent. Together with Douglas R. White, Murdock also developed the *Standard Cross-Cultural Sample* (SCCS), including 186 "culturally independent" populations (Murdock & White, 1969). Each cluster in the *Ethnographic Atlas* was represented by a single population in the SCCS. By focusing attention on a limited sample of 186 cultures, the data have steadily improved in scope and quality. The SCCS now contains well over 2,000 coded variables ranging from subsistence to religion and remains widely used in cross-cultural research.

However, the SCCS fails to provide a principled solution to Galton's problem. By using a limited sample of "culturally independent" populations, it merely moves back in time the issue of historical relatedness. The clusters in *Ethnographic Atlas* are not independent for the same logical reasons that cultures within a cluster are not. There are no such things as "independent cultures", only cultures with differing degrees of relatedness, and Murdock's clusters must also be related, even if more distantly. The SCCS populations cannot be independent on variables that evolved in or were acquired and retained by common ancestors of these clusters (Mace & Pagel, 1994). The correct way to distinguish "functional" from "historical" associations in cross-cultural studies is to account for differing degrees of relatedness across populations.

The same problem that concerns testing adaptive hypotheses using cross-cultural comparison applies to identifying instances of horizontal cultural transmission (cultural diffusion) between populations. It is impossible to evaluate the importance of different modes of cultural transmission (horizontal and vertical) in cultural macroevolution without some notion of phylogeny. The term "horizontal transmission" is actually meaningless unless we assume that branching process did indeed underpin human evolutionary history and cultural diversification (Mace & Jordan, 2011).

Theoretical foundations of the field of cultural evolution were laid in the 1970s by Luigi L. Cavalli-Sforza and Marcus Feldman who borrowed and adapted models from population genetics and epidemiology to describe the processes of (horizontal and vertical) cultural transmission, cultural change, and gene-culture coevolution (Cavalli-Sforza & Feldman, 1973, 1981). Barry Hewlett was the first to investigate the significance of different processes of cultural transmission, described by Cavalli-Sforza and Feldman, in non-industrial society, the Aka pygmies in Central African (Hewlett & Cavalli-Sforza, 1986). Robert Boyd and Peter Richerson further developed the field of cultural evolution by introducing a variety of novel ideas and methods (Boyd & Richerson, 1985). These ideas were later popularized in Boyd and Richerson's book *Not by genes alone* (Richerson & Boyd, 2005). Paraphrasing the famous essay by Theodosius Dobzhansky (Dobzhansky, 1973), Richerson and Boyd asserted that "Nothing about culture makes sense except in the light of evolution". In the 1990s the science of cultural evolution was further developed thanks to the work of Mark Pagel and Ruth Mace (Mace & Pagel, 1994) who emphasized the need to use phylogenies as tools to interpret various aspects of human cultural variation from languages to social norms. Thus, Mace and Pagel introduced formal phylogenetic comparative methods to anthropology. In the late 1990s and the early 2000s, the first phylogenetic cross-cultural studies and phylogenetic analyses of cultural traits have emerged (Holden & Mace, 2003; Sellen and Mace, 1997; Holden & Mace, 2003).

Today, the science of cultural evolution is divided into two sub-disciplines: *gene-culture coevolution* (or *dual inheritance theory*) and *cultural phylogenetics* (Laland & Brown, 2011). The first sub-discipline is largely build on research made by Cavalli-Sforza and Feldman and Richerson and Boyd, and investigates how genes and culture interact, how they adapt to one another, or how they affects one another's selective environment. Cultural phylogenetics is largely build on research made by Mace and Pagel and investigates evolution of the culture itself, treating human populations as biologists treat species in phylogenetic analysis (Pagel &

Mace, 2004). It also investigates to what extent are the patterns and processes of cultural evolution comparable to those of biological evolution.

Such research requires robust phylogenetic and statistical inference techniques. Therefore, the advances in the study of cultural macroevolution depend on our ability to infer phylogenies of human populations and to use these phylogenies while analyzing comparative data.

2. Phylogenetic comparative methods

Comparative method is an analytical approach based on the comparison of different objects with the aim to elucidate the processes that responsible for their diversity. Phylogenetic comparative method is the analytical study of species, populations, and individuals in which phylogenetic relationships are taken into account to distinguish between similarities that arose for functional reasons (i.e., selection) and those that are due to historical reasons (i.e., common ancestry) (Garamszegi, 2014).

Beginning with the late 1960s, the developments in phylogenetics were accompanied by the developments of comparative methods (until that time, comparative biology developed its statistical tools independently of phylogenetics). Phylogenetic comparative methods introduced a crucial historical dimension of biological evolution to the cross-species comparisons. These ideas and methods were extremely valuable since recognizing adaptations was the aim of many comparative studies. The most important tools for phylogenetic comparative analysis were introduced in the late 1980s and early 1990s (Felsenstein, 1985; Grafen, 1989; Harvey & Pagel, 1991; Pagel, 1994; 1999).

Phylogenetic comparative methods allow to answer questions that are simply not possible within the framework of more traditional statistical inference techniques (e.g., simple correlations or contingency tables). These methods allow, among other things, to control for phylogenetic autocorrelation in the comparative analysis, to reconstruct ancestral character states, and to detect character coevolution.

Phylogenetic comparative methods have been regularly used in primatology and evolutionary anthropology since the 1980s (Nunn, 2011). These studies include, e.g., identifying morphological ecological and behavioral determinants of body-size sexual dimorphisms in primates (Cheverud et al., 1985), investigating evolution of social organization in primates (Di Fiore & Rendall, 1994), explaining the occurrence of social monogamy in primates as an adaptation to higher infanticide risk (Opie et al., 2013), or reconstructing behavior of the chimpanzee-human last common ancestor (Duda & Zrzavý, 2013). Later, phylogenetic comparative methods became used by historical linguistics, cultural anthropology and archeology to reconstruct human population history and to tackle the questions regarding the evolution of cultural diversity within our species (Mace & Holden, 2005; Lipo et al., 2006; Mendoza Straffon, 2016) (see “Cultural phylogenetics”).

3. Human population history

During the last 50 years, the development of molecular phylogenetics and statistical inference techniques have revolutionized evolutionary biology. A gradually increasing amount of molecular sequence data has produced enormous datasets that can only be investigated using computational techniques. These techniques have been applied to genetic data and, more recently, to linguistic data in order to reconstruct of human population history.

In 1988, long before the first phylogenies of human populations based on formal phylogenetic analysis of linguistic data appeared, Cavalli-Sforza et al. (1988) published a paper directly comparing a phylogenetic tree of human populations and dendrogram of language families. Although the dendrogram was based on highly controversial linguistic classification on the level of language families and macrofamilies (Ruhlen, 1991), the paper has become very influential. It highlighted the similarities between the processes of historical inference in biology and linguistics, as well as the potential importance of linguistic, not only genetic data, for inferences about human population history. In the wake of this paper, there has been a proliferation of studies attempting to test hypotheses about human population history and coevolution between genes and languages.

3.1 Phylogenies based on genetic data

Constructing phylogenies based on patterns of molecular genetic variation in human population has proved to be a powerful approach to learn about the origins and evolutionary history of our species (Veeramah & Hammer, 2014). Population-genetic research of human evolutionary history began in the 1970s with the analyses of so called "classical markers" such as ABO blood groups and protein allomorphs. The culmination of this research has been an extensive monograph *The History and Geography of Human Genes* by Cavalli-Sforza, Paolo Menozzi, and Alberto Piazza (Cavalli-Sforza et al., 1994). With the advent of automated sequencing technology and the Polymerase chain reaction (PCR) technique, the era of research of uniparental markers has begun. Uniparental markers are parts of a genome that are passed to the offspring via a single parent; they include mitochondrial DNA (mtDNA) and the non-recombining portion of the Y chromosome (NRY). MtDNA is a circular piece of non-recombining DNA of ~16,000 bp found in the mitochondrion and inherited exclusively from the mother. NRY is the middle ~95% of the Y chromosome passed from father to son that does not undergo recombination during meiosis, thereby allowing genetic ancestry to be traced exclusively down the paternal line. The research of uniparental markers gained wide attention with the demonstration in 1987 of an African root of the human mtDNA tree, so called "Mitochondrial Eve" (Cann et al., 1987). The research of human mtDNA was later popularized in a Bryan Sykes's book *The Seven Daughters of Eve* (Sykes, 2001). The results based on mtDNA have soon been complemented by the results based on NRY. Studies based on the analysis uniparentally inherited markers are still being published, and new mtDNA and Y-chromosomal haplogroups are being regularly described and incorporated into increasingly better-resolved human mtDNA and NRY phylogenies (van Oven & Kayser, 2009; van Oven et al., 2014). It is worth noting, however, that the information based on these markers provide only a limited “snapshot” of the genealogical information in the human genome. Often, these markers are informative of sex-specific demographic processes driven by socio-cultural processes rather than of the history of the whole populations, due to their uni-parental inheritance (Wood et al., 2005; Berniell-Lee et al., 2009).

For example, the present-day Bantu-speaking populations of the Congo Basin share ancient mtDNA haplogroups with pygmy hunter-gatherers and even South African Khoisans, suggesting their autochthonous origin, while according to Y-chromosomal haplogroups, they are descendants of a population dispersal that began around 3,000–5,000 years ago in the region of the border between Nigeria and Cameroon. This discrepancy is a consequence of sex-biased admixture between Bantu-speaking farmers and pygmy hunter-gatherers. While Bantu men regularly marry pygmy women, the marriages between pygmy men and Bantu women are prevented by a set of socio-cultural taboos (Berniell-Lee et al., 2009; Verdu et al., 2009; Duda, 2015).

Population history of Icelanders is another good example. Icelandic Y-chromosomal haplogroups come from Norwegian and other Scandinavian stock, reflecting that Icelanders are the descendants of Viking explorers who settle the island in the 9th century. However, the substantial proportion of Icelandic mtDNA genepool comes from the British Isles where Vikings used to stop off on their way to Iceland to pick up the supplies and some women (Helgason et al., 2000). Later, in the 11th century, Vikings made an incursion into North America (Newfoundland and Labrador) and brought “trophy wifes” from there as evident from the presence of rare mtDNA haplogroup C1e in the genomes of present-day Icelanders (Ebenesersdottir et al., 2011).

In the early 2000s, the development of DNA hybridization microarray technology led to the first studies of human population history based on genome-wide data. These data included short tandem repeats (STRs) or microsatellites and single-nucleotide polymorphisms (SNPs). At around the same time, the first by on-line archiving molecular datasets, such as the Human Genome Diversity Project (HGDP) panel (H. M. Cann et al., 2002) or the International HapMap Project (Gibbs et al., 2003) were published. With the advent of massively parallel short-read (second-generation) sequencing and recently third-generation sequencing technologies, the whole-genome data become regularly used and they are now becoming available for an increasing number of human populations.

Notably, the field of human population history have been greatly affected by the increasing availability of ancient DNA (aDNA), from mtDNA fragments to whole genomes (Pickrell & Reich, 2014). The inclusion of aDNA samples into genomic analyses allowed to make inferences about human population history that would be impossible without this source of information. These analyses detected an archaic introgression from Neanderthal to modern human genome that contributed to the genetic ancestry of present-day non-Africans (Green et al., 2010), and discovered that a previously unknown Nenaderthal-related hominin population that was living in the Altai-Sayan region that has contributed to genetic ancestry of present-day aboriginal people from New Guinea, Australia, and the Philippines (Krause et al., 2010; Reich et al., 2010). Paleogenomic analyses have elucidated the ancestry of Native American people (Raghavan et al., 2014) and confirmed that there were at least two independent migrations from Beringia to North America (Rasmussen et al., 2010, 2014). It also confirmed some theories about ancient migrations based on linguistic evidence. For example, the analysis including aDNA helped to ascertain the role of Taiwan as a source for the Austronesian expansion (Ko et al., 2014), or to support the “steppe hypothesis” of the Indo-European origins (Haak et al., 2015) (see “Phylogenies based on linguistic data”).

As the quality and quantity of population genetic data improved, so did the quality of phylogenetic inference methods used. The early trees of human populations were based on distance-based algorithms such as UPGMA and neighbor-joining (Nei & Roychoudhury, 1993). In the late 1990s and early 2000s, distance-based methods were steadily replaced by character-based methods of phylogenetic tree inference, especially maximum-likelihood and Bayesian analysis. However, distance-based methods are being used to date and some of the largest trees of human populations based on genome-wide data were produced by neighbor-joining analysis (Tishkoff et al., 2009; Pemberton, 2013; Mallick et al., 2016).

Human population history is certainly not purely tree-like, as genetic admixture, mediated by processes such as migrations, expansions, intermarriage, trade, or slavery, have played an important role in history (Hellenthal et al., 2014; Busby et al., 2015). Recently developed phylogenetic methods allowed to visualize evolutionary history of populations using a bifurcating tree with horizontal links (“admixture edges”), accounting for both population splits and mixtures (Pickrell & Pritchard, 2012).

In addition to phylogenetic analyses that produce trees or reticulated networks, clustering methods producing so called “admixture plots” were applied to genetic data to infer human population history since the early 2000s (Rosenberg et al., 2002). Admixture plots are graphical outputs of programs STRUCTURE (Pritchard et al., 2000), FRAPPE (Tang et al., 2005), or ADMIXTURE (Alexander et al., 2009), developed to estimate individual ancestry and population structure on the basis of recombining genetic markers (i.e., SNPs, STRs, sequence haplotypes, etc.). These programs use iterative maximum-likelihood or Bayesian clustering algorithms that attribute individual genotypes of known ethnic population origin to K clusters such that Hardy-Weinberg equilibrium is maximized within clusters. Individuals are given a membership coefficient for each cluster such that the estimated membership coefficient of each individual sum to 1 across K clusters. The analysis itself assumes no grouping of information, and the individuals are arrayed by the population of origin only after the analysis. The graphical outputs of these programs are plots indicating the proportions of individual genotypes attributable to K clusters by color. The order of appearance of clusters reflects a composite of individuals’ ancestry proportions and ancestry-specific allele frequencies.

3.2 Phylogenies based on linguistic data

Linguistic data have a strong historical signal for at least two reasons. Firstly, language is a neutral trait, i.e., the words are arbitrary sounds that should have no fitness implications (Mace & Jordan, 2011, but see Blasi et al., 2016). Secondly, language, especially its basic vocabulary, is a very conservative trait and fits the idea of “cultural core”, i.e., there are strong pressures, so called conformist bias, that maintain words in distinct and consistent forms. The fidelity of cultural transmission of some words can rival that of genes (Pagel, 2009). This is supported by the fact that genetic and linguistic trees are often similar, reflecting the same underlying process of population fission and migration (Cavalli-Sforza et al., 1988; Cavalli-Sforza et al., 1992;

Penny et al.,1993; Balanovsky et al., 2011; Karafet et al., 2015; Longobardi et al., 2015), and by spatial correlation of genetic variants and linguistic groups (Barbujani & Sokal, 1990; ; Novembre et al., 2008; Jay et al., 2011). Language differences often function as barriers to gene flow and thus can shape genetic diversity (Belle & Barbujani, 2007).

The idea that language similarity can be used to trace ancient human migration dates back to 18th century, to William Jones who proposed the existence of a relationship between Indo-European languages and to Thomas Jefferson who speculated about a historical connection between languages of East Siberia and North America. In the 19th century, the first language family trees by František Ladislav Čelakovský and August Schleicher predate the first trees of biological species by Charles Darwin and Ernst Haeckel by several years (List et al., 2016; Duda, *in press*).

In the 1950s, Morris Swadesh developed formal, quantitative approaches to comparative linguistics called lexicostatistics and glottochronology. They allowed to infer the relationships between languages from a distance matrix based on percentages of shared basic vocabulary items and to estimate the time since two or more languages diverged from a common ancestor. However, there are serious theoretical and methodological problems with lexicostatistics and glottochronology. The distance-based clustering algorithm causes a loss of information and tend to group languages that evolved in a same rate rather than languages that share a recent common ancestor. The assumption of the constant rate of language change is inadequate, making the inferred time estimates unrealistic. For these and other reasons, lexicostatistics and glottochronology were heavily criticized and are now largely discredited (Atkinson & Gray, 2005).

In recent years, historical linguists have increasingly utilized phylogenetic approaches to reconstruct language histories and, by inference, population histories using character-based phylogenetic methods and lexical data (and especially basic vocabulary). Character-based methods of phylogenetic analysis have provided principled solutions to many of the problems that plagued lexicostatistics and glottochronology.

The first tree based on phylogenetic analysis of linguistic data was published in 2000 (Gray & Jordan, 2000) and was based on maximum parsimony analysis of 77 Austronesian languages and 5,185 lexical items. Its topology supported the *Express Train model* of Austronesian expansion popularized by Jared Diamond (Diamond, 1988). In 2002, the first phylogenetic tree of Bantu languages was published (Holden, 2002) based on 92 basic vocabulary items and 72 Bantu and Bantoid languages. In the next year, the first phylogenetic tree for Indo-European languages was published by Rexová et al. (2003). This tree was based on parsimony analysis of 84 languages and 200 vocabulary items of the Swadesh wordlist. Several months later, the first time-calibrated phylogeny of Indo-European languages by Gray and Atkinson (Gray & Atkinson, 2003) was published, based on Bayesian analysis of 87 Indo-European languages and 200 vocabulary items. Gray and Atkinson estimated the age of the root of the Indo-European tree between 7,800 and 9,800 years BP. This time depth supports the Anatolian theory of Indo-European origin, originally proposed by Colin Renfrew (Renfrew 1990). However, a recently published tree of Indo-European languages that uses slightly different method of time-calibration (Chang et al., 2015) supports the alternative “steppe hypothesis”, originally proposed by Marija Gimbutas (Gimbutas, 1973), in agreement with recent paleogenomic analysis (Haak et al., 2015).

Today, phylogenies of languages and dialects based on lexical data are available for the following language families and groups of languages: Austronesian (Gray & Jordan, 2000; Gray et al., 2009), Malagasy (Serva, 2012), Bantu (Holden, 2002; Holden & Gray, 2006; Rexová et al., 2003; 2006; Currie et al., 2013), Kikongo (De Schryver et al., 2015), Indo-European (Gray & Atkinson, 2003; Rexová et al., 2003; Atkinson & Gray, 2006; Bouckaert et al., 2012; Chang et al., 2015; Serva & Petroni, 2008), Slavonic (Nurbakova et al., 2013), Semitic (Kitchen et al., 2009), Turkic (Hruschka et al., 2015), North Caucasian (Balanovsky et al., 2011), East Caucasian (Karafet et al., 2015), Uralic (Honkola et al., 2013), Pama-Nyungan (Bowern & Atkinson, 2012; Malaspinas et al., 2016), Tasmanian (Bowern, 2012), Aslian (Dunn et al., 2013), Japonic (Lee & Hasegawa, 2011), Ainu (Lee & Hasegawa, 2013), Dene-Yeniseian (Sicoli & Holton, 2014), Pacific Coast Athabaskan (Spence, 2013), Uto-Aztecan (Levinson et al., 2011; Wheeler & Whiteley, 2015), Arawak (Walker & Ribeiro, 2011), Tupi-Guarani (Michael et al., 2015), Tupian (Walker et al., 2012; Galucio et al., 2015), and Tucanoan (Chacon & List, 2015).

The vast majority of published studies has concerned a single language family due to the shallow time depth (approximately 10,000 years) available for traditional methods based on the analysis of lexical data, caused by a relatively fast process of lexical change (Gray, 2005). Some researchers assume that the methods based on the analysis of structural elements of languages (e.g., phonological and morphosyntactic features, and word order) or highly conserved, most frequently used words in the basic vocabulary (e.g., special adverbs, pronouns, and numerals) will push the time depth of language phylogenies well beyond 10,000 years limit and elucidate the relationships between language families (Gray, 2005; Pagel et al., 2007; 2013). However, phylogenetic trees that include languages of more than one language family based on lexical data or structural data are still very rare. They include trees of Austronesian and non-Austronesian (Papuan) languages of Island Melanesia (Dunn et al., 2005, 2009; Hunley et al., 2008), a tree of language families of Europe (Longobardi et al., 2015), and trees of language families that constitute the proposed Eurasiatic macrofamily (Pagel et al., 2013; Jäger, 2015).

Bayesian clustering methods producing admixture plots that were originally developed to estimate population structure based on genetic data, were also applied to linguistic data, to infer the relationships within and between groups of languages where tree building methods might prove inadequate. These methods were applied to languages of Sahul (Reesink et al., 2009) and Tasmania (Bowern, 2012).

Currently, our ability to infer distant language relationships using a formal phylogenetic analysis of linguistic data is limited (Gell-Mann et al., 2009). While many language families based on comparative method are well-established, the proposed language macrofamilies, based on “mass comparison” or “multilateral comparison” methods devised by Joseph H. Greenberg (Greenberg, 1954), are all considered controversial. Greenberg’s classification of the languages of Africa (Greenberg, 1963) was relatively well-received, but other linguistic macrofamilies such as Eurasiatic/Nostratic (Bomhard & Kerns, 1994; Greenberg, 2000, 2002), Indo-Pacific (Greenberg, 1971), and Amerind (Greenberg, 1987) are highly contentious (Jäger, 2015).

4. Cultural phylogenetics

Following the work of Mace and Pagel (Mace & Pagel, 1994), some researchers have adopted conceptual and methodological tools developed in evolutionary biology to investigate a diverse range of topics of anthropology, from reconstructing ancient migrations to investigating how human populations have adapted to their environments.

Within cultural phylogenetics, some researchers focus on proximate mechanisms underlying cultural change in order to determine how tree-like cultural evolution is. This research include both case studies (Guglielmino et al., 1995; Hewlett et al., 2002; Collard et al., 2006) and simulation studies (Greenhill et al., 2009; Currie et al., 2010; Nunn et al., 2010). Others examine how mode of transmission differ for various kinds of cultural traits (Guglielmino et al., 1995; Hewlett et al., 2002) or how mode of cultural transmission depends on socio-political settings (Tehrani & Collard, 2002). Others examine how different amounts of horizontal transmission obscure cultural phylogenies (Greenhill et al., 2009; Currie et al., 2010).

Among the multitude of studies that apply phylogenetic methods to languages, cultural artifacts and other aspects of cultural variation, two subfields stand out in particular: “building phylogenies” and “using phylogenies” (Mace & Holden, 2005; Mace & Jordan, 2011).

4.1 Building phylogenies

The first subfield – building phylogenies – applies methods of phylogenetic analysis to infer historical relationships of languages, cultural artifacts or other cultural phenomena, e.g., oral and musical traditions. The authors operate under the assumption that the observed similarities among cultural traits are a function of common ancestry and that the artifacts’ “phenotypes” are a result of a process of cultural selection upon them. Constructing genealogies of cultural artifacts has a long tradition, beginning in the 19th century (Rivero, 2016; Duda, *in press).* This approach has recently experienced its renaissance thanks to work of historical linguists, cultural anthropologists and archeologists who adopted phylogenetic methods to investigate whether cultural groups are related in a tree-like fashion and what are the relationships between them (Mace et al., 2005; Lipo et al., 2006; Mendoza Straffon, 2016).

The tree building methods can help us reconcile the long-standing debates over the relative importance of branching and blending processes (i.e., vertical and horizontal transmission) in cultural macroevolution. It is possible to test how cultural datasets fit the bifurcating tree model by quantifying the amount of homoplasy and synapomorphy in these datasets using, e.g., consistency and retention index (Collard et al., 2006). Support for individual nodes on the inferred tree can be tested using bootstrap or jackknife analysis in the tree based on maximum parsimony analysis, or, in the case of Bayesian inference, by estimating posterior probabilities of each node. Often, the data fit more than one tree equally, with ambiguous relationships arising from parallel evolution or horizontal transmission (e.g., from lexical borrowing). Networks, unlike trees, allow us represent more than one evolutionary pathway in a single graph, by allowing branches to join as well as diverge. These techniques can detect and visualize conflicts in cultural datasets (e.g., NeighborNet; Tehrani, 2013) or the instances of horizontal transmission (e.g., Galled networks; Willems et al., 2016).

The tree building methods allow us to investigate whether the branching pattern of the culture-based tree reflects the history of the populations based on genes. Some studies have used methods of tree building and tree comparison methods (e.g., tanglegrams; Duda & Zrzavý, 2016) to investigate coevolution of cultural traits or gene-culture coevolution as an analogy of the host-parasite cospeciation (Riede, 2009). The direct comparison of genetic and linguistic phylogenies is a useful way to investigate gene-culture coevolution on various geographic and taxonomic scales (Hunley et al., 2007; Karafet et al., 2015; Duda & Zrzavý, 2016).

To date, tree building methods have been applied to a wide range of linguistic groups (see "Phylogenies based on linguistic data") and a wide range of cultural phenomena. The cultural artifacts studied include arrowheads (Buchanan & Collard, 2007; O'Brien et al., 2014), textiles (Tehrani & Collard, 2002; Buckley, 2012), basketry (Jordan & Shennan, 2003), pottery (Cochrane & Lipo, 2010), architecture (Jordan & O'Neill, 2010), watercrafts (Rogers et al., 2009), statues (Marwick, 2012; Tripp, 2016), storage containers (O'Neill et al., 2013), engraved stone plaques (Rivero & O'Brien, 2014), cutlery (Riede, 2009), and archeobotanical assemblages (Coward et al., 2008). Besides the languages, analyses of non-material culture include folktales (Ross et al., 2013; Tehrani, 2013), music (Windram et al., 2008; Le Bomin et al., 2016), and manuscript versions (Barbrook et al., 1998; Windram et al., 2008).

4.2 Using phylogenies

The second subfield – using phylogenies – applies methods of phylogenetic comparative analysis to test hypotheses about adaptation through cross-cultural comparison and to reconstruct evolutionary history of cultural phenomena. These studies use phylogenies either to control for phylogenetic autocorrelation in cross-cultural analyses or to directly optimize cultural evolution and adaptation to a phylogenetic tree.

As stated above, one cannot make valid inferences regarding adaptive evolution (both biological and cultural) without controlling for shared ancestry. Efforts have been made to reduce phylogenetic non-independence (Galton's problem) in cross-cultural datasets, most notably by creating the SCCS. Nonetheless, there is still serious non-independence among the populations in the SCCS (Eff, 2004; Dow & Eff, 2008), leading to high false positive rates (type I errors) in cross-cultural analyses. Phylogenetic comparative methods, e.g., phylogenetic independent contrast (Felsenstein, 1985) or phylogenetic generalized least squares (Grafen, 1989), take the non-independence of populations into account. The distribution of cultural traits that confer a selective advantage (e.g., farming) is likely to reflect both phylogeny and adaptive pressures within a particular environment. The application of phylogenetic comparative methods is all the more necessary as empirical studies show that adaptive cultural traits often cover a strong phylogenetic signal, presumably because these traits are transmitted predominantly from parents to offspring, ensuring it will have higher reproductive success (Guglielmino et al., 1995; Hewlett et al., 2002).

Phylogenetic cross-cultural analyses published to date investigated the association between cattle keeping and lactose digestion capacity in adults (Holden & Mace, 1997), the mode of subsistence and fertility (Sellen & Mace, 1997), sexual dimorphism in stature and sexual division of labor in the world’s populations (Holden & Mace, 1999), reciprocity and food sharing in human hunter-gatherers and non-human primates (Jaeggi & Gurven, 2013), men’s social status and reproductive success in nonindustrial societies (von Rueden & Jaeggi, 2016), or various cultural and environmental traits and island deforestation in Austronesian-speaking populations (Atkinson et al., 2016).

These phylogenetic cross-cultural analyses use phylogenies only to remove any confounding effects of shared ancestry that may lurk behind any functional relationships between traits (Pagel & Harvey, 1988). Only after removing or controlling for the effect of phylogeny, the true functional association between traits can be uncovered. However, there are methods that use phylogeny to investigate cultural evolution and adaptation without discarding it first. Character states based on cross-cultural data are mapped onto the tree topology, and their ancestral states and probable pattern of historical change along the branches of the tree are inferred (using maximum-parsimony, maximum-likelihood, or Bayesian optimization). These methods allow to test for correlated evolution in cultural traits, e.g., descent rules and pastoralism in Bantu-speaking populations in sub-Saharan Africa (Holden & Mace, 2003), marriage practices in Indo-European-speaking populations (Fortunato et al., 2006), or traits of religion (e.g., belief in afterlife and shamanism) in hunter-gatherers (Peoples et al., 2016). These methods avoid Galton’s problem because the units of analysis are not populations, but instances of evolutionary change (Mace & Holden, 2005; Mace & Jordan, 2011).

These types of studies include investigating the coevolution of cattle-keeping and descent rules (matrilineality and patrilineality) (Holden & Mace, 2003), descent rules and residence rules (matrilocality and patrilocality) (Opie et al., 2014), and descent rules and kinship terminology (Guillon & Mace, 2016) in Bantu-speaking populations, evolution of wealth transactions at marriage (bridewealth and dowry) (Fortunato et al., 2006), evolution of marriage practices (monogamy and polygyny) and post-marital residence rules (neo-, uxori-, and virilocality) (Fortunato, 2011a, b), and evolutionary history of folktales (da Silva & Tehrani, 2016) in Indo-European speaking societies, or evolution of post-marital residence rules (matrilocality and patrilocality) in Indo-European and Austronesian speaking societies (Jordan et al., 2009; Fortunato & Jordan, 2010), evolution of socio-political complexity in Bantu and Austronesian-speaking populations (Currie et al. 2010; Currie & Mace, 2011; Walker & Hamilton, 2011), evolution of land tenure norms (Kushnick et al., 2014), and coevolution of religious beliefs and socio-political complexity in Austronesian-speaking populations (Watts et al., 2015), coevolution of conception beliefs (partible paternity) and post-marital residence rules among indigenous societies of lowland South America (Walker et al., 2010), evolutionary history of marriage practices (courtship and arranged marriages) in hunter-gatherer societies in Africa, Eurasia and Oceania (Walker et al., 2011), and evolutionary history of religious beliefs and practices in the global sample of hunter-gatherers (Peoples et al., 2016).

5. Phylogenies as a framework for studying evolution of culture

The largest global human population-level phylogenetic tree published to date (Pemberton et al., 2013) is a neighbor-joining tree based on eight global and regional molecular datasets of autosomal STRs. It includes 267 human populations. This population sample is not sufficient, considering there are approximately 7,100 human ethno-linguistic groups living in the world today (Lewis et al., 2016). Notably, population sample used by Pemberton et al. (2013) includes recently aggregated, mixed ancestry populations (e.g., African Americans and Latin American Mestizos) while lacking populations that are essential for understanding human population history on a global scale (e.g., southern South African Khoisans, Omotic speakers, Berber, Vedda, Andamanese, Australian and Tasmanian Aboriginals, Malagasy, Ainu, populations of East Siberia, Amerind speakers of North America, Na-Dene speakers) and those featured in cross-cultural datasets (e.g., the SCCS). Large regions of the world are underrepresented or entirely missing in the sample (South Africa, North Africa and Near East, Europe, North Asia, Southeast Asia and Oceania, Australia, and North America).

Given the insufficient population samples and limited overlap between published genetic trees and cross-cultural datasets, the utility of the published genetic trees for investigating human population history and evolution of culture using phylogenetic comparative methods is limited. Another problem is that genetic phylogenies are sometimes poorly resolved, hampering the use of some of some phylogenetic comparative methods.

Unlike genetic phylogenies, linguistic phylogenies are comprehensive (up to 542 language varieties; Currie et al., 2013) and well resolved. Some cultural anthropologists believe that linguistic phylogenies might be preferable to phylogenies based on genetic data as models of human population history because languages track the inheritance of culture, and it is this inheritance that is relevant to questions of human cultural evolution. Languages typically evolve at a higher rate than genes, meaning that they can capture population diversification that occurs at shorter time scales (Pagel, 2009; Grollemund et al., 2015). Some anthropologists also assume that genetic phylogenies might be more easily obscured by genetic admixture than linguistic phylogenies might be obscured by horizontal cultural transmission (i.e., linguistic borrowings and language shifts) (Mace & Jordan, 2011).

The problem is, however, that most of the published linguistic phylogenies are geographically and taxonomically restricted, covering just one language family each (see “Phylogenies based on linguistic data”). The scope of phylogenetic comparative analyses and the “depth” of the phylogenetic reconstructions of ancestral states are therefore limited (Heyer, 2012).

Another problem is the accuracy of linguistic phylogenies as model for population history. The empirical studies show that evolutionary processes are generally more complex than simple models of gene–language coevolution predicted, and linguistic boundaries do not always function as barriers to gene flow (Steele & Kandler, 2010; Pakendorf, 2014). Moreover, the language evolution is tree-like to different extent in different parts of the world (Gray et al., 2010). Indeed, evolution of human languages is not identical to the evolution of peoples themselves. For example, Central European Hungarians form a genetic continuum with the neighboring Indo-European-speaking populations while speaking a Uralic (Finno-Ugric) language related to the Khanty and Mansi languages spoken in the vicinity of the Ural Mountains (Honkola et al., 2013; Longobardi et al., 2015). Genetic and linguistic data sometimes imply different historical scenarios and reflect different time scales (Steele & Kandler, 2010). This problem might affect results of the phylogenetic comparative analyses that use language phylogenies as a framework to study evolution of culture.

Consider the case of Central African pygmies, e.g., Aka pygmies of Congo and Central African Republic. Aka, like most Central African pygmies, are forest-dwelling hunter-gatherers. They speak Nigero-Congo (Northwest Bantu) languages (Lewis et al., 2016). Linguistically, Aka are closely related to the neighboring Bantu-speaking farmers and by the inference from linguistic phylogeny, their hunter-gatherer lifestyle and the associated cultural traits would be interpreted as an independently derived adaptation. However, genetic evidence strongly suggest that Aka, like other Western pygmies, are more closely related to Eastern pygmies and to other hunter-gatherer groups of East and South Africa, and that they have adopted Bantu languages relatively recently, following Bantu expansion some 3,000 years ago (Tishkoff et al., 2009; Bahuchet, 2012; Duda, 2015; Duda & Zrzavý, 2016). The pygmies with their unique hunter-gatherer lifestyle thus cannot be interpreted as evolving several times independently from Bantu-speaking farmers as some theories have suggested (Blench, 1999).

"Negritos" of Philippines are another good example of recent language shift occurring in a culturally distinct group. Watts et al. (2015) used phylogeny of Austronesian languages to reconstruct coevolution of religion (belief in supernatural punishment and moral high gods) and socio-political complexity and test for coevolution between them using phylogenetic tree of Austronesian languages as a framework. This tree naturally includes Agta, the negritos of Philippines, who speak Austronesian (Malayo-Polynesian) language. They are, however, genetically more related to other negrito groups and to aboriginal Papuans and Australians (Rasmussen et al., 2011; Pugach et al., 2013; Duda, 2016). Although negritos have experienced extensive Asian-related admixture following Austronesian expansion (Abdulla et al., 2009), they are descendants of an earlier, unrelated settlement wave and have adopted Austronesian languages relatively recently (Reid, 2013). The Agta peoples' belief in supernatural punishment is shared with several negrito groups from the Philippines, Malaysia and even Andaman Islands who speak unrelated languages (Blust, 2013). These beliefs likely predate the origin of Austronesian language family. The belief in supernatural punishment in Agta thus leads to the misinterpretation of the pattern of cultural evolution (what has been interpreted as an independent gain of cultural trait was in fact a language shift).

6. Supertree approach

Numerous genetic and linguistic phylogenetic trees of human populations were published but little attention has been paid to formal phylogenetic synthesis. Given the current state of the field, a possible strategy is to focus on published ("source") trees, adopting the "supertree approach" (or "taxonomic congruence approach"; Pisani & Wilkinson, 2002). Supertree methods are a part of a toolkit of phylogenetic comparative methods (Bininda-Emonds, 2014). The principle of all supertree methods is combine many smaller, even if only partially overlapping source trees to create a single, more comprehensive supertree (Bininda-Emonds, 2004). Supertrees can summarize relationship patterns from multiple independent tree topologies based on different kinds of character data (e.g., genetic and linguistic data) and obtained using different phylogenetic methods (e.g., distance-based and character-based) that can otherwise not be combined. The important feature of a supertree approach is its ability to evaluate pieces of competing evidence to identify topological conflicts caused by incongruence of data partitions or by data deficiency. Supertree methods are excellent tools for summarizing the existing phylogenetic hypotheses, based on different types of data, in a form of "synthetic consensus" phylogeny (Ruta et al., 2007). They have been used to provide some of the largest, most comprehensive phylogenies for diverse extant and extinct groups at various taxonomic levels, e.g., for mammalian species (Bininda-Emonds et al., 2007), dinosaur genera (Pisani et al., 2002) or hexapod orders (Davis et al., 2010). Supertrees have also been used to address various questions concerning evolution and adaptation, including reconstructing ancestral character states (Espinoza et al., 2004), detecting character coevolution on phylogenies (Lukas & Clutton-Brock, 2012), quantifying evolutionary rate (Clarke et al., 2016), detecting diversification rate shifts (Stadler, 2011; Bronzati et al., 2015), or predicting extinction risk (Cardillo et al., 2008).

The first method of supertree construction devised was the "matrix representation with parsimony" (MRP) method (Baum, 1992; Ragan, 1992). Numerous supertree methods have been developed over the years (e.g., Matrix Representation with Compatibility, Matrix Representation with Flipping, Average consensus Supertree, Strict Consensus Merger, MinCut Supertree, MinFlip Supertree, PhySIC_IST, Robinson–Foulds Supertree, Subtree Prune and Regraft Supertree, MultiLevel Supertree, Quartet Supertree, SuperFine+MRP, and SuperFine+MRL, or PhySIC and PhySIC_IST) but the MRP remains by far the most commonly used method of supertree construction (Bininda-Emonds, 2014). In MRP, each source tree is converted into a matrix of additive binary characters. After standardization of taxonomic nomenclature and taxonomic level across the source trees, the matrices are merged (completely or partially, depending on the degree of taxon overlap), and processed using an optimality criterion of maximum parsimony. The resulting most parsimonious tree(s) are presented in a form of strict or semi-strict consensus trees. MRP is the most tractable approach for medium to large data sets (Cotton et al., 2006). Experimental studies evaluating MRP in comparison with other supertree methods have established that for large datasets MRP generally produces trees of equal or greater accuracy than other methods (Buerki et al., 2011; Nguyen et al., 2012).

7. Aims and scope of the thesis

A robust phylogenetic tree of human populations is a necessary prerequisite for such investigations. Today, no unified picture of human population history is available, as studies that infer human population history have used different types (and conceptual classes) of genetic and linguistic markers. The available phylogenetic trees of human populations capture only a snapshot of human ethnic diversity, and the overlap between these trees and cross-cultural datasets is limited. For these reasons, phylogenetic cross-cultural analyses and phylogenetic comparative studies of cultural evolution have been limited to small samples of populations for which genetic data were available (Walker et al., 2011; Jaeggi & Gurven, 2013), or to language families such as Bantu, Indo-European, and Austronesian, with well-resolved phylogenies based on lexical data (Holden & Mace, 2003; Fortunato et al., 2006; Jordan et al., 2009).

Linguistic data seems to be unable to provide a global tree of human populations due to a limited timescale over which linguistic inference is possible (Gray, 2005; Pagel, 2009). On the other hand, genetic phylogenies, although global, might be unsuitable for studies of cultural evolution, as the population history they inform of might be older than the cultural traits under investigation (Grollemund et al., 2015). A supertree incorporating all temporal "strata" of human evolution is necessary to elevate the studies of cultural evolution to a global level (Duda & Zrzavý, 2016).

The ultimate goal of this effort is to "put pre-history back into anthropology" (Mace & Jordan, 2011). We can now approach the theories of evolutionist anthropologists once again, this time as testable hypothesis. We can indeed test whether culture, "race", and language are wholly independent or whether there is a congruence between genetic and cultural history of populations (Duda & Zrzavý, 2016). We can test whether culture evolves in a tree-like fashion, and whether this applies to a broad range of cultural traits, different environments, and different socio-political settings (Tehrani & Collard, 2002; Collard et al., 2006; Bowern et al., 2011; Rivero & O'Brien, 2014). We can also test whether society progressed from "savagery to civilization", i.e., from less complex to more complex forms of socio-political organization, at least in some regions of the World (Currie & Mace, 2011; Walker & Hamilton, 2011), or whether the religion evolved from animism to theism (Peoples et al., 2016). Within the phylogenetic framework, it is possible to ask – and sometimes answer – many of the old questions explicitly and with a new level of precision.

In **Chapter I** we inferred a composite phylogeny of human populations using the MRP supertree method. The supertree is based on 257 genetic, as well as linguistic, phylogenetic trees and 44 admixture plots from 200 published studies (1990–2014). The resulting tree topology is dominated by genetic data and includes the most basal position of South African Khoisan followed by Central African pygmies and by a paraphyletic section of all other sub-Saharan African peoples. The sub-Saharan Africans are basal to the monophyletic superclade consisting of the North African–West Eurasian assemblage and the consistently monophyletic Eastern clade (Sahul–Oceanian, East Asian, and Beringian–American peoples). We investigated robustness of the inferred supertree topology using a set of different analysis parameters. The overall topology of the supertree is surprisingly stable and well-resolved. We identified areas of topological instability and individual unstable taxa. In order to investigate gene-language coevolution on a global scale and to test for monophyly of the proposed linguistic groupings (language families and macrofamilies), we optimized linguistic data onto the topology of the supertree and used these data to constrain the topology of the supertree. Linguistic data fit rather poorly on the genetic supertree topology, supporting a view that direct coevolution between genes and languages is far from universal. Most of the controversial language macro-families were not supported by the supertree topology.

In **Chapter II** we used time-calibrated supertree in combination with the ethnographic record to reconstruct evolution of religious beliefs and behaviors in early modern humans. We used a global sample of hunter-gatherers and seven traits describing their religiosity: animism, belief in an afterlife, shamanism, ancestor worship, high gods, and worship of ancestors or high gods who are active in human affairs. We reconstructed ancestral character states and tested for correlated evolution between the characters and for the direction of cultural change. Our results indicated that the oldest trait of religion, present in the most recent common ancestor of present-day hunter-gatherers, was animism, in agreement with long-standing beliefs about the fundamental role of this trait. Belief in an afterlife emerged later, followed by shamanism and ancestor worship. Ancestor spirits or high gods who are active in human affairs were absent in early humans, indicating a deep history for the egalitarian nature of hunter-gatherer societies. There is a significant positive relationship between most characters investigated, but the trait "high gods" stands apart, suggesting that belief in a single creator deity can emerge in a society regardless of other aspects of its religion.

Chapter III describes evolutionary history of Central African pygmies. Pygmies are remarkable for their basal position on phylogenetic tree of human populations and for their hunter-gatherer lifestyle and the associated morphological, life-history, and cultural adaptations. This chapter discusses phylogenetic relationships of Central African pygmies to other populations living in sub-Saharan Africa based on genetic and linguistic evidence. It describes the effect of Bantu expansion on genetic and linguistic diversity of these hunter-gatherer groups. It also highlights cultural links between groups of pygmies in Central Africa who speak unrelated languages and between pygmies and other hunter-gatherer groups living in South and East Africa. The chapter briefly discusses evolutionary history of short-statured populations living in Southeast Asia (Andaman Islands, Malaysia, and the Philippines) and the controversies surrounding *Homo floresiensis*, a human species remarkable for its small body size and small cranial capacity, discovered in 2003 at Liang Bua cave in Flores (Indonesia). Current hypotheses about the evolution of the pygmy phenotype are briefly discussed.

Chapter IV discusses phylogenetic approaches to linguistic and cultural evolution and how can these approaches facilitate our understanding of human prehistory and evolution of cultural adaptations. Cultural phylogenetics is a scientific discipline that studies human culture using phylogenetic methods developed by evolutionary biology to reconstruct phylogenetic relationships between species or populations of organisms and to test hypotheses about evolution of biological traits. When applied to human culture, these methods can help us answer some of the fundamental questions of historical linguistics, cultural anthropology and archeology. They can reconstruct ancient human migrations, pinpoint the age of cultural phenomena, or test hypothesis about evolution of cultural traits. This chapter describe the beginnings of "tree thinking" in biological and social sciences and parallel developments in evolutionary biology and historical linguistics, cultural anthropology and archeology during the 19th and 20th century. It compares the processes of biological and cultural evolution and discusses the methods that can assess how tree-like is cultural evolution. It describes how phylogenetic methods can be applied to material and non-material culture, especially the language, and how phylogenies of human populations can be used to test hypotheses about cultural history and diversification and cultural adaptation, using phylogenetic comparative methods.

References

Abdulla, M. A., Ahmed, I., Assawamakin, A., Bhak, J., Brahmachari, S. K., Calacal, G. C., . . . Indian Genome Variation, C. (2009). Mapping human genetic diversity in Asia. *Science, 326*(5959), 1541-1545. doi:10.1126/science.1177074

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research, 19*(9), 1655-1664. doi:10.1101/gr.094052.109

Atkinson, Q. D., Coomber, T., Passmore, S., Greenhill, S. J., & Kushnick, G. (2016). Cultural and environmental predictors of pre-European deforestation on Pacific Islands. *Plos One, 11*(5), 15. doi:10.1371/journal.pone.0156340

Atkinson, Q. D., & Gray, R. D. (2005). Curious parallels and curious connections - Phylogenetic thinking in biology and historical linguistics. *Syst Biol, 54*(4), 513-526. doi:10.1080/10635150590950317

Atkinson, Q. D., & Gray, R. D. (2006). How old is the Indo-European language family? Illumination or more moths to the flame? In P. Forster & C. Renfrew (Eds.), *Phylogenetic Methods and the Prehistory of Languages Cambridge*. UK: The McDonald Institute for Archaeological Research.

Bahuchet, S. (2012). Changing language, remaining pygmy. *Human Biology, 84*(1), 11-43.

Balanovsky, O., Dibirova, K., Dybo, A., Mudrak, O., Frolova, S., Pocheshkhova, E., . . . Genographic, C. (2011). Parallel evolution of genes and languages in the Caucasus region. *Molecular Biology and Evolution, 28*(10), 2905-2920. doi:10.1093/molbev/msr126

Barbrook, A. C., Howe, C. J., Blake, N., & Robinson, P. (1998). The phylogeny of The Canterbury Tales. *Nature, 394*(6696), 839-839. doi:10.1038/29667

Barbujani, G., & Sokal, R. R. (1990). Zones of sharp genetic change in Europe are also linguistic boundaries. *Proceedings of the National Academy of Sciences of the United States of America, 87*(5), 1816-1819. doi:10.1073/pnas.87.5.1816

Baum, B. R. (1992). Combining trees as a way of combining data sets for phylogenetic inference, and the desirability of combining gene trees. *Taxon, 41*(1), 3-10. doi:10.2307/1222480

Belle, E. M. S., & Barbujani, G. (2007). Worldwide analysis of multiple microsatellites: Language diversity has a detectable influence on DNA diversity. *American Journal of Physical Anthropology, 133*(4), 1137-1146. doi:10.1002/ajpa.20622

Berniell-Lee, G., Calafell, F., Bosch, E., Heyer, E., Sica, L., Mouguiama-Daouda, P., . . . Comas, D. (2009). Genetic and demographic implications of the Bantu expansion: Insights from human paternal lineages. *Molecular Biology and Evolution, 26*(7), 1581-1589. doi:10.1093/molbev/msp069

Bininda-Emonds, O. R., Cardillo, M., Jones, K. E., MacPhee, R. D., Beck, R. M., Grenyer, R., . . . Purvis, A. (2007). The delayed rise of present-day mammals. *Nature, 446*(7135), 507-512. doi:10.1038/nature05634

Bininda-Emonds, O. R. P. (2004). The evolution of supertrees. *Trends in Ecology & Evolution, 19*(6), 315-322. doi:10.1016/j.tree.2004.03.015

Bininda-Emonds, O. R. P. (2014). An introduction to supertree construction (and partitioned phylogenetic analyses) with a view toward the distinction between gene trees and species trees. In L. Z. Garamszegi (Ed.), *Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology: Concepts and Practice* (pp. 49-76). Berlin, Heidelberg: Springer Berlin Heidelberg.

Blasi, D. E., Wichmann, S., Hammarström, H., Stadler, P. F., & Christiansen, M. H. (2016). Sound–meaning association biases evidenced across thousands of languages. *Proceedings of the National Academy of Sciences, 113*(39), 10818-10823. doi:10.1073/pnas.1605782113

Blench, R. (1999). Are the African pygmies an ethnographic fiction? In K. Biesbrouck, S. Elders, & G. Rossel (Eds.), *Central African Hunter-Gatherers in a Multidisciplinary Perspective: Challenging Elusiveness* (pp. 41-60). Leiden: CNWS, Leiden University.

Blust, R. (2013). Terror from the sky: Unconventional linguistic clues to the Negrito past. *Human Biology, 85*(1-3), 401-416.

Bomhard, A. R., & Kerns, J. C. (1994). *The Nostratic macrofamily: a study in distant linguistic relationship* (Vol. 74): Walter de Gruyter.

Borgerhoff Mulder, M., Nunn, C. L., & Towner, M. C. (2006). Cultural macroevolution and the transmission of traits. *Evolutionary Anthropology, 15*(2), 52-64. doi:10.1002/evan.20088

Bouckaert, R., Lemey, P., Dunn, M., Greenhill, S., Alekseyenko, A., & Drummond, A. (2012). Mapping the origins and expansion of the Indo-European language family. *Science, 337*. doi:10.1126/science.1219669

Bowern, C. (2012). The riddle of Tasmanian languages. *Proceedings of the Royal Society B-Biological Sciences, 279*(1747), 4590-4595. doi:10.1098/rspb.2012.1842

Bowern, C., & Atkinson, Q. (2012). Computational phylogenetics and the internal structure of Pama-Nyungan. *Language, 88*(4), 817-845.

Bowern, C., Epps, P., Gray, R., Hill, J., Hunley, K., & McConvell, P. (2011). Does lateral transmission obscure inheritance in hunter-gatherer languages? *Plos One, 6*. doi:10.1371/journal.pone.0025195

Boyd, R., & Richerson, P. J. (1985). *Culture and the Evolutionary Process*. Chicago: University of Chicago Press.

Bronzati, M., Montefeltro, F. C., & Langer, M. C. (2015). Diversification events and the effects of mass extinctions on Crocodyliformes evolutionary history. *Royal Society Open Science, 2*(5). doi:10.1098/rsos.140385

Brown, G. R., Dickins, T. E., Sear, R., & Laland, K. N. (2011). Evolutionary accounts of human behavioural diversity. *Philosophical Transactions of the Royal Society B-Biological Sciences, 366*(1563), 313-324. doi:10.1098/rstb.2010.0267

Buckley, C. D. (2012). Investigating cultural evolution using phylogenetic analysis: The origins and descent of the Southeast Asian tradition of warp ikat weaving. *Plos One, 7*(12), 20. doi:10.1371/journal.pone.0052064

Buerki, S., Forest, F., Salamin, N., & Alvarez, N. (2011). Comparative performance of supertree algorithms in large data sets using the soapberry family (Sapindaceae) as a case study. *Syst Biol, 60*(1), 32-44. doi:10.1093/sysbio/syq057

Buchanan, B., & Collard, M. (2007). Investigating the peopling of North America through cladistic analyses of Early Paleoindian projectile points. *Journal of Anthropological Archaeology, 26*(3), 366-393. doi:10.1016/j.jaa.2007.02.005

Busby, G. B. J., Hellenthal, G., Montinaro, F., Tofanelli, S., Bulayeva, K., Rudan, I., . . . Capelli, C. (2015). The role of recent admixture in forming the contemporary West Eurasian genomic landscape. *Current Biology, 25*(19), 2518-2526. doi:10.1016/j.cub.2015.08.007

Cann, H. M., de Toma, C., Cazes, L., Legrand, M. F., Morel, V., Piouffre, L., . . . Cavalli-Sforza, L. L. (2002). A human genome diversity cell line panel. *Science, 296*(5566), 261-262.

Cann, R. L., Stoneking, M., & Wilson, A. C. (1987). Mitochondrial DNA and human evolution. *Nature, 325*(6099), 31-36. doi:10.1038/325031a0

Cardillo, M., Mace, G. M., Gittleman, J. L., Jones, K. E., Bielby, J., & Purvis, A. (2008). The predictability of extinction: biological and external correlates of decline in mammals. *Proceedings of the Royal Society B-Biological Sciences, 275*(1641), 1441-1448. doi:10.1098/rspb.2008.0179

Cavalli-Sforza, L., & Feldman, M. W. (1973). Models for cultural inheritance I. Group mean and within group variation. *Theoretical population biology, 4*(1), 42-55.

Cavalli-Sforza, L. L., & Feldman, M. W. (1981). *Cultural Transmission and Evolution: A Quantitative Approach*: Princeton, New Jersey: Princeton University Press.

Cavalli-Sforza, L. L., Menozzi, P., & Piazza, A. (1994). *The History and Geography of Human Genes*. Princeton, New Jersey: Princeton university press.

Cavalli-Sforza, L. L., Minch, E., & Mountain, J. L. (1992). Coevolution of genes and languages revisited. *Proceedings of the National Academy of Sciences of the United States of America, 89*(12), 5620-5624. doi:10.1073/pnas.89.12.5620

Cavalli-Sforza, L. L., Piazza, A., Menozzi, P., & Mountain, J. (1988). Reconstruction of human evolution: bringing together genetic, archaeological, and linguistic data. *Proceedings of the National Academy of Sciences of the United States of America, 85*(16), 6002-6006. doi:10.1073/pnas.85.16.6002

Clarke, J. T., Lloyd, G. T., & Friedman, M. (2016). Little evidence for enhanced phenotypic evolution in early teleosts relative to their living fossil sister group. *Proceedings of the National Academy of Sciences of the United States of America, 113*(41), 11531-11536. doi:10.1073/pnas.1607237113

Cochrane, E. E., & Lipo, C. P. (2010). Phylogenetic analyses of Lapita decoration do not support branching evolution or regional population structure during colonization of Remote Oceania. *Philosophical Transactions of the Royal Society B-Biological Sciences, 365*(1559), 3889-3902. doi:10.1098/rstb.2010.0091

Collard, M., Shennan, S. J., & Tehrani, J. J. (2006). Branching, blending, and the evolution of cultural similarities and differences among human populations. *Evolution and Human Behavior, 27*(3), 169-184. doi:10.1016/j.evolhumbehav.2005.07.003

Cotton, J. A., Slater, C. S. C., & Wilkinson, M. (2006). Discriminating supported and unsupported relationships in supertrees using triplets. *Syst Biol, 55*(2), 345-350. doi:10.1080/10635150500481556

Coward, F., Shennan, S., Colledge, S., Conolly, J., & Collard, M. (2008). The spread of Neolithic plant economies from the Near East to northwest Europe: a phylogenetic analysis. *Journal of Archaeological Science, 35*(1), 42-56. doi:10.1016/j.jas.2007.02.022

Currie, T. E., Greenhill, S. J., Gray, R. D., Hasegawa, T., & Mace, R. (2010). Rise and fall of political complexity in island South-East Asia and the Pacific. *Nature, 467*(7317), 801-804. doi:10.1038/nature09461

Currie, T. E., Greenhill, S. J., & Mace, R. (2010). Is horizontal transmission really a problem for phylogenetic comparative methods? A simulation study using continuous cultural traits. *Philosophical Transactions of the Royal Society B-Biological Sciences, 365*(1559), 3903-3912. doi:10.1098/rstb.2010.0014

Currie, T. E., & Mace, R. (2011). Mode and tempo in the evolution of socio-political organization: reconciling 'Darwinian' and 'Spencerian' evolutionary approaches in anthropology. *Philosophical Transactions of the Royal Society B-Biological Sciences, 366*(1567), 1108-1117. doi:10.1098/rstb.2010.0318

Currie, T. E., Meade, A., Guillon, M., & Mace, R. (2013). Cultural phylogeography of the Bantu Languages of sub-Saharan Africa. *Proceedings of the Royal Society B-Biological Sciences, 280*(1762), 8. doi:10.1098/rspb.2013.0695

da Silva, S. G., & Tehrani, J. J. (2016). Comparative phylogenetic analyses uncover the ancient roots of Indo-European folktales. *Royal Society Open Science, 3*(1), 11. doi:10.1098/rsos.150645

Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. London: John Murray.

Davis, R. B., Baldauf, S. L., & Mayhew, P. J. (2010). Many hexapod groups originated earlier and withstood extinction events better than previously realized: inferences from supertrees. *Proceedings of the Royal Society B-Biological Sciences, 277*(1687), 1597-1606. doi:10.1098/rspb.2009.2299

De Schryver, G.-M., Grollemund, R., Branford, S., & Bostoen, K. (2015). Introducing a state-of-the-art phylogenetic classification of the Kikongo Language Cluster. *Africana Linguistica, 21*, 87-162.

Diamond, J. M. (1988). Express train to Polynesia. *Nature, 336*(6197), 307-308.

Di Fiore, A., & Rendall, D. (1994). Evolution of social organization: a reappraisal for primates by using phylogenetic methods. *Proceedings of the National Academy of Sciences of the United States of America, 91*(21), 9941-9945.

Dobzhansky, T. (1973). Nothing in Biology Makes Sense Except in the Light of Evolution. *American Biology Teacher, 35*(3), 125-129.

Dow, M. M., & Eff, E. A. (2008). Global, regional, and local network autocorrelation in the standard cross-cultural sample. *Cross-Cultural Research, 42*(2), 148-171. doi:10.1177/1069397107311186

Duda, P. (2015). Pygmejové pohledem evoluční biologie. In L. Hroníková & M. Schierová (Eds.), *Pygmejové: nejmenší lidé pohledem antropologie & Šebestova sbírka v Hrdličkově muzeu člověka PřF UK* (pp. 18-48). Praha: Academia.

Duda, P. (*in press*). Kulturní fylogenetika - Využití fylogenetických metod ke studiu evoluční historie jazyka a kultury. In L. Ovčáčková (Ed.), *Biologické a evoluční teorie kultury*. Praha: Academia.

Duda, P., & Zrzavý, J. (2013). Evolution of life history and behavior in Hominidae: Towards phylogenetic reconstruction of the chimpanzee-human last common ancestor. *Journal of Human Evolution, 65*(4), 424-446. doi:10.1016/j.jhevol.2013.07.009

Duda, P., & Zrzavý, J. (2016). Human population history revealed by a supertree approach. *Scientific Reports, 6*, 10. doi:10.1038/srep29890

Dunn, M. (2009). Contact and phylogeny in Island Melanesia. *Lingua, 119*(11), 1664-1678. doi:10.1016/j.lingua.2007.10.026

Dunn, M., Kruspe, N., & Burenhult, N. (2013). Time and place in the prehistory of the Aslian languages. *Human Biology, 85*(1-3), 383-399.

Dunn, M., Terrill, A., Reesink, G., Foley, R. A., & Levinson, S. C. (2005). Structural phylogenetics and the reconstruction of ancient language history. *Science, 309*(5743), 2072-2075. doi:10.1126/science.1114615

Ebenesersdottir, S. S., Sigurosson, A., Sanchez-Quinto, F., Lalueza-Fox, C., Stefansson, K., & Helgason, A. (2011). A new subclade of mtDNA haplogroup C1 found in Icelanders: Evidence of pre-Columbian contact? *American Journal of Physical Anthropology, 144*(1), 92-99. doi:10.1002/ajpa.21419

Eff, E. A. (2004). Does Mr. Galton still have a problem? Autocorrelation in the standard cross-cultural sample. *World Cultures*, *15*(2), 153-170.

Espinoza, R. E., Wiens, J. J., & Tracy, C. R. (2004). Recurrent evolution of herbivory in small, cold-climate lizards: Breaking the ecophysiological rules of reptilian herbivory. *Proceedings of the National Academy of Sciences of the United States of America, 101*(48), 16819-16824. doi:10.1073/pnas.0401226101

Felsenstein, J. (1985). Phylogenies and the comparative method. *American Naturalist, 125*(1), 1-15. doi:10.1086/284325

Fortunato, L. (2011a). Reconstructing the history of marriage strategies in Indo-European-speaking societies: Monogamy and polygyny. *Human Biology, 83*(1), 87-105.

Fortunato, L. (2011b). Reconstructing the history of residence strategies in Indo-European-speaking societies: Neo-, uxori-, and virilocality. *Human Biology, 83*(1), 107-128.

Fortunato, L., Holden, C., & Mace, R. (2006). From bridewealth to dowry? *Human Nature, 17*(4), 355-376. doi:10.1007/s12110-006-1000-4

Fortunato, L., & Jordan, F. (2010). Your place or mine? A phylogenetic comparative analysis of marital residence in Indo-European and Austronesian societies. *Philosophical Transactions of the Royal Society B-Biological Sciences, 365*(1559), 3913-3922. doi:10.1098/rstb.2010.0017

Galucio, A. V., Meira, S., Birchall, J., Moore, D., Gabas Júnior, N., Drude, S., . . . Rodrigues, C. R. (2015). Genealogical relations and lexical distances within the Tupian linguistic family. *Boletim do Museu Paraense Emílio Goeldi. Ciências Humanas, 10*(2), 229-274.

Garamszegi, L. Z. (Ed.) (2014). *Modern Phylogenetic Comparative Methods and their Application in Evolutionary Biology*. Heidelberg: Springer-Verlag.

Gell-Mann, M., Peiros, I., & Starostin, G. (2009). Distant language relationship: The current perspective. *Journal of Language Relationship, 1*, 13-30.

Gibbs, R. A., Belmont, J. W., Hardenbol, P., Willis, T. D., Yu, F. L., Yang, H. M., . . . Int HapMap, C. (2003). The International HapMap Project. *Nature, 426*(6968), 789-796. doi:10.1038/nature02168

Gimbutas, M. (1973). Old Europe c. 7000–3500 B.C.: The earliest European civilization before the infiltration of the Indo-European peoples. *JIES, 1*.

Grafen, A. (1989). The phylogenetic regression. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences, 326*(1233), 119-157. doi:10.1098/rstb.1989.0106

Gray, R. (2005). Pushing the time barrier in the quest for language roots. *Science, 309*(5743), 2007-2008. doi:10.1126/science.1119276

Gray, R. D., & Atkinson, Q. D. (2003). Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature, 426*(6965), 435-439. doi:10.1038/nature02029

Gray, R. D., Bryant, D., & Greenhill, S. J. (2010). On the shape and fabric of human history. *Philosophical Transactions of the Royal Society B-Biological Sciences, 365*(1559), 3923-3933. doi:10.1098/rstb.2010.0162

Gray, R. D., Drummond, A. J., & Greenhill, S. J. (2009). Language phylogenies reveal expansion pulses and pauses in Pacific settlement. *Science, 323*(5913), 479-483. doi:10.1126/science.1166858

Gray, R. D., & Jordan, F. M. (2000). Language trees support the express-train sequence of Austronesian expansion. *Nature, 405*(6790), 1052-1055. doi:10.1038/35016575

Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., . . . Pääbo, S. (2010). A Draft Sequence of the Neandertal Genome. *Science, 328*(5979), 710-722. doi:10.1126/science.1188021

Greenberg, J. H. (1954). Studies in African linguistic classification: VIII. Further remarks on method; revisions and corrections. *Southwestern Journal of Anthropology, 10*, 405-415.

Greenberg, J. H. (1963). *The Languages of Africa*. Bloomington: Indiana University Press.

Greenberg, J. H. (1971). The Indo-Pacific hypothesis. In T. A. Sebeok (Ed.), *Current trends in linguistics* (Vol. 8, pp. 807-871). The Hague, The Netherlands: Mouton.

Greenberg, J. H. (1987). *Language in the Americas*: Stanford University Press.

Greenberg, J. H. (2000). *The Eurasiatic Language Family: Indo-European and Its Closest Relations. Volume I: Grammar* (Vol. 1). Stanford: Stanford University Press.

Greenberg, J. H. (2002). *Indo-European and its Closest Relatives: The Eurasiatic Language Family Volume II: Lexicon* (Vol. 2). Stanford: Stanford University Press.

Greenhill, S. J., Currie, T. E., & Gray, R. D. (2009). Does horizontal transmission invalidate cultural phylogenies? *Proceedings of the Royal Society B-Biological Sciences, 276*(1665), 2299-2306. doi:10.1098/rspb.2008.1944

Grollemund, R., Branford, S., Bostoen, K., Meade, A., Venditti, C., & Pagel, M. (2015). Bantu expansion shows that habitat alters the route and pace of human dispersals. *Proceedings of the National Academy of Sciences*, 201503793.

Guglielmino, C. R., Viganotti, C., Hewlett, B., & Cavallisforza, L. L. (1995). Cultural variation in Africa: Role of mechanisms of transmission and adaptation. *Proceedings of the National Academy of Sciences of the United States of America, 92*(16), 7585-7589. doi:10.1073/pnas.92.16.7585

Guillon, M., & Mace, R. (2016). A phylogenetic comparative study of Bantu kinship terminology finds limited support for its co-evolution with social organisation. *Plos One, 11*(3), e0147920. doi:10.1371/journal.pone.0147920

Haak, W., Lazaridis, I., Patterson, N., Rohland, N., Mallick, S., Llamas, B., . . . Reich, D. (2015). Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature, 522*. doi:10.1038/nature14317

Harvey, P. H., & Pagel, M. D. (1991). *The comparative method in evolutionary biology* (Vol. 239). Oxford, UK: Oxford University Press.

Helgason, A., Sigurðardóttir, S., Gulcher, J. R., Ward, R., & Stefánsson, K. (2000). MtDNA and the origin of the Icelanders: Deciphering signals of recent population history. *The American Journal of Human Genetics, 66*(3), 999-1016. doi:10.1086/302816

Hellenthal, G., Busby, G. B. J., Band, G., Wilson, J. F., Capelli, C., Falush, D., & Myers, S. (2014). A genetic atlas of human admixture history. *Science, 343*(6172), 747-751. doi:10.1126/science.1243518

Hewlett, B. S., & Cavalli-Sforza, L. L. (1986). Cultural transmission among Aka pygmies. *American Anthropologist, 88*(4), 922-934.

Hewlett, B. S., de Silvestri, A., & Guglielmino, C. R. (2002). Semes and genes in Africa. *Current Anthropology, 43*(2), 313-321. doi:10.1086/339379

Heyer, E. (2012). Interview with Laura Fortunato, winner of the 2011 Gabriel W. Lasker prize. *Human Biology, 84*(3), 227-234.

Holden, C., & Mace, R. (1997). Phylogenetic analysis of the evolution of lactose digestion in adults. *Human Biology, 69*(5), 605-628.

Holden, C., & Mace, R. (1999). Sexual dimorphism in stature and women's work: A phylogenetic cross-cultural analysis. *American Journal of Physical Anthropology, 110*(1), 27-45. doi:10.1002/(sici)1096-8644(199909)110:1<27::aid-ajpa3>3.0.co;2-g

Holden, C. J. (2002). Bantu language trees reflect the spread of farming across sub-Saharan Africa: a maximum-parsimony analysis. *Proceedings of the Royal Society B-Biological Sciences, 269*(1493), 793-799. doi:10.1098/rspb.2002.1955

Holden, C. J., & Gray, R. D. (2006). Rapid radiation, borrowing and dialect continua in the Bantu languages. In P. Forster & C. Renfrew (Eds.), *Phylogenetic Methods and the Prehistory of Languages* (pp. 19-31). Cambridge: McDonald Institute for Archaeological Research.

Holden, C. J., & Mace, R. (2003). Spread of cattle led to the loss of matrilineal descent in Africa: a coevolutionary analysis. *Proceedings of the Royal Society B-Biological Sciences, 270*(1532), 2425-2433. doi:10.1098/rspb.2003.2535

Honkola, T., Vesakoski, O., Korhonen, K., Lehtinen, J., Syrjanen, K., & Wahlberg, N. (2013). Cultural and climatic changes shape the evolutionary history of the Uralic languages. *Journal of Evolutionary Biology, 26*(6), 1244-1253. doi:10.1111/jeb.12107

Hruschka, D. J., Branford, S., Smith, E. D., Wilkins, J., Meade, A., Pagel, M., & Bhattacharya, T. (2015). Detecting regular sound changes in linguistics as events of concerted evolution. *Current Biology, 25*(1), 1-9. doi:10.1016/j.cub.2014.10.064

Hunley, K., Dunn, M., Lindstrom, E., Reesink, G., Terrill, A., Healy, M. E., . . . Friedlaender, J. S. (2008). Genetic and linguistic coevolution in northern island Melanesia. *Plos Genetics, 4*(10). doi:10.1371/journal.pgen.1000239

Hunley, K. L., Cabana, G. S., Merriwether, D. A., & Long, J. C. (2007). A formal test of linguistic and genetic coevolution in Native Central and South America. *American Journal of Physical Anthropology, 132*(4), 622-631. doi:10.1002/ajpa.20542

Chacon, T. C., & List, J.-M. (2015). Improved computational models of sound change shed light on the history of the Tukanoan languages. *Journal of Language Relationship• Вопросы языкового родства*, 177-203.

Chang, W., Cathcart, C., Hall, D., & Garrett, A. (2015). Ancestry-constrained phylogenetic analysis supports the Indo-European steppe hypothesis. *Language, 91*(1), 194-244.

Cheverud, J. M., Dow, M. M., & Leutenegger, W. (1985). The quantitative assessment of phylogenetic constraints in comparative analyses: sexual dimorphism in body weight among primates. *Evolution, 39*(6), 1335-1351. doi:10.2307/2408790

Jaeggi, A. V., & Gurven, M. (2013). Reciprocity explains food sharing in humans and other primates independent of kin selection and tolerated scrounging: a phylogenetic meta-analysis. *Proceedings of the Royal Society B-Biological Sciences, 280*(1768), 8. doi:10.1098/rspb.2013.1615

Jäger, G. (2015). Support for linguistic macrofamilies from weighted sequence alignment. *Proceedings of the National Academy of Sciences, 112*(41), 12752-12757. doi:10.1073/pnas.1500331112

Jay, F., Francois, O., & Blum, M. G. B. (2011). Predictions of Native American population structure using linguistic covariates in a hidden regression framework. *Plos One, 6*(1). doi:10.1371/journal.pone.0016227

Jordan, F. M., Gray, R. D., Greenhill, S. J., & Mace, R. (2009). Matrilocal residence is ancestral in Austronesian societies. *Proceedings of the Royal Society B-Biological Sciences, 276*(1664), 1957-1964. doi:10.1098/rspb.2009.0088

Jordan, P., & O'Neill, S. (2010). Untangling cultural inheritance: language diversity and long-house architecture on the Pacific northwest coast. *Philosophical Transactions of the Royal Society B-Biological Sciences, 365*(1559), 3875-3888. doi:10.1098/rstb.2010.0092

Jordan, P., & Shennan, S. (2003). Cultural transmission, language, and basketry traditions amongst the California Indians. *Journal of Anthropological Archaeology, 22*(1), 42-74. doi:10.1016/s0278-4165(03)00004-7

Karafet, T. M., Bulayeva, K. B., Nichols, J., Bulayev, O. A., Gurgenova, F., Omarova, J., . . . Hammer, M. F. (2015). Coevolution of genes and languages and high levels of population structure among the highland populations of Daghestan. *J Hum Genet*. doi:10.1038/jhg.2015.132

Kitchen, A., Ehret, C., Assefa, S., & Mulligan, C. J. (2009). Bayesian phylogenetic analysis of Semitic languages identifies an Early Bronze Age origin of Semitic in the Near East. *Proceedings of the Royal Society B: Biological Sciences, 276*(1668), 2703-2710. doi:10.1098/rspb.2009.0408

Ko, A. M. S., Chen, C. Y., Fu, Q. M., Delfin, F., Li, M. K., Chiu, H. L., . . . Ko, Y. C. (2014). Early Austronesians: Into and Out Of Taiwan. *American Journal of Human Genetics, 94*(3), 426-436. doi:10.1016/j.ajhg.2014.02.003

Krause, J., Fu, Q., Good, J. M., Viola, B., Shunkov, M. V., Derevianko, A. P., & Paabo, S. (2010). The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. *Nature, 464*(7290), 894-897. doi:10.1038/nature08976

Kushnick, G., Gray, R. D., & Jordan, F. M. (2014). The sequential evolution of land tenure norms. *Evolution and Human Behavior, 35*(4), 309-318. doi:10.1016/j.evolhumbehav.2014.03.001

Laland, K. N., & Brown, G. R. (2011). *Sense and nonsense: Evolutionary perspectives on human behaviour*: Oxford University Press.

Le Bomin, S., Lecointre, G., & Heyer, E. (2016). The evolution of musical diversity: The key role of vertical transmission. *Plos One, 11*(3), 17. doi:10.1371/journal.pone.0151570

Lee, S., & Hasegawa, T. (2011). Bayesian phylogenetic analysis supports an agricultural origin of Japonic languages. *Proceedings of the Royal Society B-Biological Sciences, 278*(1725), 3662-3669. doi:10.1098/rspb.2011.0518

Lee, S., & Hasegawa, T. (2013). Evolution of the Ainu language in space and time. *Plos One, 8*(4), 6. doi:10.1371/journal.pone.0062243

Levinson, S. C., Greenhill, S. J., Gray, R. D., & Dunn, M. (2011). Universal typological dependencies should be detectable in the history of language families. *Linguistic Typology, 15*, 509-534. doi:10.1515/LITY.2011.034

Lewis, P. M., Simons, Gary F., & Fennig, C. D. (2016). Ethnologue: Languages of the World, Nineteenth edition. Retrieved from http://www.ethnologue.com

Lipo, C. P., O’Brien, M. J., Collard, M., & Shennan, S. (Eds.). (2006). *Mapping our Ancestors: Phylogenetic Approaches in Anthropology and Prehistory*. New York: Aldine.

List, J.-M., Pathmanathan, J. S., Lopez, P., & Bapteste, E. (2016). Unity and disunity in evolutionary sciences: process-based analogies open common research avenues for biology and linguistics. *Biology Direct, 11*(1), 39. doi:10.1186/s13062-016-0145-2

Longobardi, G., Ghirotto, S., Guardiano, C., Tassi, F., Benazzo, A., Ceolin, A., & Barbujani, G. (2015). Across language families: Genome diversity mirrors linguistic variation within Europe. *American Journal of Physical Anthropology, 157*(4), 630-640. doi:10.1002/ajpa.22758

Lukas, D., & Clutton-Brock, T. (2012). Cooperative breeding and monogamy in mammalian societies. *Proceedings of the Royal Society B-Biological Sciences, 279*(1736), 2151-2156. doi:10.1098/rspb.2011.2468

Mace, R., & Holden, C. J. (2005). A phylogenetic approach to cultural evolution. *Trends in Ecology & Evolution, 20*(3), 116-121. doi:10.1016/j.tree.2004.12.002

Mace, R., Holden, C. J., & Shennan, S. (Eds.). (2005). *The Evolution of Cultural Diversity: A Phylogenetic Approach*. London: University College London Press.

Mace, R., & Jordan, F. M. (2011). Macro-evolutionary studies of cultural diversity: a review of empirical studies of cultural transmission and cultural adaptation. *Philosophical Transactions of the Royal Society B-Biological Sciences, 366*(1563), 402-411. doi:10.1098/rstb.2010.0238

Mace, R., & Pagel, M. (1994). The comparative method in anthropology. *Current Anthropology, 35*(5), 549-564. doi:10.1086/204317

Malaspinas, A.-S., Westaway, M. C., Muller, C., Sousa, V. C., Lao, O., Alves, I., . . . Willerslev, E. (2016). A genomic history of Aboriginal Australia. *Nature, 538*(7624), 207-+. doi:10.1038/nature18299

Mallick, S., Li, H., Lipson, M., Mathieson, I., Gymrek, M., Racimo, F., . . . Reich, D. (2016). The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature, 538*(7624), 201-+. doi:10.1038/nature18964

Marwick, B. (2012). A cladistic evaluation of ancient Thai bronze Buddha images: six tests for a phylogenetic signal in the Griswold collection. In D. Bonatz, A. Reinecke, & M. A. Tjoa-Bonatz (Eds.), *Connecting Empires* (pp. 159-176). Singapore: National University of Singapore Press.

Mendoza Straffon, L. (Ed.) (2016). *Cultural Phylogenetics: Concepts and Applications in Archaeology* (Vol. 4). Heidelberg: Springer-Verlag.

Mesoudi, A., Whiten, A., & Laland, K. N. (2006). Towards a unified science of cultural evolution. *Behavioral and brain sciences, 29*(4), 329-+. doi:10.1017/s0140525x06009083

Mesoudi, A., Whiten, A., Laland, K. N., & Harrison, R. (2004). Perspective: Is human cultural evolution Darwinian? Evidence reviewed from the perspective of The Origin of Species. *Evolution, 58*(1), 1-11. doi:10.1554/03-212

Michael, L., Chousou-Polydouri, N., Keith, B., Donnelly, E., Meira, S., Wauters, V., & O'Hagan, Z. (2015). A Bayesian phylogenetic classification of Tupí-Guaraní. *LIAMES-Línguas Indígenas Americanas, 15*(2), 193-221.

Morgan, L. (1871). *Ancient Society*. New York: Henry Holt.

Murdock, G. P. (1957). World ethnographic sample. *American Anthropologist, 59*(4), 664-687.

Murdock, G. P. (1967). *Ethnographic Atlas*: University of Pittsburgh Press.

Murdock, G. P., & White, D. R. (1969). Standard cross-cultural sample. *Ethnology, 8*(4), 329-369.

Naroll, R. (1961). Two Solutions to Galton's Problem. *Philosophy of Science, 28*(1), 15-39. doi:doi:10.1086/287778

Naroll, R. (1965). Galton's problem: The logic of cross-cultural analysis. *Social Research*, 428-451.

Nei, M., & Roychoudhury, A. K. (1993). Evolutionary relationships of human populations on a global scale. *Molecular Biology and Evolution, 10*(5), 927-943.

Nguyen, N., Mirarab, S., & Warnow, T. (2012). MRL and SuperFine+MRL: new supertree methods. *Algorithms Mol Biol, 7*(1), 3. doi:10.1186/1748-7188-7-3

Novembre, J., Johnson, T., Bryc, K., Kutalik, Z., Boyko, A. R., Auton, A., . . . Bustamante, C. D. (2008). Genes mirror geography within Europe. *Nature, 456*(7218), 98-U95. doi:10.1038/nature07331

Nunn, C. L. (2011). *The Comparative Approach in Evolutionary Anthropology and Biology*. Chicago: University of Chicago Press.

Nunn, C. L., Arnold, C., Matthews, L., & Mulder, M. B. (2010). Simulating trait evolution for cross-cultural comparison. *Philosophical Transactions of the Royal Society B-Biological Sciences, 365*(1559), 3807-3819. doi:10.1098/rstb.2010.0009

Nurbakova, D., Rusakov, S., & Alexandrov, V. (2013). Quantifying uncertainty in phylogenetic studies of the Slavonic languages. *Procedia Computer Science, 18*, 2269-2277. doi:10.1016/j.procs.2013.05.398

O'Brien, M. J., Boulanger, M. T., Buchanan, B., Collard, M., Lyman, R. L., & Darwent, J. (2014). Innovation and cultural transmission in the American Paleolithic: Phylogenetic analysis of eastern Paleoindian projectile-point classes. *Journal of Anthropological Archaeology, 34*, 100-119. doi:10.1016/j.jaa.2014.03.001

O’Neill, S., Ellen, R., Lycett, S., & Johns, S. (2013). Co-evolution between bentwood box traditions and languages on the Pacific Northwest Coast. In R. Ellen, S. Lycett, & S. Johns (Eds.), *Understanding Cultural Transmission in Anthropology: A Critical Synthesis* (pp. 165-190). Oxford: Berghahn.

Opie, C., Atkinson, Q. D., Dunbar, R. I. M., & Shultz, S. (2013). Male infanticide leads to social monogamy in primates. *Proceedings of the National Academy of Sciences of the United States of America, 110*(33), 13328-13332. doi:10.1073/pnas.1307903110

Opie, C., Shultz, S., Atkinson, Q. D., Currie, T., & Mace, R. (2014). Phylogenetic reconstruction of Bantu kinship challenges Main Sequence Theory of human social evolution. *Proceedings of the National Academy of Sciences of the United States of America, 111*(49), 17414-17419. doi:10.1073/pnas.1415744111

Pagel, M. (1994). Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. *Proceedings of the Royal Society B-Biological Sciences, 255*(1342), 37-45. doi:10.1098/rspb.1994.0006

Pagel, M. (1999). The maximum likelihood approach to reconstructing ancestral character states of discrete characters on phylogenies. *Syst Biol, 48*(3), 612-622. doi:10.1080/106351599260184

Pagel, M. (2009). Human language as a culturally transmitted replicator. *Nature Reviews Genetics, 10*(6), 405-415. doi:10.1038/nrg2560

Pagel, M., Atkinson, Q. D., Calude, A. S., & Meade, A. (2013). Ultraconserved words point to deep language ancestry across Eurasia. *Proceedings of the National Academy of Sciences of the United States of America, 110*(21), 8471-8476. doi:10.1073/pnas.1218726110

Pagel, M., Atkinson, Q. D., & Meade, A. (2007). Frequency of word-use predicts rates of lexical evolution throughout Indo-European history. *Nature, 449*(7163), 717-U717. doi:10.1038/nature06176

Pagel, M., & Mace, R. (2004). The cultural wealth of nations. *Nature, 428*(6980), 275-278. doi:10.1038/428275a

Pagel, M. D., & Harvey, P. H. (1988). Recent developments in the analysis of comparative data. *Quarterly Review of Biology, 63*(4), 413-440. doi:10.1086/416027

Pakendorf, B. (2014). Coevolution of languages and genes. *Current opinion in genetics & development, 29*, 39-44. doi:10.1016/j.gde.2014.07.006

Pemberton, T. J., DeGiorgio, M., & Rosenberg, N. A. (2013). Population structure in a comprehensive genomic data set on human microsatellite variation. *G3-Genes Genomes Genetics, 3*(5), 891-907. doi:10.1534/g3.113.005728

Penny, D., Watson, E. E., & Steel, M. A. (1993). Trees from languages and genes are very similar. *Syst Biol, 42*(3), 382-384. doi:10.2307/2992472

Peoples, H. C., Duda, P., & Marlowe, F. W. (2016). Hunter-gatherers and the origins of religion. *Human Nature-an Interdisciplinary Biosocial Perspective, 27*(3), 261-282. doi:10.1007/s12110-016-9260-0

Pickrell, J. K., & Pritchard, J. K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *Plos Genetics, 8*(11). doi:10.1371/journal.pgen.1002967

Pickrell, J. K., & Reich, D. (2014). Toward a new history and geography of human genes informed by ancient DNA. *Trends in Genetics, 30*(9), 377-389. doi:10.1016/j.tig.2014.07.007

Pisani, D., & Wilkinson, M. (2002). Matrix representation with parsimony, taxonomic congruence, and total evidence. *Syst Biol, 51*(1), 151-155. doi:10.1080/106351502753475925

Pisani, D., Yates, A. M., Langer, M. C., & Benton, M. J. (2002). A genus-level supertree of the Dinosauria. *Proceedings of the Royal Society B-Biological Sciences, 269*(1494), 915-921. doi:10.1098/rspb.2001.1942

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics, 155*(2), 945-959.

Pugach, I., Delfin, F., Gunnarsdottir, E., Kayser, M., & Stoneking, M. (2013). Genome-wide data substantiate Holocene gene flow from India to Australia. *Proceedings of the National Academy of Sciences of the United States of America, 110*(5), 1803-1808. doi:10.1073/pnas.1211927110

Ragan, M. A. (1992). Phylogenetic Inference Based on Matrix Representation of Trees. *Molecular Phylogenetics and Evolution, 1*(1), 53-58. doi:10.1016/1055-7903(92)90035-f

Raghavan, M., Skoglund, P., Graf, K. E., Metspalu, M., Albrechtsen, A., Moltke, I., . . . Willerslev, E. (2014). Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature, 505*(7481), 87-+. doi:10.1038/nature12736

Rasmussen, M., Anzick, S. L., Waters, M. R., Skoglund, P., DeGiorgio, M., Stafford, T. W., . . . Willerslev, E. (2014). The genome of a Late Pleistocene human from a Clovis burial site in western Montana. *Nature, 506*(7487), 225-+. doi:10.1038/nature13025

Rasmussen, M., Guo, X. S., Wang, Y., Lohmueller, K. E., Rasmussen, S., Albrechtsen, A., . . . Willerslev, E. (2011). An Aboriginal Australian genome reveals separate human dispersals into Asia. *Science, 334*(6052), 94-98. doi:10.1126/science.1211177

Rasmussen, M., Li, Y., Lindgreen, S., Pedersen, J. S., Albrechtsen, A., Moltke, I., . . . Willerslev, E. (2010). Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature, 463*(7282), 757-762. doi:http://www.nature.com/nature/journal/v463/n7282/suppinfo/nature08835_S1.html

Reesink, G., Singer, R., & Dunn, M. (2009). Explaining the linguistic diversity of Sahul using population models. *Plos Biology, 7*(11). doi:10.1371/journal.pbio.1000241

Reid, L. A. (2013). Who are the Philippine negritos? Evidence from language. *Human Biology, 85*(1-3), 329-358.

Reich, D., Green, R. E., Kircher, M., Krause, J., Patterson, N., Durand, E. Y., . . . Paabo, S. (2010). Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature, 468*(7327), 1053-1060. doi: 10.1038/nature09710

Renfrew, C. (1990). *Archaeology and Language: The Puzzle of Indo-European Origins*. Cambridge: Cambridge University Press.

Rexová, K., Bastin, Y., & Frynta, D. (2006). Cladistic analysis of Bantu languages: a new tree based on combined lexical and grammatical data. *Naturwissenschaften, 93*(4), 189-194. doi:10.1007/s00114-006-0088-z

Rexová, K., Frynta, D., & Zrzavý, J. (2003). Cladistic analysis of languages: Indo-European classification based on lexicostatistical data. *Cladistics-the International Journal of the Willi Hennig Society, 19*(2), 120-127. doi:10.1016/s0748-3007(02)00147-0

Riede, F. (2009). Tangled trees: Modelling material culture evolution as host-associate co-speciation. In S. Shennan (Ed.), *Pattern and process in cultural evolution* (pp. 85-99). Berkeley, CA: University of California Press.

Richerson, P. J., & Boyd, R. (2005). *Not by Genes Alone* (pp. 342). Chicago: University of Chicago Press.

Rivero, D. G. (2016). Darwinian archaeology and cultural phylogenetics. In L. Mendoza Straffon (Ed.), *Cultural Phylogenetics: Concepts and Applications in Archaeology* (Vol. 4, pp. 43-72). Heidelberg: Springer-Verlag.

Rivero, D. G., & O'Brien, M. J. (2014). Phylogenetic analysis shows that Neolithic slate plaques from the southwestern Iberian Peninsula are not genealogical recording systems. *Plos One, 9*(2), 18. doi:10.1371/journal.pone.0088296

Rogers, D. S., Feldman, M. W., & Ehrlich, P. R. (2009). Inferring population histories using cultural data. *Proceedings of the Royal Society B-Biological Sciences, 276*(1674), 3835-3843. doi:10.1098/rspb.2009.1088

Rosenberg, N. A., Pritchard, J. K., Weber, J. L., Cann, H. M., Kidd, K. K., Zhivotovsky, L. A., & Feldman, M. W. (2002). Genetic structure of human populations. *Science, 298*(5602), 2381-2385. doi:10.1126/science.1078311

Ross, R. M., Greenhill, S. J., & Atkinson, Q. D. (2013). Population structure and cultural geography of a folktale in Europe. *Proceedings of the Royal Society B-Biological Sciences, 280*(1756), 9. doi:10.1098/rspb.2012.3065

Ruhlen, M. (1991). *Guide to the World's Languages: Classification* (Vol. 1): Stanford University Press.

Ruta, M., Pisani, D., Lloyd, G. T., & Benton, M. J. (2007). A supertree of Temnospondyli: cladogenetic patterns in the most species-rich group of early tetrapods. *Proceedings of the Royal Society B-Biological Sciences, 274*(1629), 3087-3095. doi:10.1098/rspb.2007.1250

Sellen, D. W., & Mace, R. (1997). Fertility and mode of subsistence: A phylogenetic analysis. *Current Anthropology, 38*(5), 878-889. doi:10.1086/204677

Serva, M. (2012). The settlement of Madagascar: What dialects and languages can tell us. *Plos One, 7*(2), 6. doi:10.1371/journal.pone.0030666

Serva, M., & Petroni, F. (2008). Indo-European languages tree by Levenshtein distance. *EPL (Europhysics Letters), 81*(6). doi:10.1209/0295-5075/81/68005

Sicoli, M. A., & Holton, G. (2014). Linguistic phylogenies support back-migration from Beringia to Asia. *Plos One, 9*(3), 8. doi:10.1371/journal.pone.0091722

Spence, J. (2013). The phylogenetic status of Pacific Coast Athabaskan: a computational assessment. *Thirty-Ninth Annual Meeting of the Berkeley Linguistics Society*. Berkeley, CA: Berkeley Linguistics Society.

Stadler, T. (2011). Mammalian phylogeny reveals recent diversification rate shifts. *Proceedings of the National Academy of Sciences, 108*(15), 6187-6192. doi:10.1073/pnas.1016876108

Steele, J., & Kandler, A. (2010). Language trees not equal gene trees. *Theory in Biosciences, 129*(2-3), 223-233. doi:10.1007/s12064-010-0096-6

Sykes, B. (2001). *The Seven Daughters of Eve: The Science That Reveals Our Genetic Ancestry:* W.W.Norton & Company.

Tang, H., Peng, J., Wang, P., & Risch, N. J. (2005). Estimation of individual admixture: Analytical and study design considerations. *Genetic Epidemiology, 28*(4), 289-301. doi:10.1002/gepi.20064

Tehrani, J., & Collard, M. (2002). Investigating cultural evolution through biological phylogenetic analyses of Turkmen textiles. *Journal of Anthropological Archaeology, 21*(4), 443-463. doi:10.1016/s0278-4165(02)00002-8

Tehrani, J. J. (2013). The Phylogeny of Little Red Riding Hood. *Plos One, 8*(11), e78871. doi:10.1371/journal.pone.0078871

Tishkoff, S. A., Reed, F. A., Friedlaender, F. R., Ehret, C., Ranciaro, A., Froment, A., . . . Williams, S. M. (2009). The genetic structure and history of Africans and African Americans. *Science, 324*(5930), 1035-1044. doi:10.1126/science.1172257

Tripp, A. (2016). A cladistics analysis exploring regional patterning of the anthropomorphic figurines from the Gravettian. In L. Mendoza Straffon (Ed.), *Cultural Phylogenetics: Concepts and Applications in Archaeology* (Vol. 4, pp. 179-202). Heidelberg: Springer-Verlag.

van Oven, M., & Kayser, M. (2009). Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Human Mutation, 30*(2), E386-E394. doi:10.1002/humu.20921

van Oven, M., Van Geystelen, A., Kayser, M., Decorte, R., & Larmuseau, M. H. D. (2014). Seeing the wood for the trees: A minimal reference phylogeny for the human Y chromosome. *Human Mutation, 35*(2), 187-191. doi:10.1002/humu.22468

Veeramah, K. R., & Hammer, M. F. (2014). The impact of whole-genome sequencing on the reconstruction of human population history. *Nature Reviews Genetics, 15*(3), 149-162. doi:10.1038/nrg3625

Verdu, P., Austerlitz, F., Estoup, A., Vitalis, R., Georges, M., Thery, S., . . . Heyer, E. (2009). Origins and genetic diversity of pygmy hunter-gatherers from Western Central Africa. *Current Biology, 19*(4), 312-318. doi:10.1016/j.cub.2008.12.049

von Rueden, C. R., & Jaeggi, A. V. (2016). Men's status and reproductive success in 33 nonindustrial societies: Effects of subsistence, marriage system, and reproductive strategy. *Proceedings of the National Academy of Sciences of the United States of America, 113*(39), 10824-10829. doi:10.1073/pnas.1606800113

Walker, R. S., Flinn, M. V., & Hill, K. R. (2010). Evolutionary history of partible paternity in lowland South America. *Proceedings of the National Academy of Sciences of the United States of America, 107*(45), 19195-19200. doi:10.1073/pnas.1002598107

Walker, R. S., & Hamilton, M. J. (2011). Social complexity and linguistic diversity in the Austronesian and Bantu population expansions. *Proceedings of the Royal Society B-Biological Sciences, 278*(1710), 1399-1404. doi:10.1098/rspb.2010.1942

Walker, R. S., Hill, K. R., Flinn, M. V., & Ellsworth, R. M. (2011). Evolutionary history of hunter-gatherer marriage practices. *Plos One*, *6*(4). doi:10.1371/journal.pone.0019066

Walker, R. S., & Ribeiro, L. A. (2011). Bayesian phylogeography of the Arawak expansion in lowland South America. *Proceedings of the Royal Society B-Biological Sciences, 278*(1718), 2562-2567. doi:10.1098/rspb.2010.2579

Walker, R. S., Wichmann, S., Mailund, T., & Atkisson, C. J. (2012). Cultural phylogenetics of the Tupi language family in lowland South America. *Plos One, 7*(4). doi:10.1371/journal.pone.0035025

Watts, J., Greenhill, S. J., Atkinson, Q. D., Currie, T. E., Bulbulia, J., & Gray, R. D. (2015). Broad supernatural punishment but not moralizing high gods precede the evolution of political complexity in Austronesia. *Proceedings of the Royal Society B: Biological Sciences, 282*(1804). doi:10.1098/rspb.2014.2556

Wheeler, W. C., & Whiteley, P. M. (2015). Historical linguistics as a sequence optimization problem: the evolution and biogeography of Uto-Aztecan languages. *Cladistics, 31*(2), 113-125. doi:10.1111/cla.12078

Willems, M., Lord, E., Laforest, L., Labelle, G., Lapointe, F.-J., Di Sciullo, A. M., & Makarenkov, V. (2016). Using hybridization networks to retrace the evolution of Indo-European languages. *BMC Evolutionary Biology, 16*(1), 180. doi:10.1186/s12862-016-0745-6

Windram, H. F., Shaw, P., Robinson, P., & Howe, C. J. (2008). Dante's Monarchia as a test case for the use of phylogenetic methods in stemmatic analysis. *Literary and Linguistic Computing, 23*(4), 443-463. doi: 10.1093/llc/fqn023

Windram, H. F., Charlston, T., & Howe, C. J. (2014). A phylogenetic analysis of Orlando Gibbons's Prelude in G. *Early Music, 42*(4), 515-+. doi:10.1093/em/cau100

Wood, E. T., Stover, D. A., Ehret, C., Destro-Bisol, G. D., Spedini, G., McLeod, H., . . . Hammer, M. F. (2005). Contrasting patterns of Y chromosome and mtDNA variation in Africa: evidence for sex-biased demographic processes. *European Journal of Human Genetics, 13*(7), 867-876. doi:10.1038/sj.ejhg.5201408

Chapter I

Human population history revealed by a supertree approach

Duda, P., & Zrzavý, J.

Scientific Reports, *6*, 10.

doi:10.1038/srep29890

www.nature.com/scientificreports

SCIENTIFIC REPORTS

OPEN

Human population history revealed by a supertree approach

Pavel Duda[1,2] & **Jan Zrzavý**[1]

Received: 16 March 2016
Accepted: 23 June 2016
Published: 19 July 2016

Over the past two decades numerous new trees of modern human populations have been published extensively but little attention has been paid to formal phylogenetic synthesis. We utilized the "matrix representation with parsimony" (MRP) method to infer a composite phylogeny (supertree) of modern human populations, based on 257 genetic/genomic, as well as linguistic, phylogenetic trees and 44 admixture plots from 200 published studies (1990–2014). The resulting supertree topology includes the most basal position of S African Khoisan followed by C African Pygmies, and the paraphyletic section of all other sub-Saharan peoples. The sub-Saharan African section is basal to the monophyletic clade consisting of the N African–W Eurasian assemblage and the consistently monophyletic Eastern superclade (Sahul–Oceanian, E Asian, and Beringian–American peoples). This topology, dominated by genetic data, is well-resolved and robust to parameter set changes, with a few unstable areas (e.g., West Eurasia, Sahul–Melanesia) reflecting the existing phylogenetic controversies. A few populations were identified as highly unstable "wildcard taxa" (e.g. Andamanese, Malagasy). The linguistic classification fits rather poorly on the supertree topology, supporting a view that direct coevolution between genes and languages is far from universal.

Evolutionary history of modern human populations is an extensively studied topic of great complexity. Human population history is certainly not purely phylogenetic, or tree-like[1], as genetic admixture, mediated by processes such as migrations, expansions, intermarriage, trade, or slavery, have played an important role in shaping human history[2]. There is, however, a strong hierarchical signal that can be hypothesized as phylogeny in both genetic[3,4] and cultural (especially linguistic) data[5,6]. It is worth noting that even using such terms as "genetic admixture" and "horizontal gene flow" implies an assumption of an underlying tree-like model[7]. Recently developed phylogenetic methods applied to both genetic[8,9] and linguistic data[10] allow us to visualize evolutionary history of populations using a bifurcating tree with horizontal links ("admixture edges"), accounting for both population splits and mixtures.

Today, no unified picture of modern human evolution based on genetic data is available, as studies that infer human population history have used different types of genetic markers, from "classical polymorphisms" (such as AB0 blood groups and protein allomorphisms) and uniparental markers (the mitochondrial DNA and the non-recombining portion of the Y chromosome) to genome-wide allele frequency data and data based on whole-genome sequencing[11]. Moreover, individual studies only partially overlap taxonomically. Even the largest published tree (267 populations) based on genome-wide data[12] lacks several population groups important for a comprehensive description of human population history on a global scale (e.g., populations of N Africa, Anatolia, Balkans, E Europe, Indonesia, N Asia, Beringia, and N America). A recent meta-analysis of human genomic diversity projects[13] has also pointed to the lack of several key population groups (e.g., Hadza, Sandawe, Fulani, Chadic speakers, Australian Aboriginals, populations of Indonesia, Polynesia, and Northern America).

The language phylogenies published to date include up to 542 language varieties[14] but usually cover just one language family each (mostly Bantu, Indo-European, or Austronesian). Formal attempts to reconstruct genealogical relationships between languages beyond the level of the families have been rare so far[15,16], and nearly all of the proposed linguistic macrofamilies such as Eurasiatic/Nostratic[17–19], Indo-Pacific[20], and Amerind[21] are considered controversial[16].

Although a large body of comparative data currently exists for a phylogenetic synthesis, integration of all kinds of raw data using a "supermatrix approach" (or "total evidence approach"[22]) remains unfeasible for the human population, particularly due to the distance-based (instead of character-based) nature of some source data and lack of widely overlapping datasets. In light of these problems, a possible strategy is to focus on published

[1]Department of Zoology, Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic. [2]Center for Theoretical Study, Charles University and Academy of Sciences of the Czech Republic, Prague, Czech Republic. Correspondence and requests for materials should be addressed to P.D. (email: dudapa01@.gmail.com)

("source") trees, adopting the "supertree approach" (or "taxonomic congruence approach"[23]). The primary application of supertrees is to summarize existing phylogenetic hypotheses in a form of a synthetic consensus which can be used to identify and evaluate topological conflicts caused by incongruent or missing data[24]. In the "matrix representation with parsimony" (MRP) method[25,26], each source tree is converted into a matrix of additive binary characters; the individual matrices are eventually merged into a single character matrix which is then analyzed by the maximum parsimony (MP) method to obtain a composite phylogeny. The resulting supertree is analogous to a consensus tree when the source trees have different sets of taxa[24].

The aims of this study are:

(1) to provide a well corroborated phylogenetic hypothesis on human population group-level relationships, based on both genetic and linguistic data;
(2) to assess for the first time the utility of admixture plots, produced by STRUCTURE, FRAPPE, and ADMIXTURE software, as sources of hierarchical information during the supertree construction;
(3) to assess the stability of the inferred supertree topology and to identify populations whose phylogenetic position is particularly unstable;
(4) to compare the topologies based on genetic and linguistic data, and evaluate their relative influence on the supertree topology; and
(5) to test for congruence between proposed linguistic groupings (language families and macrofamilies) and supertree topology and to infer the relationships between language families by constraining the supertree topology with linguistic classification.

Results and Discussion

Supertree construction. Altogether 257 source trees (obtained by using both distance-based and character-based methods) and 44 admixture plots from 200 published studies (1990–2014) contributed to the resulting supertree dataset. They included trees based on genomic data, including both genome-wide allele frequency data and whole-genome sequences (51 trees from 33 studies), genetic trees based on autosomal data (26, 19), Y-chromosomal data (9, 9), mtDNA (25, 20), human leukocyte antigen (HLA) system (75, 57), "classical polymorphisms" (27, 8), language trees based on lexical or structural data (44, 33), admixture plots based on genomic data (43, 36), and one admixture plot based on linguistic structural data.

The resulting supertree dataset (unpublished) included 973 populations and 5 great apes or archaic hominins that featured in the source trees (see Supplementary methods). Two datasets were then created based on restricted samples of this dataset. The first dataset consisted of 186 populations and included all world regions and major linguistic groups that are reasonably well represented throughout the source trees ("representative dataset" hereafter) (Supplementary Table S2). The second dataset consisted of 52 populations from the Human Genome Diversity Project (HGDP) panel[3] that are best represented throughout the source trees, plus three additional populations to represent Australia, Micronesia and Polynesia ("HGDP dataset" hereinafter).

To investigate robustness of the inferred supertree topology, we used a method inspired by the "sensitivity analysis" of Wheeler[27]. The analysis was carried out by successively reweighting and rerooting the data partitions, adjusting an effect of different data partitions on the resulting supertree topology. In this study, a sensitivity analysis has been used for the first time for supertree inference. We used 16 sets of parameters for both representative and HGDP samples, based on combinations of four weighting and four rooting schemes (see Methods).

Sixteen sets of the most parsimonious (MP) trees, recovered in a sensitivity analysis of the representative dataset, were analyzed using the *Iter*PCR method[28,29] in order to identify unstable ("wildcard"[30]) taxa which cause large polytomies in the supertree, hampering the interpretation of phylogenetic results (see Methods). Four wildcards that decreased resolution of the supertree by five or more nodes (see below) were excluded from the dataset, and the pruned version of the representative dataset (182 populations) was used for subsequent analyses.

Supertree topologies and topological stability. Given the expected conflict across different types of data, the resulting supertree topologies based on the representative dataset (Fig. 1a–d and Supplementary Figs S1–S16) were surprisingly well-resolved (Supplementary Table S3).The parameter set 1.A maximizes congruence between data partitions, providing the shortest supertree with the highest CI and RI values (Supplementary Table S3). The resulting supertree topologies are, overall, robust to parameter set changes. Similarity of the resulting supertrees measured by subtree prune and regraft (SPR) distances is 99–74% (Supplementary Table S4a,b). The contribution of admixture plots to the resulting supertree topology was relatively small. The topology of the combined supertree based on parameter set 1.A, where the effect of admixture plots was maximized, was 85% similar to the parameter set 2.B where the effect of admixture plots was minimized while all other data partitions played an equal role (Supplementary Table S4a,b; Supplementary Figs S1 and S6). The admixture plots alone provided a more symmetrical topology with three superclades: the basal African, followed by the N African–W Eurasian, and the Eastern superclade (Supplementary Fig. S19). However, the hierarchical clustering of populations in admixture plots is not always comparable to the order of branching events in human population history. The early divergence of some populations (e.g., Hadza, Dogon, Basque, and Tibetan) implied by the admixture plots could reflect isolation and random genetic drift rather than early divergence. Some relationships probably reflect relatively recent admixture (e.g., Bantu populations of S Africa[4]).

In all 16 topologies provided by the representative dataset (Fig. 1 and Supplementary Figs S1–S16) sub-Saharan Africa is located nearest to the root of the tree, followed by N Africa, the Near East, Europe, S and C Asia, Oceania, E Asia and America. The general branching order is largely consistent with the previously published global human population-level phylogenies, despite major differences in sampling and phylogenetic inference methods used[3,4,12,31]. All 16 topologies based on representative dataset agreed upon the most basal position

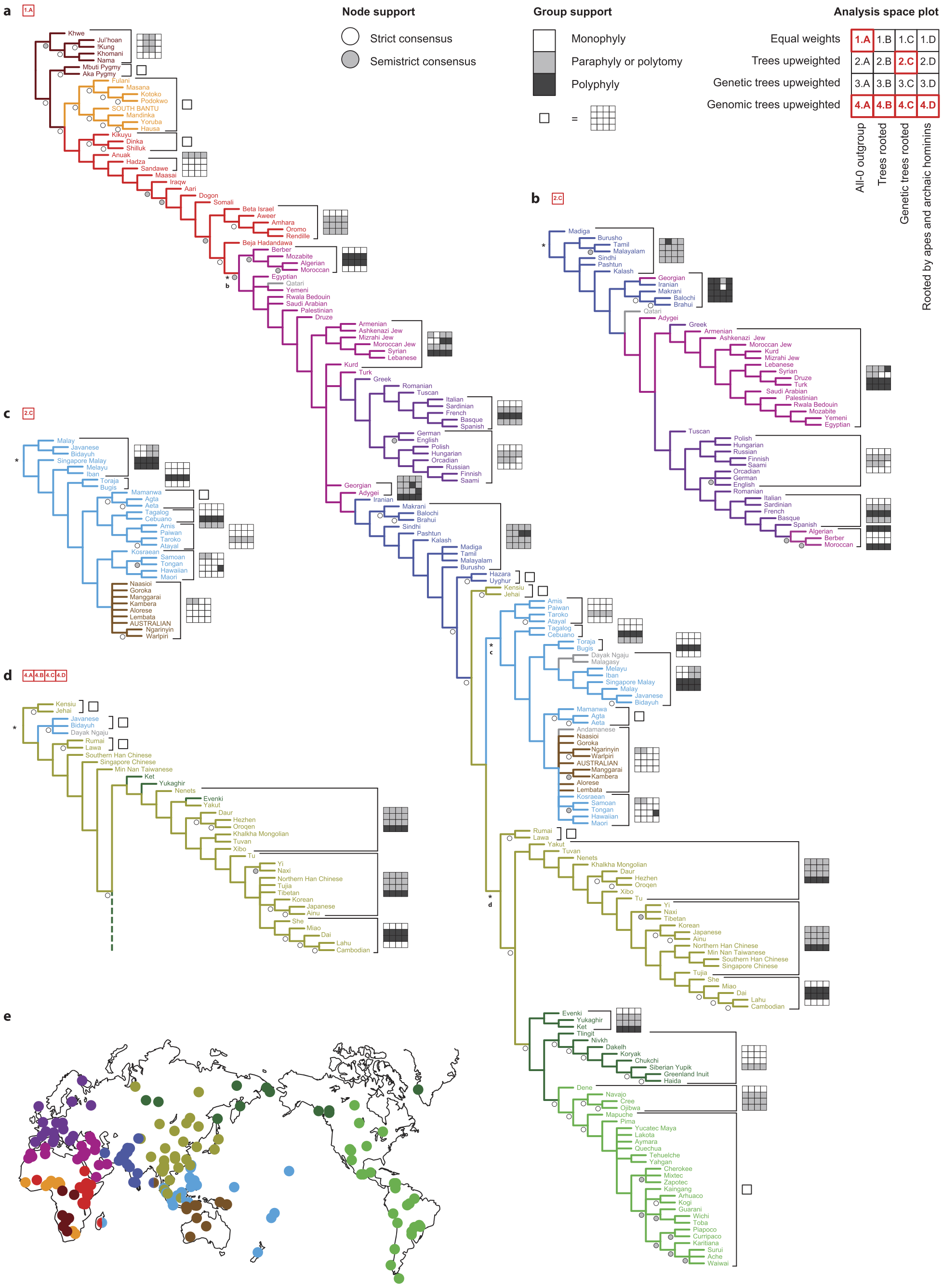


Figure 1. (**a**) Semistrict consensus supertree of 186 human populations (outgroups not shown) based on the representative dataset and parameter set 1.A of the sensitivity analysis (all data partitions were weighted equally and all sources were considered rooted). SOUTH BANTU = Ndebele + Swati + Xhosa + Zulu (often occurred as a composite population in the source trees); AUSTRALIAN consists of Australian Aboriginal populations of unspecified ethnic origin. The color code corresponds to the recovered monophyletic or paraphyletic groups of populations. The wildcard taxa (Qatari, Andamanese, Malagasy, Dayak Ngaju) are displayed (in gray) in the most basal of all positions they acquired when included into the dataset, but were not taken into account when assessing node and group support. The circles indicate presence of the nodes in the strict (white) and semistrict (gray) consensus of 16 supertrees derived from the sensitivity analysis (a circle is absent if the respective node is absent even in the semistrict consensus). The analysis space plots (square grids) describe presence of the selected clades/groups in the supertree under individual parameter sets as either: a monophyletic clade (white); a paraphyletic group or an unresolved section compatible with monophyly or paraphyly (gray); a polyphyletic assemblage (black). Completely white grids (= the group present under all parameter sets) are substituted by small white squares. (**b, c**) Alternative topology for the N African–W Eurasian assemblage and the Sahul–Oceanian clade as recovered in parameter set 2.C. (**d**) Alternative topology for the E Asia clade as recovered in parameter sets 4.A–4.D. The nodes where the alternative topologies (**b, c, d**) begin in the supertree 1.A (**a**) are denoted by asterisks. (**e**) Geographic locations of 186 human populations plotted on the world map using QGIS v.2.8 (the color code corresponds to the trees).

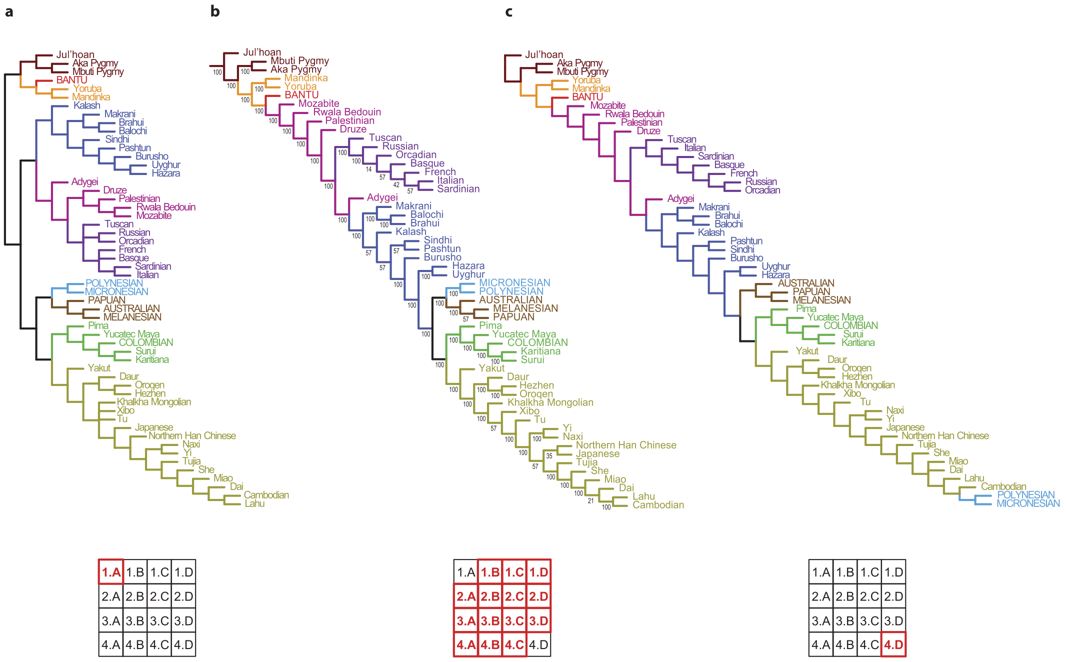


Figure 2. (**a**) Semistrict consensus supertree of 55 human populations based on HGDP dataset and parameter set 1.A of the sensitivity analysis. Populations were renamed to correspond to those used in the HGDP panel. BANTU = Kikuyu; POLYNESIAN = Samoan + Maori; MICRONESIAN = Kosraean; MELANESIAN = Naasioi; PAPUAN = Goroka; COLOMBIAN = Piapoco + Curripaco. The color code corresponds to Fig. 1. (**b**) Frequency-differences consensus of 14 supertrees based on parameter sets 1.B–4.C of the sensitivity analysis. (**c**) Semistrict consensus supertree based on parameter set 4.D of the sensitivity analysis. The geographic color code corresponds to Fig. 1.

of S African Khoisan followed by C African Pygmies, and the clade consisting of Fulani Afro-Asiatic (Cushitic) speaking populations as a sister group to Niger-Congo speaking populations (including Bantu). The next paraphyletic section of the supertree included Niger-Congo (Bantu), Nilo-Saharan, and Afro-Asiatic (Cushitic, Omotic, and Semitic) peoples and the click-speaking Hadza and Sandawe hunter-gatherers of E Africa, The clustering of Chadic speaking populations of C Africa with Niger-Congo speaking populations of this region rather than with Afro-Asiatic speaking populations of E Africa was consistent with the previously published genomic trees[3,4]. So was the position of Hadza and Sandawe within the ethno-linguistically heterogeneous E African section of the supertree[4,12]. This section was basal to the monophyletic clade including N African and Eurasian peoples. The latter consisted of the largely unresolved N African–W Eurasian assemblage (N African, Near Eastern, European, and S Asian peoples) and the consistently monophyletic Eastern superclade (Sahul–Oceanian, E Asian, and Beringian–American peoples). The most remarkable differences between individual topologies derived from different parameter sets concerned W Eurasia, Mainland and Island SE Asia and Oceania, and E Asia. In the N African–W Eurasian assemblage, there are highly unstable relationships among its constituent sections (N Africa, Near East, Europe, and S Asia), most of which tend to be para- or even polyphyletic (Fig. 1a,b). In Mainland and Island SE Asia and Oceania, different parameter sets imply a different source of the expansion into the area (either from Taiwan or from Malay Peninsula) and a varying degree of admixture of Austronesians with Sahul–Melanesian peoples (Fig. 1a,c). In E Asia, there is an unstable relationship between populations of E and SE Asia and a highly unstable position of some Siberian peoples (Evenki, Ket, Yukaghir) who were either recovered at the basal position within the E Asian clade or within the Beringian–American clade (Fig. 1d).

Sensitivity analysis of the HGDP dataset produced three distinct topologies (Fig. 2). They were, for the most part, congruent with the supertrees based on the representative dataset, although they included a few clades that were not recovered in the representative-dataset supertrees. The topology recovered under parameter set 1.A (Fig. 2a), was the most symmetrical and included monophyletic superclades as follows: sub-Saharan African (with Khoisan–Pygmy and Bantu–E African subclades), N African–W Eurasian (with S Asian, N African–Near Eastern and European subclades), and Eastern (with Sahul–Oceanian, American, and E Asian subclades). The topologies recovered under parameter sets 2.A–4.C (Fig. 2b) were fully compatible with the representative-dataset supertrees, and in agreement with other studies using similar population samples[3,31], regardless of the tree-building techniques used. In the topology recovered under parameter set 4.D (Fig. 2c), the "Oceania" clade situated in the base of the Eastern superclade in most supertrees, was recovered as polyphyletic. The Sahul–Melanesian subclade remained basal to the rest of the Eastern superclade, while the Micronesian–Polynesian ("Remote Oceanian") subclade was deeply nested within the E Asian populations.

The most important point of conflict among the alternative supertree topologies thus concerned the position of Sahul–Melanesian and Micronesian–Polynesian peoples. Phylogenetic affinities of Sahul–Melanesian peoples varied greatly between the source trees. While in multiple studies, Sahul–Melanesia was placed basally, often as a sister-group to E Eurasia as a whole[3,4,12], in others they were nested deeply within SE Asia[31–33]. These topological conflicts reflect the complex population history of Island SE Asia, from early "out-of-Africa" migration via the "southern route"[34] through later interactions with Mainland SE Asia[9,35] up to the putative "express-train" migration of the Austronesian speakers from Taiwan via the Philippines, Greater and Lesser Sunda Islands, and Melanesia to Micronesia and Polynesia[36–38]. The phylogenetic placement of Sahul–Melanesia is further complicated by possible gene flow from India to Australia around the mid-Holocene[39].

The supertree topology is notably pectinante in agreement with the previously published global human population-level phylogenies[3,12,31]. There were just a few apparent major radiations, namely, Bantu and related sub-Saharan populations (Fig. 1a), European or W Eurasian (Fig. 1a,b), SE Asian–Oceanian (with or without the Sahul–Melanesian peoples) (Fig. 1a,c), E Asian (Fig. 1a,d), and Beringian–American. Individual small clades or even individual terminal taxa tended to branch off from the major migration route in E Africa, Near East, and S Asia. This topology is consistent with a serial founder effect model, which suggested that human populations have remained in the locations they first colonized after the out-of-Africa expansion, exchanging migrants only at a low rate with their immediate neighbors, until the long-range migrations began to happen.

Wildcard taxa. Twenty-four populations, either terminal taxa or small clades, were identified as wildcards in topologies recovered under one or more parameter sets of the sensitivity analysis (Supplementary Table S5). The populations responsible for the greatest loss of resolution (5 nodes or more) throughout the sensitivity analyses were Andamanese (a wildcard taxon in 14 parameter sets, decreasing resolution by 1–21 nodes; see below), Malagasy (12: 3–23; see below), Dayak Ngaju (2: 1 and 10; identified as either Island or Mainland SE Asians), and Qatari (2: 21; highly unstable position within N African–W Eurasian section of the supertree).

The unstable position of some populations provides clues about conflicts within the dataset, which reflects either the paucity of data or complex population history of the peoples in question. For example, the unstable position of Malagasy reflects a relatively recent (ca. 1,200 ya) migration of Austronesian-speaking people across the Indian Ocean, followed by admixture with E Africans. While linguistic evidence places Malagasy language within Barito group of W Malayo-Polynesian (Austronesian) languages[36], Malagasy population exhibit genetic affinities to both SE Asian and E African populations[40].

The case of Andaman Islanders is much more complicated. They were recovered either as Sahul–Melanesians or S Asians, or at the base of E Asia under different parameter sets (Supplementary Figs S17 and S18). Position of Andamanese within the Sahul–Melanesian clade is based on the analysis of structural features of language using a Bayesian clustering algorithm[38]. Initial genetic studies suggested that Andamanese are descendants of an early "out-of-Africa" migration[41], while later studies proposed a more recent S or E Indian origin[42]. Recent studies agree that Andamanese represent an isolated, relatively basal lineage, with possible genetic affinities to both Sahul–Melanesia and S Asia[33,39]. The relatedness of Andamanese to Sahul–Melanesians, particularly the Papuans, has recently been substantiated also by genomic data[43,44].

Assessment of gene–language coevolution. The question of coevolution of genes and languages is considered fundamental but rarely studied by formal phylogenetic methods. Although the genetic and linguistic evolution may often be correlated, the assumption of direct coevolution between genes and languages is evidently misleading[45]. Evolutionary processes shaping genetic diversity are not directly analogous to those shaping linguistic diversity[46] and, consequently, genetic and linguistic data often imply different historical scenarios[47].

The supertree dataset included 45 linguistic source trees and one linguistic admixture plot from 34 studies (compared to 213 genetic/genomic source trees and 43 genomic admixture plots from 170 studies). The 535 (~9%) parsimoniously informative characters based on these sources contributed only marginally to the resulting supertree topology (Supplementary Fig. 21). In fact, only a few language families have so far been analyzed phylogenetically, and hence numerous areas of the supertree included no linguistic data at all. Only a few small clades were supported by linguistic characters. The supertree topology was, in general, dominated by the genetic/genomic data.

In order to test for monophyly of the proposed linguistic macrofamilies we created two datasets based on formal linguistic classifications (Supplementary Table S6) to be both optimized on, and to constrain, the topology of the supertree based on the representative dataset. The first dataset was based on linguistic classification in *Ethnologue*[48] on the level of language families ("*Ethnologue*" hereinafter). The second dataset ("Greenberg–Ruhlen" hereinafter) included additional characters based on linguistic classification by Ruhlen[49] on the level of linguistic macrofamilies, and by Greenberg & Ruhlen[50] on the level of linguistic stocks within the Amerind macrofamily. The hunter–gatherer populations, who speak the languages of neighboring agriculturalist or pastoralist groups as a result of a relatively recent language shift (C African Pygmies, "Negritos" of Malaysia and Philippines) were not scored for linguistic characters (see Methods).

Optimization of the datasets based on linguistic classification on the representative-dataset supertree showed rather poor fit of the classification on the supertree topology (*Ethnologue* and Greenberg–Ruhlen datasets' CI's were 0.27 and 0.25, respectively, for the purely genetic, and 0.31 and 0.28, respectively, for the combined supertree). Within *Ethnologue* dataset, the best fitting language families were Austronesian, S African Khoisan, Afro-Asiatic (especially Semitic languages), and Indo-European (Supplementary Table S7a). Within Greenberg–Ruhlen dataset, the macrofamily which is by far the most consistent with the supertree topology is Amerind, followed by Austric (and its constituent language families Austronesian and Austroasiatic), Afro-Asiatic, Khoisan, and Indo-Hittite. We do not consider the good fit of Amerind on the supertree topology as support for the Amerind hypothesis[21], but rather a consequence of geographic and genetic coherence of the presumably Amerind-speaking populations. The linguistic stocks within the Amerind macrofamily are not consistent with the supertree topology (Supplementary Table S7b). The other controversial linguistic macrofamilies, such as Eurasiatic/Nostratic, Macro-Altaic, Dene–Yeniseian, and Dene–Caucasian, fitted poorly on the supertree

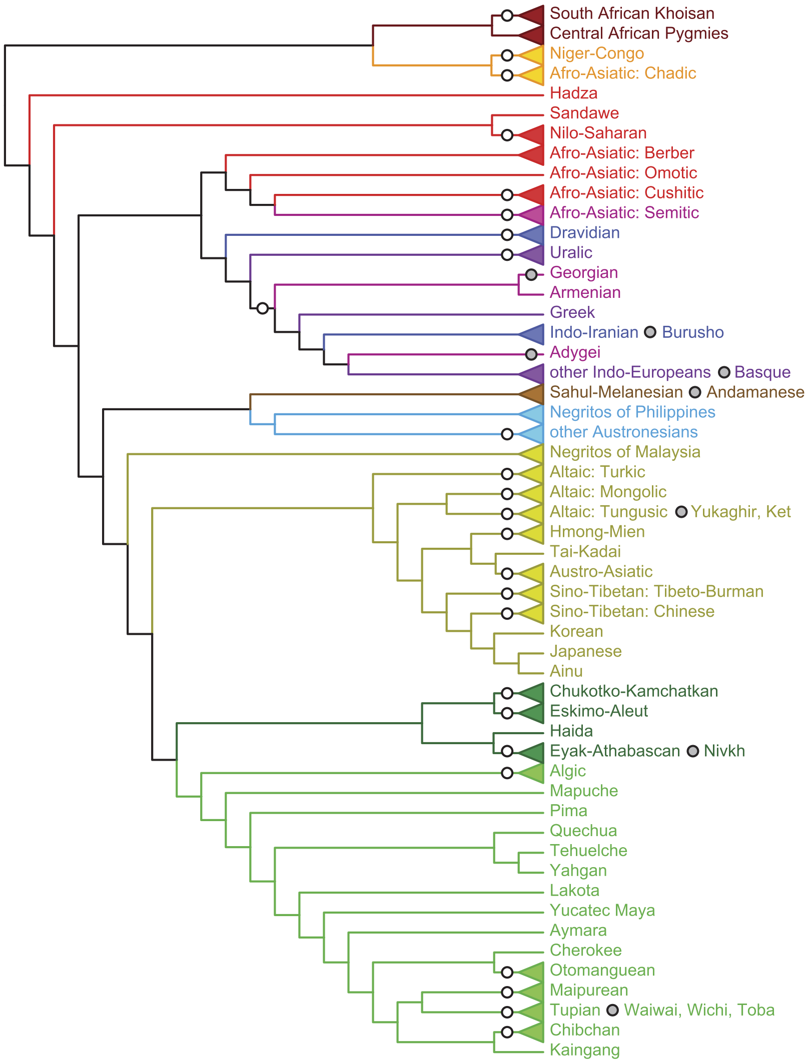


Figure 3. The supertree constrained by *Ethnologue* classification. White circles indicate topological constraints. Grey circles indicate an unconstrained taxon or clade (usually a language isolate) recovered within a constrained one. The geographic color code corresponds to Fig. 1.

topology. The poor fit of Macro-Altaic and the families that constitutes it (especially the Turkic) is in agreement with the fact that there is only a weak unifying genetic signal for the Turkic-speaking populations across Eurasia[51]. The expansion of Turkic languages has probably been largely mediated by language replacements rather than demic expansion.

The supertrees constrained by *Ethnologue* (Fig. 3) and Greenberg–Ruhlen datasets (Supplementary Fig. S22) summarized relationships between groups of populations speaking related languages based on genetic data (congruence between the purely genetic supertree and the supertree constrained by the Greenberg–Ruhlen dataset is illustrated by a tanglegram and by "anticonsensus" trees; Supplementary Figs S22–S24). The *Ethnologue* dataset included only those language families that are relatively non-controversial, and the supertree constrained by *Ethnologue* classification (Fig. 3) is in many respects similar to the combined supertree, although it is more symmetrical. Importantly, it includes several monophyletic clades that are not based on the linguistic topological constraint used. In sub-Saharan Africa, there is a clade including peoples speaking S African Khoisan, Niger-Congo, and Chadic languages. There is also a large clade including peoples speaking Afro-Asiatic (excl. Chadic), Dravidian, Uralic, and Indo-European languages, a group roughly coextensive with the hypothesized Eurasiatic/Nostratic macrofamilies[17–19]; however, the Altaic, Chukotko-Kamchatkan and Eskimo-Aleut-speaking peoples (that have been hypothesized to belong to the Eurasiatic/Nostratic macrofamily as well) are not closely related genetically. The relationships between individual language families of the hypothesized Eurasiatic macrofamily[17,18] (i.e., Dravidian, Kartvelian, Indo-European, Uralic, Altaic, Chukchi-Kamchatkan and Eskimo-Aleut) are largely consistent with those inferred in by Pagel *et al.*[15]. There is, however, no support for monophyly of the hypothesized Eurasiatic languages as a whole (Supplementary Table S7c and Supplementary Fig. S22a). Another large clade includes Altaic, Austric (excl. Austronesian), and Sino-Tibetan languages, together with Korean–Japanese and Ainu. Whereas Ainu was considered related either to Eurasiatic[17,18,49] or Austric[16] macrofamily, Korean and Japanese were seen as distant relatives of Altaic languages[49]. On the contrary, the close relationships between Korean–Japanese–Ainu and Sino-Tibetan peoples have no linguistic basis. The clade which includes Australian, Papuan, Melanesian, and Andamanese populations is somewhat reminiscent of the controversial Indo-Pacific macrofamily[20]. The Na-Dene, Eskimo–Aleut, and Chukotko-Kamchatkan populations are closely related to Amerind, a connection that has no linguistic basis. Within America, there is a conspicuous basal placement of populations of the Southern Cone (Andean languages according to Greenbeg–Ruhlen; Supplementary Fig. S22b), which could be indicative of an early western route used during the initial colonization of Americas.

The hunter-gatherer groups that were deliberately not scored for linguistic characters were recovered outside of the clades they belong to based on their linguistic affiliation. C African Pygmies were recovered as a sister group to S African Khoisan (or within the Khoisan family in the supertree constrained by the Greenberg–Ruhlen classification; Supplementary Fig. S22a), providing evidence for shared ancestry among these geographically diverse groups of hunter-gatherers[4]. E African Hadza (but not Sandawe) were recovered at a basal position within Africa just above the S African Khoisans and C African Pygmies. "Negritos" of Malaysia were placed as a sister group to the whole Eastern superclade, and "Negritos" of Philippines were recovered as a sister group to Austronesian language family (or within the clade including languages of Australia and Indo-Pacific languages in the supertree constrained by the Greenberg–Ruhlen classification; Supplementary Fig. S22a), providing evidence for the hypothesis that the "Negrito" populations represent the descendants of the early migration into the area[34], with lasting genetic affinities to Sahul–Melanesia[39,43,44].

Interestingly, when the supertree topology was constrained to include linguistic-compatible clades, as if the language families were indeed monophyletic, it tended to form more inclusive clades, which are more or less compatible with the proposed linguistic macrofamilies (Eurasiatic/Nostratic, Indo-Pacific, and especially Amerind). It is possible that the "macrofamilies" are consistent genetically and geographically rather than linguistically; however, the possibility that historical linguistics is able to reconstruct the most basal relationships between modern human populations should be re-assessed critically[6].

The supertree can provide a robust framework for studies concerning evolution of culture[52]. Such a framework is needed because most cross-cultural comparative studies published to date used language phylogenies[5,7]. Although language phylogenies provide an excellent proxy for population histories in some regions (e.g., Remote Oceania), this is not universally the case[45,47]. Linguistic data seems to be unable to provide a global tree of human populations due to a limited timescale over which linguistic inference is possible[6]. On the other hand, genetic phylogenies, although global, could be unsuitable for studies of cultural evolution, as the population history they inform of can be older than the cultural traits under investigation. A time-calibrated supertree, incorporating all time "strata" of human evolution (and informed by ancient DNA), is needed to elevate the studies of cultural evolution to a global level.

Methods

Data. Source trees published in peer-reviewed journals, edited volumes and monographs between 1990 and 2013 (most of them post 2007) were collected (including papers "in press" by the end of 2013). Altogether 257 source trees (obtained by using both distance-based and character-based methods) and 44 admixture plots from 200 published studies contributed to the present dataset. Only trees that were inferred by formal phylogenetic methods and based on original analyses were utilized. The protocol for inclusion and rejection of source trees was guided by the issues of sufficient taxonomic coverage and data quality (see Supplementary Methods). In order to ameliorate the problem of data non-independence and duplication, we used a protocol for source-trees retention and exclusion proposed by Bronzati *et al.*[53] (see Supplementary Methods). In addition to the trees, we also utilized admixture plots, produced by software like STRUCTURE, FRAPPE, and ADMIXTURE, as a source of hierarchical information for supertree construction (see the section "Matrix representation with parsimony" and Supplementary Methods).

Taxonomic nomenclature and taxonomic level. To synthesize published phylogenies from different sources, the names of terminal taxa from the source trees were standardized using ISO 639-3 codes from *Ethnologue*[48], a standard, widely recognized taxonomic reference. Information on geographic range of a population in question, sampling location(s) of genotyped individual(s), linguistic affiliation and ethnonyms were utilized in order to standardize the taxonomy among individual sources. Where higher-level taxa (e.g., population or linguistic groupings above the level of ethno-linguistic groups listed in *Ethnologue*) were used in the source studies, they were either replaced by a single population based on information from the original study, or, when this information was insufficient or unavailable, by "type" population(s) (Supplementary Table S1). Lower-level taxa (e.g., local populations or language dialects) took on the names of the corresponding ethno-linguistic groups listed in *Ethnologue* (see Supplementary Methods). Populations of well-known recent mixed ancestry (e.g., "African American", "US Hispanic", "Cape Mixed Ancestry"), colonial populations (e.g., Boer), creole languages (e.g., Haitian), and loosely specified higher level taxa (e.g., "African", "Native North American") were not included. The only exceptions were Australian Aboriginals of unspecified ethnic population origin that were merged together and analyzed as a single terminal taxon, named "AUSTRALIAN" (see Supplementary methods).

Population samples. Two datasets were created. The first dataset of 186 populations included all world regions and major linguistic groups that were reasonably well represented throughout the source trees ("representative dataset") (Supplementary Table S2). The representative dataset included 5,987 parsimoniously informative characters. The second dataset consisted of 51 populations from the Human Genome Diversity Project (HGDP) panel[3], plus three additional populations to represent Australia, Micronesia and Polynesia ("HGDP dataset"). The HGDP dataset included 3,070 parsimoniously informative characters (see Supplementary Methods).

Matrix representation with parsimony. The matrix representation with parsimony (MRP) method[24,25] is based on creating, merging and reanalyzing matrix representations of the source trees: each source tree was converted into a partial matrix of additive binary characters. Taxa descended from a given node were coded as "1" (=present); those that did not were coded as "0" (=absent); all taxa that were not present in the given source trees were coded as "–" (=inapplicable). Each admixture plot was converted into a matrix representation such that each population was coded as "1" (=present) or as "0" (=absent) based on the proportions of individual genotypes attributable to each cluster. Limited attribution to a given cluster (less than 10%) was neglected, and ambiguous sections of a plot (borderline proportions or different proportions in individuals within a single population) were coded as "?" (=unknown). The resulting matrix of additive binary characters was analyzed by the MP method to obtain a tree which corresponds to clustering implied by the admixture plot. The trees based on admixture plots typically contain unresolved sections due to membership of some populations in several clusters, but they still preserve enough valuable branching information. The merged character matrix consisting of matrix representations of trees and admixture plots was analyzed by the MP method to obtain a supertree presented in the form of a strict or semistrict consensus tree.

Phylogenetic analysis. Phylogenetic analyses were performed in TNT ver. 1.1[54] under "new technology search" with search level 10 using sectorial, ratchet, and tree fusing searches, obtaining trees from a 10,000-replicate random addition sequence, followed by additional branch swapping using the tree-bisection and reconnection method (see Supplementary Methods). The datasets were analyzed without any topological constraints (i.e., without any assumptions on geographic regions or language families).

Sensitivity analysis and wildcard taxa identification. To investigate robustness of the inferred supertree topology, we used a method inspired by Wheeler's "sensitivity analysis"[27] (see Supplementary Methods). We used 16 parameter sets (for each population sample in parallel), based on combinations of four weighting and four rooting schemes as follows: either (1) all data partitions were weighted equally, or (2) all trees were upweighted by the factor of 1,000 relative to admixture plots, or (3) genetic/genomic trees were upweighted by the factor of 1,000 relative to language trees and all admixture plots, or (4) genomic trees were upweighted by the factor of 1,000 relative to all remaining data partitions; and either (A) all rooted source trees and admixture plots were treated as rooted (by inserting a hypothetical "all-0" outgroup), or (B) only rooted source trees were treated as rooted, or (C) only rooted genetic/genomic trees were treated as rooted, or (D) only source trees/admixture plots featuring great ape and/or archaic hominin outgroups (*Gorilla gorilla*, *Pan paniscus*, *P. troglodytes*, Denisova hominin, *Homo neanderthalensis*) were treated as rooted. When performing sensitivity analysis on the HGDP dataset, the data partitions were either downweighted as above (1:1,000) or completely deactivated to test whether the weighting scheme was sufficient to minimize the effect of a data partition on the resulting topology. Sixteen sets of the most parsimonious trees, recovered in the sensitivity analysis of the representative dataset, were analyzed using *Iter*PCR script[28] as implemented in TNT[29], to improve resolution of the consensus tree by identifying taxa of unstable positions ("wildcard taxa"[30]). Alternative positions of the identified wildcards were investigated using pruned strict consensus (*nelsen//*) in TNT. Four wildcards that decreased resolution of the supertree by five or more nodes were excluded from the dataset, and the pruned version of the representative dataset (182 populations) was used for subsequent analyses (Supplementary Methods).

Linguistic constraints. In order to infer the relationships of language families and macrofamilies, we created two datasets based on linguistic classification. The first dataset included 37 parsimoniously informative characters based on *Ethnologue* classification[48] on the level of language families. The second dataset included additional 26 parsimoniously informative characters based on on classification by Ruhlen[49] and Greenberg & Ruhlen[50] on the level of linguistic macrofamilies and linguistic stocks within the Amerind macrofamily; Supplementary Table S6; Supplementary Methods). Characters based on *Ethnologue*[48], Ruhlen[49] and Greenberg & Ruhlen[50] were fully congruent, with no hard conflict between them. Hunter–gatherer populations, speaking languages of neighboring groups, were scored as "unknown" ("?") in both datasets. These included C African Pygmies (Mbuti Pygmy, Aka Pygmy) who speak Niger-Congo or Nilo-Saharan languages[55], the "Negritos" of Malaysia (Jehai and Kensiu) who speak Austro-Asiatic languages[56,57], and the "Negritos" of Philippines (Agta, Aeta, and Mamanwa) who speak Malayo-Polynesian (Austronesian) languages[58]. Similarly, Ashkenazi Jews who used to speak Indo-European (Germanic) Yiddish were not scored for linguistic characters. (Supplementary Methods). Language-constrained supertrees (Fig. 3; Supplementary Fig. S22a,b) were inferred by analyzing all data partitions rooted by "all-0" outgroup together with the *Ethnologue* and Greenberg–Ruhlen datasets; the data partitions based on linguistic sources and the datasets based on linguistic classification were upweighted by the factor of 1,000 relative to genetic data partitions.

Supertrees comparison and phylogenetic signal. The resulting supertree topologies were compared using the SPR distance measure (*sprdiff*) and the "anticonsensus" measure (*tcomp*) in TNT software. Topology of the supertree constrained by the Greenberg–Ruhlen classification was compared with the purely genetic supertree using a tanglegram computed in Dendroscope ver. 3.2.10[59]. In order to assess the support for proposed linguistic groupings (macrofamilies, stocks, and families), consistency index (CI) and retention index (RI) values were calculated in Mesquite ver. 3.02[60] for each character in the linguistic classification datasets optimized onto the purely genetic and combined supertree topologies (based on parameter set 1.A, see the section "Sensitivity analysis and wildcard taxa identification"). The resulting CI values were compared to the minimum possible CI values (for a binary character, $CI_{min} = 1/N$, where N taxa were scored positively for presence of a character), which made the values directly comparable for language families represented by different number of populations (Supplementary Methods).

Plotting. Plotting of sampling locations on the world map was performed using an open source software QGIS v.2.8 (http://qgis.org/en/site/) with open-source map.

References

1. Templeton, A. R. Biological races in humans. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* **44,** 262–271, doi: http://dx.doi.org/10.1016/j.shpsc.2013.04.010 (2013).
2. Hellenthal, G. *et al.* A Genetic Atlas of Human Admixture History. *Science* **343,** 747–751, doi: 10.1126/science.1243518 (2014).
3. Li, J. Z. *et al.* Worldwide human relationships inferred from genome-wide patterns of variation. *Science* **319,** 1100–1104, doi: 10.1126/science.1153717 (2008).
4. Tishkoff, S. A. *et al.* The genetic structure and history of Africans and African Americans. *Science* **324,** 1035–1044, doi: 10.1126/science.1172257 (2009).
5. Mace, R. & Holden, C. J. A phylogenetic approach to cultural evolution. *Trends in ecology & evolution* **20,** 116–121, doi: 10.1016/j.tree.2004.12.002 (2005).
6. Pagel, M. Human language as a culturally transmitted replicator. *Nature Reviews Genetics* **10,** 405–415, doi: 10.1038/nrg2560 (2009).
7. Mace, R. & Jordan, F. M. Macro-evolutionary studies of cultural diversity: a review of empirical studies of cultural transmission and cultural adaptation. *Philosophical Transactions of the Royal Society B-Biological Sciences* **366,** 402–411, doi: 10.1098/rstb.2010.0238 (2011).
8. Pickrell, J. K. & Pritchard, J. K. Inference of population splits and mixtures from genome-wide allele frequency data. *PLOS Genetics* **8,** doi: 10.1371/journal.pgen.1002967 (2012).
9. Lipson, M. *et al.* Reconstructing Austronesian population history in Island Southeast Asia. *Nature Communications* **5,** 7, doi: 10.1038/ncomms5689 (2014).
10. Nelson-Sathi, S. *et al.* Networks uncover hidden lexical borrowing in Indo-European language evolution. *Proceedings of the Royal Society B-Biological Sciences* **278,** 1794–1803, doi: 10.1098/rspb.2010.1917 (2011).
11. Veeramah, K. R. & Hammer, M. F. The impact of whole-genome sequencing on the reconstruction of human population history. *Nature Reviews Genetics* **15,** 149–162, doi: 10.1038/nrg3625 (2014).
12. Pemberton, T. J., DeGiorgio, M. & Rosenberg, N. A. Population structure in a comprehensive genomic data set on human microsatellite variation. *G3-Genes Genomes Genetics* **3,** 891–907, doi: 10.1534/g3.113.005728 (2013).
13. Shriner, D., Tekola-Ayele, F., Adeyemo, A. & Rotimi, C. N. Genome-wide genotype and sequence-based reconstruction of the 140,000 year history of modern human ancestry. *Scientific Reports* **4,** doi: 10.1038/srep06055 (2014).
14. Currie, T. E., Meade, A., Guillon, M. & Mace, R. Cultural phylogeography of the Bantu Languages of sub-Saharan Africa. *Proceedings of the Royal Society B-Biological Sciences* **280,** 8, doi: 10.1098/rspb.2013.0695 (2013).
15. Pagel, M., Atkinson, Q. D., Calude, A. S. & Meade, A. Ultraconserved words point to deep language ancestry across Eurasia. *Proceedings of the National Academy of Sciences of the United States of America* **110,** 8471–8476, doi: 10.1073/pnas.1218726110 (2013).
16. Jäger, G. Support for linguistic macrofamilies from weighted sequence alignment. *Proceedings of the National Academy of Sciences* **112,** 12752–12757, doi: 10.1073/pnas.1500331112 (2015).
17. Greenberg, J. H. *The Eurasiatic Language Family: Indo-European and Its Closest Relations*. Volume I: Grammar. Vol. 1 (Stanford University Press, 2000).
18. Greenberg, J. H. *Indo-European and its Closest Relatives: The Eurasiatic Language Family* Volume II: Lexicon. Vol. 2 (Stanford University Press, 2002).
19. Bomhard, A. R. & Kerns, J. C. *The Nostratic macrofamily: a study in distant linguistic relationship*. Vol. 74 (Walter de Gruyter, 1994).
20. Greenberg, J. H. In *Current trends in linguistics* Vol. 8 (ed T. A. Sebeok) 807–871 (Mouton, 1971).
21. Greenberg, J. H. *Language in the Americas*. (Stanford University Press, 1987).
22. Kluge, A. G. A concern for evidence and a phylogenetic hypothesis of relationships among Epicrates (Boidae, Serpentes). *Systematic Zoology* **38,** 7–25, doi: 10.2307/2992432 (1989).
23. Pisani, D. & Wilkinson, M. Matrix representation with parsimony, taxonomic congruence, and total evidence. *Systematic Biology* **51,** 151–155, doi: 10.1080/106351502753475925 (2002).
24. Bininda-Emonds, O. R. P. The evolution of supertrees. *Trends in ecology & evolution* **19,** 315–322, doi: 10.1016/j.tree.2004.03.015 (2004).
25. Baum, B. R. Combining trees as a way of combining data sets for phylogenetic inference, and the desirability of combining gene trees. *Taxon* **41,** 3–10, doi: 10.2307/1222480 (1992).
26. Ragan, M. A. Phylogenetic Inference Based on Matrix Representation of Trees. *Molecular Phylogenetics and Evolution* **1,** 53–58, doi: 10.1016/1055-7903(92)90035-f (1992).
27. Wheeler, W. C. Sequence alignment, parameter sensitivity, and the phylogenetic analysis of molecular data. *Systematic Biology* **44,** 321–331, doi: 10.2307/2413595 (1995).
28. Pol, D. & Escapa, I. H. Unstable taxa in cladistic analysis: identification and the assessment of relevant characters. *Cladistics* **25,** 515–527, doi: 10.1111/j.1096-0031.2009.00258.x (2009).
29. Goloboff, P. A. & Szumik, C. A. Identifying unstable taxa: Efficient implementation of triplet-based measures of stability, and comparison with Phyutility and RogueNaRok. *Molecular phylogenetics and evolution* **88,** 93–104 (2015).
30. Nixon, K. C. & Wheeler, Q. D. In *Extinction and Phylogeny* (eds M. J. Novacek & Q. D. Wheeler) 119–143 (Columbia University Press, 1993).
31. Abdulla, M. A. *et al.* Mapping human genetic diversity in Asia. *Science* **326,** 1541–1545, doi: 10.1126/science.1177074 (2009).
32. Ayub, Q. *et al.* Reconstruction of human evolutionary tree using polymorphic autosomal microsatellites. *American Journal of Physical Anthropology* **122,** 259–268, doi: 10.1002/ajpa.10234 (2003).
33. Chaubey, G. & Endicott, P. The Andaman Islanders in a regional genetic context: Reexamining the evidence for an early peopling of the archipelago from South Asia. *Human Biology* **85,** 153–171 (2013).
34. Macaulay, V. *et al.* Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Science* **308,** 1034–1036, doi: 10.1126/science.1109792 (2005).
35. Jinam, T. A. *et al.* Evolutionary history of continental Southeast Asians:"Early train" hypothesis based on genetic analysis of mitochondrial and autosomal DNA data. *Molecular Biology and Evolution* **29,** 3513–3527, doi: 10.1093/molbev/mss169 (2012).
36. Gray, R. D., Drummond, A. J. & Greenhill, S. J. Language phylogenies reveal expansion pulses and pauses in Pacific settlement. *Science* **323,** 479–483, doi: 10.1126/science.1166858 (2009).
37. Friedlaender, J. S. *et al.* The genetic structure of Pacific islanders. *PLOS Genetics* **4,** doi: 10.1371/journal.pgen.0040019 (2008).
38. Reesink, G., Singer, R. & Dunn, M. Explaining the linguistic diversity of Sahul using population models. *PLOS Biology* **7,** doi: 10.1371/journal.pbio.1000241 (2009).

39. Pugach, I., Delfin, F., Gunnarsdottir, E., Kayser, M. & Stoneking, M. Genome-wide data substantiate Holocene gene flow from India to Australia. *Proceedings of the National Academy of Sciences of the United States of America* **110,** 1803–1808, doi: 10.1073/pnas.1211927110 (2013).
40. Regueiro, M. *et al.* Austronesian genetic signature in East African Madagascar and Polynesia. *Journal of Human Genetics* **53,** 106–120, doi: 10.1007/s10038-007-0224-4 (2008).
41. Thangaraj, K. *et al.* Genetic affinities of the Andaman Islanders, a vanishing human population. *Current Biology* **13,** 86–93, doi: 10.1016/s0960-9822(02)01336-2 (2003).
42. Reich, D., Thangaraj, K., Patterson, N., Price, A. L. & Singh, L. Reconstructing Indian population history. *Nature* **461,** 489–U450, doi: 10.1038/nature08365 (2009).
43. Aghakhanian, F. *et al.* Unravelling the Genetic History of Negritos and Indigenous populations of Southeast Asia. *Genome Biol. Evol.* **7,** 1206–1215, doi: 10.1093/gbe/evv065 (2015).
44. Basu, A., Sarkar-Roy, N. & Majumder, P. P. Genomic reconstruction of the history of extant populations of India reveals five distinct ancestral components and a complex structure. *Proceedings of the National Academy of Sciences*, doi: 10.1073/pnas.1513197113 (2016).
45. Pakendorf, B. Coevolution of languages and genes. *Current Opinion in Genetics & Development* **29,** 39–44, doi: 10.1016/j.gde.2014.07.006 (2014).
46. Creanza, N. *et al.* A comparison of worldwide phonemic and genetic variation in human populations. *Proceedings of the National Academy of Sciences of the United States of America* **112,** 1265–1272, doi: 10.1073/pnas.1424033112 (2015).
47. Steele, J. & Kandler, A. Language trees not equal gene trees. *Theory in Biosciences* **129,** 223–233, doi: 10.1007/s12064-010-0096-6 (2010).
48. Lewis, M., Simons, G. & Fennig, C. *Ethnologue: Languages of the World, Seventeenth Edition*. (SIL international Dallas, TX, 2013).
49. Ruhlen, M. *Guide to the World's Languages: Classification*. Vol. 1 (Stanford University Press, 1991).
50. Greenberg, J. H. & Ruhlen, M. *An Amerind Etymological Dictionary*. (Stanford University, Department of Anthropological Sciences, 2007).
51. Yunusbayev, B. *et al.* The genetic legacy of the expansion of Turkic-speaking nomads across Eurasia. *PLOS Genetics* **11,** 24, doi: 10.1371/journal.pgen.1005068 (2015).
52. Peoples, H. C., Duda, P. & Marlowe, F. W. Hunter-Gatherers and the Origins of Religion. *Human Nature*, 1–22, doi: 10.1007/s12110-016-9260-0 (2016).
53. Bronzati, M., Montefeltro, F. C. & Langer, M. C. A species-level supertree of Crocodyliformes. *Historical Biology* **24,** 598–606, doi: 10.1080/08912963.2012.662680 (2012).
54. Goloboff, P. A., Farris, J. S. & Nixon, K. C. TNT, a free program for phylogenetic analysis. *Cladistics* **24,** 774–786 (2008).
55. Bahuchet, S. Changing language, remaining pygmy. *Human Biology* **84,** 11–43 (2012).
56. Burenhult, N., Kruspe, N. & Dunn, M. In *Dynamics of Human Diversity: The Case of Mainland Southeast Asia* (ed N. J. Enfield) Ch. 11, 257–275 (Pacific Linguistics, 2011).
57. Dunn, M., Kruspe, N. & Burenhult, N. Time and place in the prehistory of the Aslian languages. *Human Biology* **85,** 383–399 (2013).
58. Reid, L. A. Who are the Philippine negritos? Evidence from language. *Human Biology* **85,** 329–358 (2013).
59. Huson, D. H. & Scornavacca, C. Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. *Systematic Biology* **61,** 1061–1067, doi: 10.1093/sysbio/sys062 (2012).
60. Maddison, W. P. & Maddison, D. R. Mesquite: a modular system for evolutionary analysis v. 3.02 (2015).

Acknowledgements

This research was supported by Grant Agency of the University of South Bohemia (042/2013/P; 140/2013/P). We would like to thank Olaf R.P. Bininda-Emonds for an extensive discussion of the supertree method and to Phillip Endicott, Brian P. McEvoy, Qiaomei Fu, Diana M. Morlote, and Robert S. Walker for providing the source trees. We thank Pavel Flegontov for critical comments on an earlier version of the manuscript and to Conor Redmont for proofreading the text.

Author Contributions

P.D. and J.Z. conceived and designed the study. P.D. searched the literature and collected the data. P.D. and J.Z. analyzed the data. P.D. prepared the artworks. P.D. and J.Z. wrote the paper. Both authors reviewed and approved the paper.

Additional Information

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Duda, P. *et al.* Human population history revealed by a supertree approach. *Sci. Rep.* **6**, 29890; doi: 10.1038/srep29890 (2016).

Chapter II

Hunter-gatherers and the origins of religion

Peoples, H. C., **Duda, P.**, & Marlowe, F. W.

Human Nature-an Interdisciplinary Biosocial Perspective, 27(3), 261-282.

doi:10.1007/s12110-016-9260-0

Hum Nat (2016) 27:261–282
DOI 10.1007/s12110-016-9260-0

Hunter-Gatherers and the Origins of Religion

Hervey C. Peoples[1,2] · Pavel Duda[3] ·
Frank W. Marlowe[1]

Abstract Recent studies of the evolution of religion have revealed the cognitive underpinnings of belief in supernatural agents, the role of ritual in promoting cooperation, and the contribution of morally punishing high gods to the growth and stabilization of human society. The universality of religion across human society points to a deep evolutionary past. However, specific traits of nascent religiosity, and the sequence in which they emerged, have remained unknown. Here we reconstruct the evolution of religious beliefs and behaviors in early modern humans using a global sample of hunter-gatherers and seven traits describing hunter-gatherer religiosity: animism, belief in an afterlife, shamanism, ancestor worship, high gods, and worship of ancestors or high gods who are active in human affairs. We reconstruct ancestral character states using a time-calibrated supertree based on published phylogenetic trees and linguistic classification and then test for correlated evolution between the characters and for the direction of cultural change. Results indicate that the oldest trait of religion, present in the most recent common ancestor of present-day hunter-gatherers, was animism, in agreement with long-standing beliefs about the fundamental role of this trait. Belief in an afterlife emerged, followed by shamanism and ancestor worship. Ancestor spirits or high gods who are active in human affairs were absent in early humans, suggesting a deep history for the egalitarian nature of hunter-gatherer societies. There is a significant positive relationship between most characters investigated, but the trait "high gods" stands apart, suggesting that belief in a single creator deity can emerge in a society regardless of other aspects of its religion.

Electronic supplementary material The online version of this article (doi:10.1007/s12110-016-9260-0) contains supplementary material, which is available to authorized users.

✉ Hervey C. Peoples
hervey@humanquestion.com

[1] Department of Archaeology and Anthropology, University of Cambridge, Cambridge, UK

[2] P.O. Box 8850, Longboat Key, FL 34228, USA

[3] Department of Zoology, University of South Bohemia, České Budějovice, Czech Republic

Keywords Religion · Evolution · Hunter-gatherers · Animism · High gods · Cultural phylogenetics

Religion is unique to humans. Belief in supernatural agents and the entailed religious practices occur in virtually all human cultures (Brown 1991; Johnson 2005; Murdock 1965; Murdock and White 1980). The universality of religion across human societies (Brown 1991) suggests a deep evolutionary past. The human ability to create complex culture ("cultural capacity") may be older than the first split in the modern human lineage (Lind et al. 2013), and some aspects of religion may have been present before the appearance of anatomically modern humans (Rossano 2006). Religion has generally been assumed to have emerged among anatomically modern humans in Africa during the Upper Paleolithic, and to have played a vital role in the subsequent out-of-Africa expansion (Balme et al. 2009; Rossano 2009a).

Explanations for a natural emergence of religion have been debated for hundreds, if not thousands, of years (Martin and Wiebe 2012; Wiebe 2008). That debate has now turned to empirical approaches and testable hypotheses, many of them grounded in the framework of evolutionary theory (Alcorta and Sosis 2005; Atran and Norenzayan 2004; Barrett and Lanman 2008; Boyer and Bergstrom 2008; Irons 2001; Rossano 2006; Tremlin 2006). During the past decade, evolutionary psychologists have identified and described the activity of cognitive biases that enable us to accept the counter-intuitive concepts and beliefs of religion (Atran and Henrich 2010; Barrett and Lanman 2008; Tremlin 2006). Research into the dynamics of religion has revealed the nature of ritual behavior to promote high levels of cooperation (Atran and Henrich 2010; Fischer et al. 2013; Sosis and Ruffle 2003; Xygalatas et al. 2013). The contribution of belief in morally punishing high gods to enhancing prosociality and ensuring growth and stabilization of society has also been demonstrated (Norenzayan and Shariff 2008; Norenzayan et al. 2016; Peoples and Marlowe 2012).

This research has often focused on characteristics of the large prosocial religions that have emerged during the 10,000–12,000 years since the advent of agriculture (Matthews 2012; Norenzayan 2013). Little attention has been paid to the religion of hunter-gatherers whose religious beliefs and behaviors have been evolving during the vast majority of human history (Lee and DeVore 1968). Despite established speculations about various beliefs and behaviors that may represent an original form of religion, specific traits of nascent religiosity, and the sequence in which they emerged, have remained unknown.

There have been as many attempts to define religion as to explain its origins. Broadly defined, religion is a set of beliefs and behaviors based on a shared worldview that separates the sacred, or supernatural, from the profane (Durkheim 1965 [1912]). In this study of its origins, we view religion as a biocultural adaptation (Alcorta and Sosis 2005; Harris and McNamara 2008; Sanderson and Roberts 2008).

How old are religious concepts and their engendered behaviors? Ritual behavior is widespread among humans today, operating in a variety of social environments, both religious and secular (Brown 1991; Cronk 2005; Spencer 1870). Behavior reminiscent of ritual can be seen in many animals, including the ecstatic "rain dances" and directional drumming of chimpanzees (Goodall 1986; Nishida et al. 1999), which along with bonobos are our closest extant evolutionary kin. It is highly likely that archaic hominins would have exhibited ritualistic behavior in some form, but evidence for nascent religiosity remains difficult to infer from the archaeological record (Rossano 2006, 2009b).

Whether early hominins held religious beliefs prior to the emergence of language is unknown. We should not dismiss the possible presence of non-linguistic religious thought and sentiment among early members of the genus *Homo*. However, a case can be made that transmission of religious concepts from one individual to another requires complex mental imaging, and a capacity for symbolic thought and communication that might include ritual, dancing, singing, gestures, art and ornamentation, as well as language (Deacon and Cashman 2010; Mithen 1998). Some genetic and anatomical changes enabling speech and language can be traced to the most recent common ancestor of Neanderthals and anatomically modern humans about half a million years ago (Barney et al. 2012; Dediu and Levinson 2013; Krause et al. 2007). It has been suggested that rudimentary language could be older than symbolic thought (Barnard 2012).

The archaeological record documents the presence of artifacts and technology dating from ca. 400 kya that would probably have required a level of symbolic communication close to that of language (Zilhão 2007). Early evidence for the processing and use of red ochre, often considered a marker for symbolic behavior, dates to the Middle Pleistocene >400 kya in Africa (Barham 2002) and to >284 kya (Deino and McBrearty 2002) in the presence of Levallois blade technology.

Finds at Pinnacle Point in southern Africa (Marean et al. 2007) demonstrate the use and processing of pigment among anatomically modern humans as early as 165,000 years ago (McDougall et al. 2005). Ochre nodules bearing engraved abstract patterns and perforated shell beads found at Blombos Cave in South Africa, dating to 75,000–100,000 years ago (d'Errico et al. 2005; Henshilwood et al. 2009), suggest symbolically mediated behavior. These and other similar finds lend substantial support to the theory of progressive development of symbolic behavior and complex imagery along with the evolution of modern human anatomy (Barnard 2012; Conard 2010; Deino and McBrearty 2002; d'Errico and Stringer 2011; McBrearty and Brooks 2000; Zilhão 2007). Pleistocene hunter-gatherers would most likely have possessed both the cognitive and communicative skills to share religious beliefs and practices prior to dispersal out of Africa more than 60 kya (Fu et al. 2013; Henn et al. 2012; Johansson 2011; Lind et al. 2013). Although present-day hunter-gatherers are not direct analogues of those early societies and may not be direct, unbroken descendants of ancestral hunter-gatherers, they can provide a window onto traits selected for in the Pleistocene (Marlowe 2005), including traits of early religion.

The uniqueness of "natural" religions of hunter-gatherers, and likely those of our Paleolithic ancestors, cannot be overemphasized when compared with the "world" religions that have emerged along with the advent of agriculture. Many hunter-gatherer societies have little or no concept of religion per se, though a religious dimension often permeates normal activities and is continuous with daily life (Lee 1989). Hunter-gatherer religions are seldom religions of protest or evangelism (Woodburn 1997). Instead, each society focuses on maintaining its unique beliefs and culture, along with a sense of self-worth and the general health and well-being of the group (Woodburn 1997, 2005). Simple egalitarian hunter-gatherer groups generally hold fewer religious beliefs and participate in less ritual (Marlowe 2010) than more complex groups. But hunter-gatherers do have religion, embodied in sacred healing dances and rituals marking life events. Although there is considerable variation in specific religious traits among hunter-gatherer societies, a cross-cultural view reveals underlying similarities in cosmology, ritual, and belief (Rossano 2007). These often include gods and spirits with limited powers who are typically *not* omniscient and usually lack concern for morality and human affairs (Marlowe 2010; Norenzayan et al. 2016; Peoples and Marlowe 2012; Swanson 1960; Woodburn 1997), as Marshall describes:

> The concept of sin as an offence against the gods is vague among the !Kung. Man's wrong-doing against man is not left to #Gao!na's punishment nor is it considered to be his concern. Man corrects or avenges such wrong-doings himself in his social context. (Marshall 1962:245)

What were the specific traits of early religion? How did traits of nascent religiosity evolve and interact over time? Recently phylogenetic comparative methods have been increasingly applied to the study of the evolution of material and non-material culture (Mace and Holden 2005; Mace and Jordan 2011), including religion (Matthews 2012; Watts et al. 2015). Reconstructing ancestral character states on phylogenies based on genetic or linguistic data has proven valuable in revealing the history of various sociocultural phenomena (Currie et al. 2010; Opie et al. 2014; Walker et al. 2011, 2012). Although religious beliefs are regarded as one of those cultural traits that are historically labile and prone to cultural borrowing (Guglielmino et al. 1995), cross-cultural research suggests that religion (and mythology) can be surprisingly stable across time and space, and shared religious beliefs can be indicative of deep ancestry (Berezkin 2008; Blust 2013). The use of phylogenetic methods is important for understanding not only the origins of religious traits, but also the behavioral systems that emerged from them that have determined patterns of social constraint and have impacted believers and non-believers alike.

Importantly, homology is not limited to morphology and its genetic and/or developmental underpinnings. Behavior, which is often evolutionarily conserved, is also a proper subject of homology relations and can be used in phylogenetic reconstruction (Hall 2013; Powell and Shea 2014; Rendall and Di Fiore 2007). A behavioral homology need not have a particular structural basis. We do not argue for homology of the particular religious beliefs (e.g., different afterlife beliefs across hunter-gatherer societies that possess this trait) but for the homology of the fundamental religious concepts (e.g., the concept of afterlife itself) and their continuity. Even characters that cannot be hypothesized as strictly homologous among sampled cultures can be analyzed because they can represent non-homologous psychological-behavioral responses to identical selective pressures (see Murdock 1965 for similar reasoning in anthropology).

In this study we investigate early evolution of religion by reconstructing ancestral states for seven characters describing religious beliefs and behaviors in a global sample of 33 hunter-gatherer societies (Fig. 1). Using a time-calibrated supertree based on published genetic and linguistic phylogenetic trees, and linguistic classification as a proxy for population history, we reconstruct ancestral character states and test for correlated evolution between the characters and for the direction of cultural change.

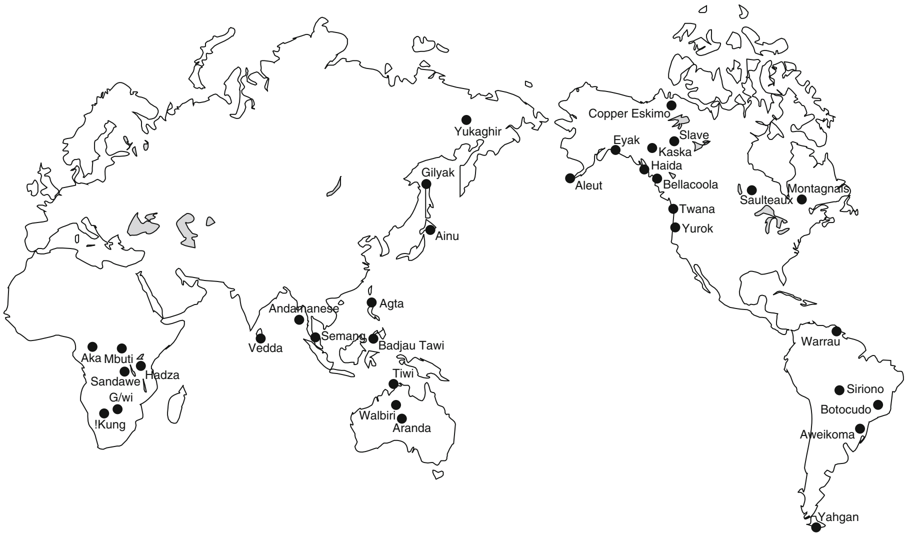


Fig. 1 Geographic locations of the 33 hunter-gatherer societies in the study sample

Methods

Phylogenetic reconstruction of ancestral character states is a two-step process. It requires a set of characters (data matrix) with known character states (e.g., present or absent) based on traits of interest, and a phylogenetic tree that represents the historical relationships between the populations under investigation.

Population Sample

We used primary ethnographic sources to create a data matrix of characters describing religiosity in a sample of 33 hunter-gatherer societies. The study sample consists of 28 hunter-gatherer societies selected from the Standard Cross-Cultural Sample (SCCS) (Murdock and White 1980), and five additional hunter-gatherer societies selected from both the *Ethnographic Atlas* (EA) (Murdock 1967) and original ethnographic descriptions of the societies.

We used ratings of other researchers to obtain a subsample of 28 hunter-gatherer societies from the SCCS. The variables used to define the sample were SCCS variable 1: Intercommunity Trade as Food Source; variable 3: Agriculture Contribution to Local Food Supply; variable 5: Animal Husbandry Contribution to Food Supply (Murdock and Morrow 1970); and variable 858: Subsistence Type–Ecological Classification (coded by D. White after Paige and Paige 1981). We defined our sample of hunter-gatherers as follows: contribution to local food supply less than 10% agriculture (SCCS v3), less than 10% animal husbandry (SCCS v5), and trade accounting for less than 50% and no more than any single local source (SCCS v1). We excluded mounted hunters (SCCS v858 = 5, mounted hunting) to more accurately represent pre-agricultural hunter-gatherer societies. We excluded eight SCCS hunter-gatherer societies from the sample either because of the lack of phylogenetic information (Ingalik, Micmac, E. Pomo, Yokuts [Lake], Paiute North, Klamath, and Kutenai) or because their main source of subsistence stated in the EA does not classify them as hunter-gatherers (Shavante depend 16–25% on agriculture). We added five additional societies chosen from the EA and original ethnographic sources in order to balance the geographic distribution of the sample. These additional societies are the Aka and G/wi (Africa), Agta (Philippines) and Walbiri (Australia) hunter-gatherers, and the Sandawe (Africa), who are now mainly herders and farmers but share a deep genetic and historical relationship with the Hadza (Tishkoff et al. 2007). Figure 1 shows the geographic locations of the 33 hunter-gatherer societies in the study.

Trait Definition and Character Matrix Construction

Original coding of data in all 33 sample societies for the traits of animism (Tylor 1871), belief in an afterlife (Bering 2006), shamanism (Eliade 1964; Winkelman 1990), and ancestor worship (Sheils 1975; Spencer 1870; Steadman et al. 1996; Swanson1960) was based on principal ethnographic source descriptions (White 1989) and additional ethnologies referenced in the character matrix.

Tylor defined animism as a general belief in the “animation of all nature” (1871:258) and fundamental to religion. Animism includes a “belief in personal souls” (1871:260) as well as “a sense of spiritual beings. . . inhabiting trees and rocks and waterfalls” (1871:260). We define animism as the belief that all “natural” things, such as plants, animals, and even such phenomena as thunder, have intentionality (or a vital force) and can have influence on human lives. Animism is coded as present or absent in each society, based on assessments of principal ethnographers. Belief in an afterlife is defined as belief in survival of the individual personality beyond death (Bering 2006) and is coded as either present or absent.

A global definition of shamanism remains contentious (Sidky 2010). We define shamanism as the presence in a society of a “shaman” (male or female), a socially recognized part-time ritual intercessor, healer, and problem solver (Sidky 2010; Winkelman 1990). Shamans often use their power over spirit helpers during performances involving altered states of consciousness (Eliade 1964; Winkelman 2010) to benefit individuals and the group as a whole (Eliade 1964; Winkelman 1990, 2010). We view shamans as a general category of individuals often found in hunter-gatherer societies who mediate between the earthly and spirit worlds to promote cohesion and physical and mental well-being in the society (Eliade 1964; Sidky 2010; Winkelman 1990, 2010). Shamanism is coded as present or absent.

Ancestor worship is defined as belief that the spirits of dead kin remain active in another realm where they may influence the living, and can be influenced by the living (Sheils 1975; Spencer 1870; Steadman et al. 1996; Swanson 1960). Ancestor worship is coded as four states in the SCCS: a belief in ancestor spirits can be (1) absent in a society or the spirits can be (2) present but inactive in human affairs. In other societies, ancestor spirits are believed to be active in human affairs but (3) may not or (4) may be influenced by humans through prayer and sacrifice (Sheils 1975; Spencer 1870; Steadman et al. 1996; Swanson 1960). “High gods” is EA variable 34 and SCCS variable 238, defined by Swanson (1960) as single, all-powerful creator deities who may be active in human affairs and supportive of human morality. The variable is coded as four states. It differentiates between societies in which a creator deity is (1) absent, (2) present but inactive in human affairs, (3) active in human affairs but does not support a moral agenda, or (4) active and morally punishing. The SCCS provides coding for high gods in 28 of the 33 societies in our sample. Original coding in the additional five societies, based on principal ethnographic sources, completed the coding for all 33 societies.

Based on the five traits of hunter-gatherer religion described above, we created a set of seven characters of hunter-gatherer religiosity. The character matrix was coded as follows: animism (absent, present); belief in an afterlife (absent, present); shamanism (absent, present); ancestor worship (absent, present); high gods (absent, present). Two additional traits, active ancestor worship (absent, present) and active high gods (absent, present), were derived from the basic traits of ancestor worship and high gods. These two additional characters recognize those societies that not only hold a belief in ancestor spirits or high gods but also believe that those spirits or high gods are active in human affairs. "Active ancestor worship" denotes the absence or presence of ancestral spirits actively meddling in human affairs, who may or may not be influenced through prayer or sacrifice (Sheils 1975; Spencer 1870; Steadman et al. 1996; Swanson 1960). Similarly, "active high gods" identifies societies that believe in a high god that is also *active* in human affairs, and may or may not be morally punishing (Swanson 1960).

The resulting character matrix has 33 terminals (hunter-gatherer populations) and 7 characters (Fig. 2). (Character matrix and coding references are provided in the ESM, Tables A1a and b.)

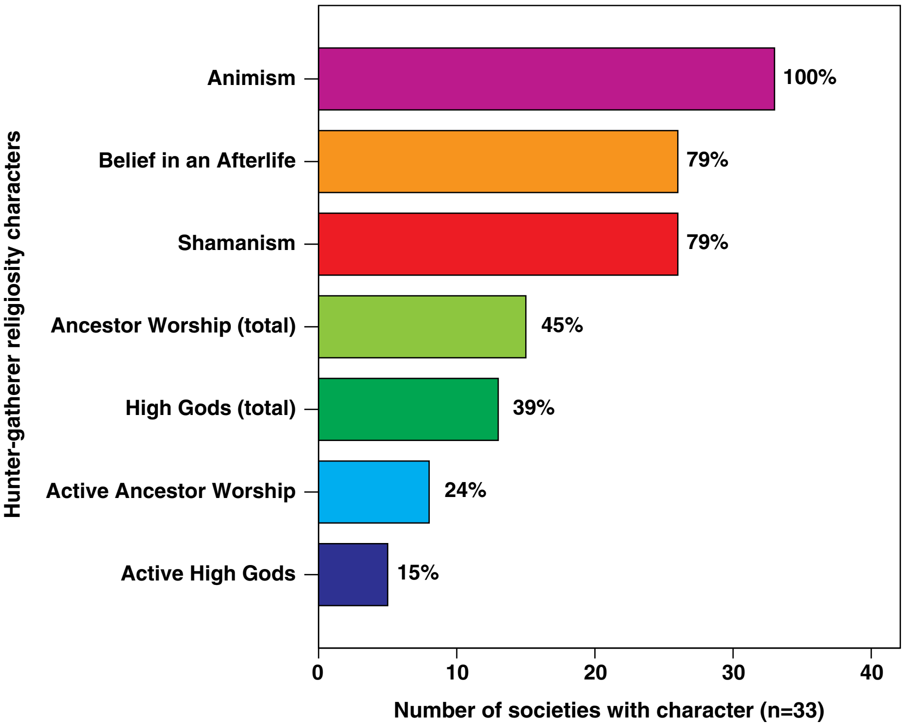


Fig. 2 Distribution of the seven characters describing hunter-gatherer religiosity in the study sample

Phylogenetic Supertree Inference

We used a dataset described in Duda and Zrzavý (2016) to generate a phylogeny of 33 hunter-gatherer populations using the Matrix Representation with Parsimony (hereinafter MRP) supertree method (Baum 1992; Ragan 1992). The dataset is based on 301 genetic and linguistic source trees from 199 studies published in journals, edited volumes, and books from 1990 to ca. 2014. It also includes a character set based on language classifications of Greenberg (1987), Ruhlen (1991), and Greenberg and Ruhlen (2007) on the level of linguistic stocks and families.

Data from White (2009) and *Ethnologue* (Lewis et al. 2013) were used to match the hunter-gatherer populations in the study sample with the populations present in the supertree dataset.

Populations absent from or underrepresented in the supertree dataset were replaced by either a more inclusive population group or by a genetically closely related population that was used as a proxy for the hunter-gatherer population in question. Positions of four North American populations (Kaska, Eyak, Twana, and Yurok) were based solely on linguistic classification (ESM Table A2).

In order to overcome the problem of the lack of genetic data and the conflicting signals caused by recent genetic admixture and language shifts in some hunter-gatherer populations in the study sample, the characters based on linguistic classifications were up-weighted by a factor of 100 to serve as a topological constraint or “scaffold.” This scaffold tree constrains the topology for a subset of populations for which linguistic affiliation can be determined (i.e., those scored for characters). (Details on supertree dataset manipulations are given in the Methods section of the ESM.)

The linguistic scaffold tree included 20 phylogenetically informative characters for the 33 populations in the study sample. Note that this linguistic scaffold implies relatively few internal groupings (clades based on linguistic classification), particularly among the Old World hunter-gatherers (see ESM Figure A1).

Semi-rooted coding (sensu Bininda-Emonds et al. 2005) was employed. The supertree was rooted by an “all-0” hypothetical outgroup that preserves the rooting information for rooted source trees; for unrooted source trees the hypothetical outgroup was scored as “?”.

The complete dataset included 974 taxa + outgroup. The analysis was performed in TNT 1.1 (Goloboff et al. 2008) under “new technology search” with search level 10 using sectorial, ratchet, and tree fusing searches, obtaining trees from a 10,000-replicate random addition sequence, treating gaps as missing data and all character changes as equal and non-additive. The recovered trees were then subjected to additional branch swapping with up to 10,000 trees held during each step. The resulting supertree is a semi-strict consensus tree reduced to 33 hunter-gatherer populations + outgroup. The topology of this supertree is fully resolved (ESM Figure A2).

Time-Calibrating the Supertree

Time-calibrated branch lengths were obtained from published time estimates of divergence events and colonization events in human population history based on molecular, linguistic, and archaeological data (ESM Table A3b). These estimates were used as time constraints on the nodal ages of the supertree. In order to test the robustness of reconstructions of ancestral character states and given the considerable variance in molecular-based time estimates of divergence dates and discrepancies between estimates based on molecular, linguistic, and archaeological data, two sets of time estimates were used. (Two sets of divergence dates for time-calibrating the supertree are given in ESM Table A3a.)

The first set of dates assumes a greater time depth of the supertree and consists mostly of molecular-based estimates of divergence dates. The second set of divergence dates is based on molecular-based estimates, archaeological data, and glottochronology and is close to the minimum time estimates. We emphasize the results based on the second set of dates since these represent the time when the populations in our sample last shared close cultural contact, which arguably suits our analyses better than the estimates of deeper, molecular-based divergence dates between the populations in question.

The age of the nodes for which time estimates were available was fixed using the Node Age Constraint tool in Mesquite 3.02 (Maddison and Maddison 2015). The time-calibrated supertree was inferred using the combination of Enforce Minimum Node Age Constraints and Arbitrarily Ultrametricize functions in Mesquite 3.02 (Maddison and Maddison 2015).

Phylogenetic Reconstruction of Ancestral Character States

The set of characters was mapped onto the tree topology. All character states in the outgroup were scored as 0, absent (i.e., plesiomorphic character states for all characters). Maximum parsimony and maximum likelihood reconstruction of ancestral character states were performed in Mesquite 3.02 (Maddison and Maddison 2015). The Markov k-state 1 parameter model (Mk1) that assumes an equal rate of change between all character states (Lewis 2001) was used for maximum likelihood reconstruction. An asymmetric likelihood ratio test (Pagel 1999b) indicated that the asymmetric two-parameter model does not offer a significant improvement over the Mk1 model for any of the seven characters in question, thus validating the Mk1 model. Each character was mapped onto a set of topologies using the Trace Character History function.

The statistical support for the ancestral state reconstructions was determined using a likelihood decision threshold of $T = 2$, indicating support at least 7.4 times greater for the character state in question than for the alternative character state(s) (Schluter et al. 1997).

Testing for Correlated Character Evolution

In order to test hypotheses about temporal ordering of character state changes and coevolution of traits we used Pagel's test for correlated discrete character evolution (Pagel 1994, 1999a) implemented in the Pagel94 module in Mesquite 3.02 (Maddison and Maddison 2015). This method uses a continuous-time Markov model to infer character changes along each branch of a phylogenetic tree in order to establish the most likely temporal ordering and direction of evolutionary change and the most probable evolutionary pathway between two discrete binary characters. Evolutionary change in each character along the tree branches is modelled as a Markov process, in which the likelihood of character change is dependent on its current character state.

Two models are fitted: an independent, four-parameter model (L_i) in which evolution in each character is independent of the state of the other character, and a dependent, eight-parameter model (L_d) in which the probability of change in one trait is dependent on the state of the other trait. (For example, the probability of a culture gaining shamanism can differ between cultures with a belief in an afterlife and cultures without one.) A likelihood ratio (LR) is used to compare the log likelihoods of the independent and dependent models. The advantage of this method is that its use is not conditioned on the ability to unequivocally reconstruct ancestral character states (Nunn 2011).

The supertree with the preferred set of divergence dates (i.e., shallower divergences) was used for correlation analysis. The probability that a model of dependent evolution fits the data significantly better than the model of independent evolution was estimated with a likelihood-ratio test involving 1000 Monte Carlo simulations. A likelihood-ratio test generates a null distribution of likelihood ratios, against which the significance of the observed LR is tested. For each simulation, maximum-likelihood estimates of model parameters were optimized using 500 iterations. If the dependent model fits the data significantly better than the independent model, this indicates that the state of one character affects the probability of change in the other, and that the two characters probably coevolve.

Results

The resulting supertree topology (see ESM Figure A2) indicates a deep split between sub-Saharan African and non-African hunter-gatherers. Within Africa, South African Khoisan who speak !Kung and G/wi are more closely related to Central African Pygmies than to click-speaking East African Hadza and Sandawe. Outside Africa, the Vedda of Sri Lanka is the deepest-rooting lineage, followed by Andaman Islanders and Australian Aboriginals, presumably the remnants of early out-of-Africa expansion via the “southern route” (Macaulay et al. 2005). A large clade follows, consisting of two groups: hunter-gatherers of East Asia and those in Beringia and America. Within East Asia, two sister groups appear: Southeastern (Negritos of Malaysia and Philippines and Badjau Tawi) and Northeastern (the “Paleo-Asiatic” peoples). The Beringian-American clade consists of related Eskimo-Aleut and Na-Dene speakers, and the Amerind speakers of North and South America as a sister group of the two.

The reconstructions of ancestral states for religious beliefs and behaviors show several consistent patterns using maximum parsimony and maximum likelihood methods, and topologies with two alternative sets of divergence dates. The presence of animistic concepts in the religions of all sample societies (Fig. 2) is in accord with Tylor’s (1871) theory that animism is fundamental to religion. The presence of animism in the last common ancestor (LCA) of present-day hunter-gatherers is significantly supported (proportional likelihood = 0.99, $p < 0.05$*).[1] Reconstructions of ancestral states for the six remaining characters describing religion in our hunter-gatherer sample are shown in Fig. 3. (Reconstructions of ancestral states for each node based on maximum parsimony and maximum likelihood using two sets of divergence dates are given in ESM Table A4.)

---

[1] Asterisk (*) indicates significant result $p \leq 0.05$

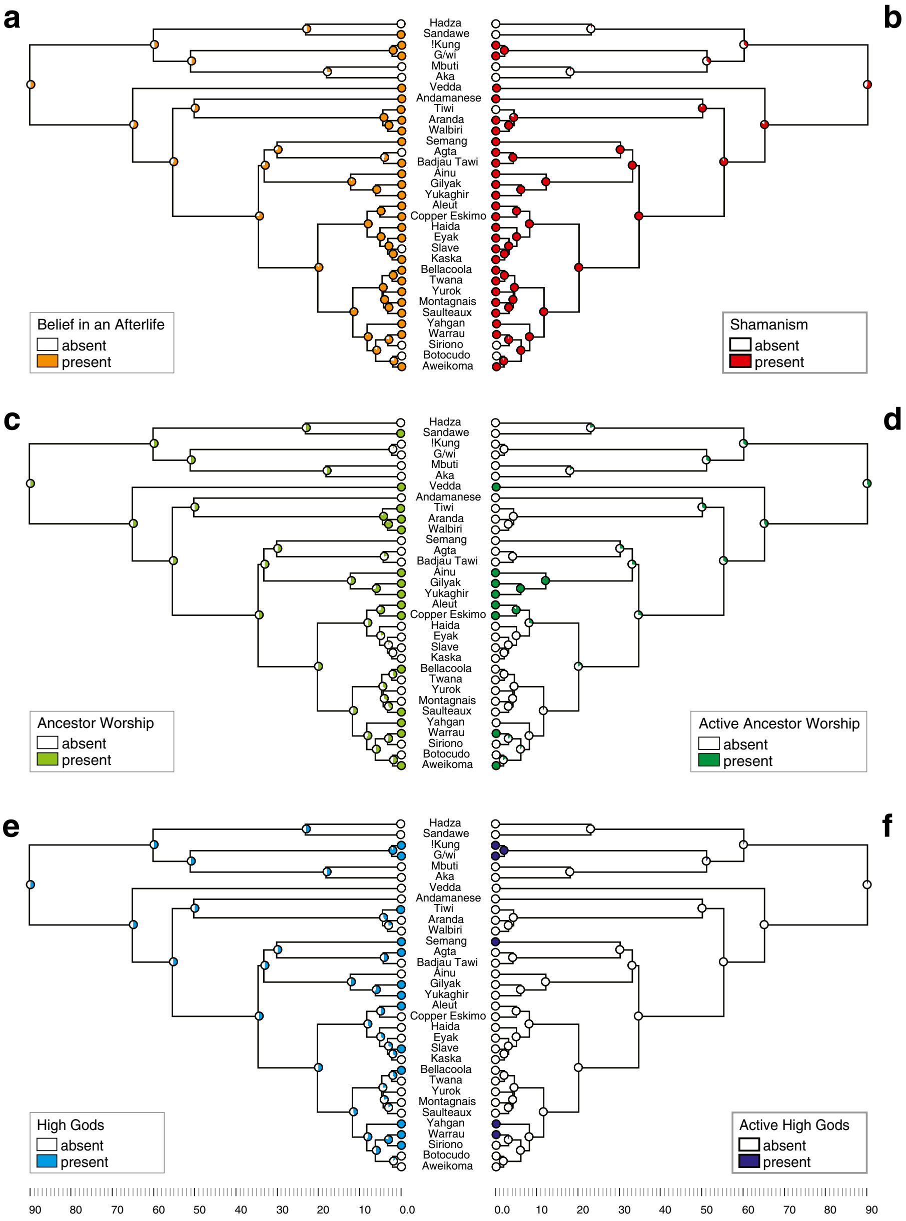


Fig. 3 Maximum likelihood reconstructions of ancestral states for six characters describing hunter-gatherer religiosity. **a** Belief in an Afterlife. **b** Shamanism. **c** Ancestor Worship. **d** Active Ancestor Worship. **e** High Gods. **f** Active High Gods. The scale indicates time depth in kya. (see ESM Table A4 for details)

Belief in an afterlife and shamanism are present among 79% of sample societies (Fig. 2) and have similar, although not identical, distribution across societies (Fig. 3a, b). These characters are less common among African hunter-gatherers. The reconstructed ancestral state in the deepest node for belief in an afterlife and shamanism is equivocal (proportional likelihood = 0.5 and 0.56, respectively; ESM Table A4a) according to maximum likelihood. Maximum parsimony favors the absence of shamanism. We cannot determine whether belief in an afterlife and shamanic practices were present in the LCA of present-day hunter-gatherers. Among present-day African hunter-gatherers (the deepest-rooting clades) the !Kung and G/wi hold a belief in an afterlife, whereas Hadza, Mbuti, and Aka either never acquired it or had the trait and then lost it (Fig. 3a). The !Kung and G/wi have shamans, but Hadza, Sandawe, Aka, and Mbuti do not (Fig. 3b). The presence of healers among the Mbuti suggests that the Mbuti may once have had shamans, but they lost the trait at some point (Winkelman 1990).

Ancestor worship is present in 45% and active ancestor worship in 24% of sample societies (Fig. 2). The ancestral state of "ancestor worship" is equivocal (proportional likelihood = 0.5; ESM Table A4b). Ancestral presence of active ancestor worship is somewhat less likely. However, this is not significantly supported (proportional likelihood = 0.43429948; ESM Table A4b) on the topology with shallower divergence dates. In contrast, its ancestral absence is significantly supported (0.10739427*; ESM Table A4b) on the topology with deeper divergence dates. Maximum parsimony reconstruction favors the absence of active ancestor worship (ESM Table A4b). These results suggest ancestor worship could have been present among ancestral hunter-gatherers, but probably not the active form.

Only 39% of sample societies have the trait "high gods," and even fewer (15%) have active high gods (Fig. 2). The equivocal results based on maximum likelihood and absence according to maximum parsimony for the ancestral state of "high gods" suggests possible presence of belief in a single creator deity among ancestral hunter-gatherers, albeit one that is not active in human affairs (proportional likelihood = 0.05*) (see ESM Table A4b for details).

The consistency and retention indices (CI, RI) calculated for each character and the whole character matrix quantifies the degree of character "fit" on the tree. CI (with values from 0 to 1) measures the amount of homoplasy on a tree; RI (also 0–1) measures the degree to which shared derived character states are exhibited on a tree. The resulting CI and RI values for the whole character matrix are low (0.17 and 0.31, respectively), indicating that most characters are highly labile. The characters displaying the worst fit on phylogeny are ancestor worship and high gods (CI = 0.1), and the reconstructed ancestral states for both are equivocal (Fig. 3c, e; ESM Table A4b).

The results of Pagel's test for correlated evolution indicate a significant positive relationship between most traits investigated. The dependent models for the evolution of selected pairs of characters showing support for correlated evolution are shown in Fig. 4. (The results of Pagel's test for correlated evolution, including transition rates for independent and dependent model of character change, log likelihood values for each model, and *p* values, are given in ESM Table A5.)

Belief in an afterlife and shamanism emerge in the presence of the fundamental trait of animism. Once these two traits are gained, they are unlikely to be lost (Fig. 4a).

Coevolution of belief in an afterlife and shamanism is significantly supported ($L_i = -35.71$, $L_d = -28.75$, $LR = 6.96$, $p = 0.00^*$) (Fig. 4c). The transitional probabilities indicate that belief in an afterlife evolves more likely in the absence of shamanism than shamanism would evolve in the absence of a belief in an afterlife. This suggests that belief in an afterlife is likely to have emerged first from the base of animistic beliefs, and later shamanism evolved in the presence of belief in an afterlife (Fig. 4c). It also indicates that shamanism is likely to be lost in the absence of belief in an afterlife (Fig. 4c; ESM Table A5).

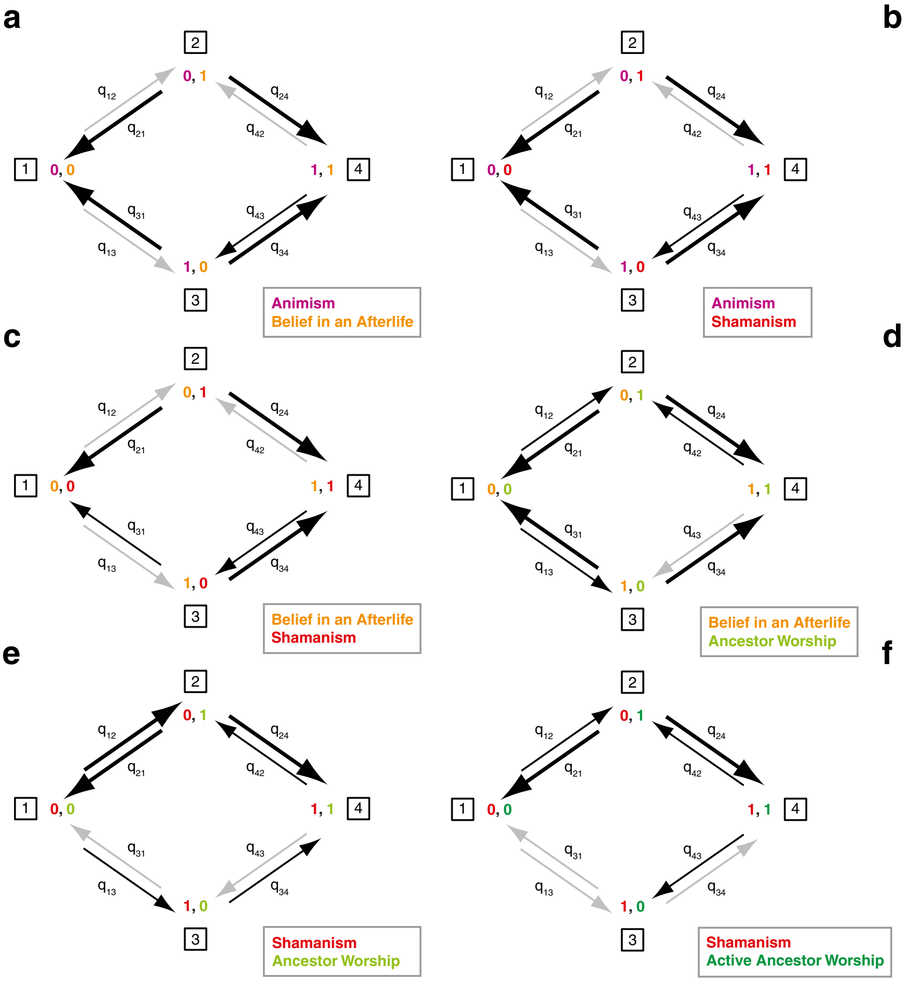


Fig. 4 Transitions between character states for selected pairs of characters showing significantly higher likelihood of the dependent model of evolution, indicating that these traits coevolve. Widths of arrows are proportional to rates of change (see ESM Table A5 for details)

There is significant support for coevolution of belief in an afterlife and ancestor worship (L_i = −40.48, L_d = 34.66, LR = 5.82, p = 0.003*) (Fig. 4d). Belief in an afterlife evolves prior to ancestor worship, and its presence stimulates the subsequent evolution of ancestor worship. Ancestor worship is unlikely to be lost in the presence of belief in an afterlife (Fig. 4d).

There is also significant support for coevolution of both shamanism and ancestor worship (L_i = −40.48, L_d = −36.60, LR = 3.88, p = 0.01*) and shamanism and active ancestor worship (L_i = −34.74, L_d = −28.72, LR = 6.02, 0.001*). Shamanism seems to have a deep history and continuity, whereas ancestor worship, although it could have evolved very early in the history of modern humans, is a highly labile trait (Fig. 3c). Active ancestor worship probably appeared later (Fig. 3d). Ancestor worship without shamanism seems to be an unstable cultural state that results either in a loss of worshipful relationship with dead kin or in the appearance of the shaman. Ancestor worship with shamanism, on the other hand, appears to be a stable cultural state, rarely lost once achieved, and the same is seen for active ancestor worship with shamanism (Fig. 4e, f).

There is no support for coevolution of any pair of characters that includes high gods and active high gods, with the obvious exception of the two traits themselves ($L_i = -33.58$, $L_d = -27.93$, LR = 5.65, $p = 0.001$). Surprisingly, not even belief in an afterlife shows any correlation with high gods. Belief in an afterlife evolved prior to high gods, as is evident from reconstruction of ancestral states and the transition rates based on Pagel's test for correlated evolution. But these pairs of characters do not coevolve: in other words, the probability of change in one is not affected by the state of the other (see ESM Table A5 for details).

Discussion

Our results reflect Tylor's (1871) belief that animism was the earliest and most basic trait of religion because it enables humans to think in terms of supernatural beings or spirits. Animism is not a religion or philosophy, but a feature of human mentality, a by-product of cognitive processes that enable social intelligence, among other capabilities. It is a widespread way of thinking among hunter-gatherers (Bird-David 1999; Charlton 2007; Klingensmith 1953; Piaget 1929). Animistic thought is a natural by-product of the human capacity for intentionality or "theory of mind mechanism" (Dunbar 2003). This innate cognitive trait allows us to attribute a vital force to animate and inanimate elements in the environment (Piaget 1929; Tylor 1871). Once that vital force is assumed, attribution of other human characteristics will follow. Animistic beliefs are generally adaptive in the environments that prevail in hunter-gatherer societies (Bird-David 1999; Charlton 2002). Animistic thinking would have been present in early hominins, certainly earlier than language (Coward 2015; Dunbar 2003).

It can be inferred from the analyses, or indeed from the universality of animism, that the presence of animistic belief predates the emergence of belief in an afterlife. Once animistic thought is prevalent in a society, interest in the whereabouts of spirits of the dead could reasonably lead to the concept of an unseen realm where the individual personality of the deceased lives on. The afterlife might be a rewarding continuation of life on earth, or a realm of eternal punishment for those who break social norms. Belief in an afterlife may have generated a sense of "being watched" by the spirits of the dead, prompting archaic forms of social norms (Bering 2006) actualized in the role of the shaman.

Shamanism significantly correlates with belief in an afterlife, which emerged first. Shamanism then evolved in the presence of belief in a realm of spirits of the dead. If belief in an afterlife is lost, shamanism is also likely to be lost. The single exception to this in our sample is the Slave, who have shamanism without belief in an afterlife. Although shamanism has been described as the universal religion of Paleolithic hunter-gatherers (Eliade 1964; Winkelman 1990), it is not a religion per se, but a complex of beliefs and behaviors that focus on communication with the ancestral spirits, as well as the general world of spirits in the realm of the afterlife. Shamans are healers, ritual leaders, and influential members of society whose keen insight and success in solving social problems (Rossano 2007; Winkelman 1990) can lead to wealth, power, and access to mates. Shamanism acts as a mechanism to reinforce social norms, encouraging group cooperation through ritual and social bonding, and calming anxiety during times of resource stress (Hayden 1987; Rossano 2007; Winkelman 1990; Winkelman 2010). Shamans, as Vitebsky (2000) puts it, are both spiritual leaders and social workers. It would be reasonable to argue that shamans, who draw their power from communication with the world of spirits, would have initially emerged in strongly animistic societies that believed in an afterlife. Communication with omniscient and perhaps judgmental spirits of known deceased, including ancestors, would have been a useful tool in the work of the shaman.

As humans migrated out of Africa more than 60 kya (Henn et al. 2012; Macaulay et al. 2005), the shaman's curing skills and group rituals would have enhanced survival through physical and emotional healing, enforcement of group norms, and resource management. At the time of the rapid population dispersal of AMH out of Africa along the southern route into Wallacea and Sahul, the physical stress of travel and encounters with unfamiliar cultures in areas already occupied by other hominins would have driven the need for use of both material and non-material culture, including religion, to negotiate identities and relationships among and between groups (Coward 2015). Evidence for more complex information exchange systems, planning depth and authority, and increased symbolization appears in the archaeological record of Wallacea and Australia prior to 40 kya (Balme et al. 2009). These suggestions are in line with the elevated likelihood support of the ancestral presence of shamanism in the deepest out-of-Africa nodes (Fig. 2b). Our results support Rossano's (2009a) hypothesis that shamanism preceded and is more basic than ancestor worship, although the presence of shamanism in the LCA of present-day hunter-gatherers is not supported.

Despite established speculation by Spencer (1870) and Tylor (1871) that universal ancestor worship was the rudimentary beginning of religion, our analysis shows that worship of dead kin is neither widespread among hunter-gatherers nor the oldest trait of religion. Fewer than half of the societies in our sample believe that dead kin can influence the living (Fig. 2). In many hunter-gatherer societies the concept of ancestor spirits is absent, or present but they are inactive in human affairs (Sheils 1975; Swanson 1960). For example, among the! Kung there is a general fear of active spirits of "the dead," who are often the ghosts of recently deceased kin. But the concept of having a worshipful relationship with their own ancestors is absent (Marshall 1962). Greater likelihood of the presence of active ancestor worship has been linked to societies with unilineal descent where important decisions are made by the kin group (Sheils 1975; Swanson 1960). Ancestor worship is an important source of social control that strengthens cohesion among kin and maintains lineal control of power and property (Sheils 1975; Steadman et al. 1996; Swanson 1960), particularly in the more complex hunter-gatherer societies. In contrast, immediate-return hunter-gatherer societies (Woodburn 1982) seldom recognize dead ancestors who may intervene in their lives. The social structure of these societies does not usually consist of strong kin ties, and individuals do not depend on help from close kin, living or dead (Barnard and Woodburn 1988).

The minimum requirement for veneration of dead ancestors is animism and belief in the survival of the personal identity beyond death. In our analyses, ancestor worship is significantly positively related with belief in an afterlife and shamanism. Belief in an afterlife evolves prior to shamanism and ancestor worship. There is significant support for coevolution of shamanism with ancestor worship and active ancestor worship. Belief in an afterlife with shamanism appears to be a stable cultural state, rarely lost once achieved. Ancestor worship is also less likely to be lost in the presence of belief in an afterlife with shamanism.

This is not to say that the reduction of complexity of religious beliefs and behaviors cannot occur in simple hunter-gatherers. The presence of belief in an afterlife and shamanism is significantly supported in the LCA of Beringian-American as well as North and South American hunter-gatherers (Fig. 3a,b; ESM Table A4), suggesting that the absence of these traits among the Siriono in Bolivia and Botocudo in southern Brazil is due to independent secondary losses. In the Siriono, this loss was probably part of a substantial decrease in cultural complexity during the expansion of Tupi language speakers across lowland South America (Walker et al. 2012).

It can be argued that those societies under higher resource stress, encountering difficulties with resource extraction that demands cooperative effort, would benefit most from the shaman's skills (Hayden 1987; Rossano 2007; Winkelman 1990, 2010). This is supported by the prevalence of shamanism among hunter-gatherer societies of Eurasia, and corresponding support for the presence of shamanism among their ancestors (Fig. 2b). In "Paleo-Asiatic" peoples (Ainu, Gilyak, and Yukaghir) and in Eskimo-Aleut peoples of the circumpolar region, the presence of shamanism combines with active ancestor worship. The presence of both traits in the LCAs of these groups is significantly supported (Fig. 2d). High gods were not the first supernatural entities to monitor morality (Geertz 2014). The power and leadership of the shaman was often based on reaffirming traditional social behavior that was presumed to have been carried out by the ancestors, and still desired and monitored by punishing ancestral spirits, even in those societies where spirits of dead kin were not considered a part of the religion (Steadman and Palmer 2008).

Belief in high gods appears to be a rather "stand-alone" phenomenon in the evolution of hunter-gatherer religion. Prior studies have shown that among the four modes of subsistence (hunter-gatherers, pastoralists, horticulturalists, and agriculturalists) hunter-gatherers are least likely to adopt morally punishing active high gods, if any high gods at all (Botero et al. 2014; Norenzayan 2013; Peoples and Marlowe 2012; Swanson 1960). This pattern is reflected in the distribution and the reconstructed evolution of high gods and active high gods in our sample (Fig. 2; Fig. 3e, f). Early egalitarian hunter-gatherers would rarely have acknowledged an active high god (Norenzayan 2013; Peoples and Marlowe 2012) and would be the least likely to accept or benefit from the supernatural meddling and social constraints of deities who would be seen as "high rulers" (Peoples and Marlowe 2012). The leaders of complex hunter-gatherer societies whose subsistence relies on collective effort should be more likely to benefit from the coercive power of a punishing high god. Our analysis does not support the prevalence of either type of high god among ancestral hunter-gatherers, and the evolution of high gods does not correlate with any of the other traits of hunter-gatherer religion, including ancestor worship.

In a study by Guglielmino et al. (1995) analyzing the transmission pattern of cultural traits in sub-Saharan Africa, high gods (Swanson 1960) was among a group of traits (taboos, ritual mutilation, premarital norms, etc.) that were consistently the least correlated with either language or ecology, suggesting they evolve rapidly and are prone to cultural borrowing. On the other hand, according to a more recent study by

Currie and Mace (2014), high gods are among those cultural traits that evolve at relatively slow rates in Bantu and Austronesian societies. Our results are consistent with Guglielmino et al. (1995) since high gods (and ancestor worship) show the worst fit on phylogeny (CI = 0.1) of all studied characters. This suggests that the presence of high gods and some other traits related to religion and ritual are influenced by more socioculturally oriented factors, and it lends support to the idea that these types of traits may be more labile. Such traits would be more readily gained or lost as the adaptively relevant sociopolitical environment changes (Irons 1998). To some extent this finding may explain the independent pattern of emergence of high gods in our study.

Ancestral spirits and local gods with limited powers of supernatural monitoring may have come relatively easily to the minds of early human hunter-gatherers. These types of supernatural entities operate in a different realm from omniscient and powerful creator gods (high gods), who have been shown to be related to a culture of some type of control or decision-making structure (Peoples and Marlowe 2012; Radin 1937; Swanson 1960). The absence of belief in active gods and spirits in the LCA of present-day hunter-gatherers, according to the reconstructions, indicates a deep evolutionary past for the egalitarian ethos of most simple hunter-gatherer societies, whose small mobile populations of self-sufficient individuals make collective action problems less of an issue. Those societies would be the least likely to accept or benefit from the personal restraints of active ancestors or active high gods.

Conclusion

In this study we used a suite of phylogenetic comparative methods to investigate the early evolution of religion. We reconstructed ancestral states for seven characters describing religious beliefs and behaviors in a global sample of 33 hunter-gatherer societies and tested for correlated evolution between these characters and for the direction of cultural change.

Our results indicate that the oldest trait of religion, shared by the most recent common ancestor of present-day hunter-gatherers, was animism. This supports longstanding beliefs about the antiquity and fundamental role of this component of human mentality, which enables people to attribute intent and lifelike qualities to inanimate objects and would have prompted belief in beings or forces in an unseen realm of spirits**.** Reconstructions are equivocal on whether or not the religion of the LCA of present-day hunter-gatherers included belief in an afterlife, shamanism, ancestor worship, and the concept of a single creator deity, or a high god. Belief in either ancestral spirits or creator deities who remain active in human affairs was not present in ancestral hunter-gatherer societies, according to the reconstructions. This may be indicative of a deep past for the egalitarian nature of hunter-gatherer societies, to whom high gods would appear to be rulers (Peoples and Marlowe 2012).

The majority of traits of religion we investigated exhibit a correlated pattern of character change on phylogeny. The results suggest that belief in an afterlife, shamanism, and ancestor worship evolve in concerted fashion as an integrated system of beliefs and practices. However, neither high gods nor active high gods exhibit correlated evolution with the rest of the religious traits, including ancestor worship, despite Spencer's and Tylor's suggestions.

This is in line with a variety of evidence from other studies (Botero et al. 2014; Norenzayan 2013; Peoples and Marlowe 2012; Radin 1937; Swanson 1960) suggesting that if a society acquires belief in an omniscient and potentially morally punishing creator deity, it does so regardless of other aspects of its religion but more as a reflection of its social and political structure.

Acknowledgments We thank Toomas Kivisild and Nathaniel Dominy for critical discussion; Olaf Bininda-Emonds and Jan Zrzavý for discussion of the methods; and Pavel Flegontov for critical comments on an earlier version of the manuscript. Pavel Duda was supported by the Grant Agency of the University of South Bohemia (042/2013/P; 140/2013/P).

References

Alcorta, C. S., & Sosis, R. (2005). Ritual, emotion, and sacred symbols: the evolution of religion as an adaptive complex. *Human Nature, 16*(4), 323–359.

Atran, S., & Henrich, J. (2010). The evolution of religion: how cognitive by-products, adaptive learning heuristics, ritual displays, and group competition generate deep commitments to prosocial religions. *Biological Theory, 5*, 18–30.

Atran, S., & Norenzayan, A. (2004). Religion's evolutionary landscape: counterintuition, commitment, compassion, communion. *Behavioral and Brain Sciences, 27*, 713–770.

Balme, J., Davidson, I., McDonald, J., Stern, N., & Veth, P. (2009). Symbolic behaviour and the peopling of the southern arc route to Australia. *Quaternary International, 202*(1), 59–68.

Barham, L. S. (2002). Systematic pigment use in the Middle Pleistocene of south-central Africa. *Current Anthropology, 43*(1), 181–190.

Barnard, A. (2012). *Genesis of symbolic thought*. Cambridge, UK: Cambridge University Press.

Barnard, A., & Woodburn, J. (1988). Property, power and ideology in hunter-gathering societies: an introduction. In T. Ingold (Ed.), *Hunters and gatherers: Property, power and ideology* (Vol. 2, pp. 4–31). Oxford: Berg.

Barney, A., Martelli, S., Serrurier, A., & Steele, J. (2012). Articulatory capacity of Neanderthals, a very recent and human-like fossil hominin. *Philosophical Transactions of the Royal Society B: Biological Sciences, 367*(1585), 88–102.

Barrett, J. L., & Lanman, J. A. (2008). The science of religious beliefs. *Religion, 38*(2), 109–124.

Baum, B. R. (1992). Combining trees as a way of combining data sets for phylogenetic inference, and the desirability of combining gene trees. *Taxon*, 3–10.

Berezkin, Y. (2008). Why are people mortal? World mythology and the "out-of-Africa scenario." In P. N. Peregrine & M. Feldman (Eds.), *Ancient human migrations: A multidisciplinary approach* (pp. 74–94). Salt Lake City: University of Utah Press.

Bering, J. M. (2006). The folk psychology of souls. *Behavioral and Brain Sciences, 29*(5), 453–498.

Bininda-Emonds, O. R., Beck, R. M., & Purvis, A. (2005). Getting to the roots of matrix representation. *Systematic Biology, 54*(4), 668–672.

Bird-David, N. (1999). "Animism" revisited: personhood, environment, and relational epistemology. *Current Anthropology, 40*(S1), S67–S91.

Blust, R. (2013). Terror from the sky: unconventional linguistic clues to the Negrito past. *Human Biology, 85*(3), 401–416.

Botero, C. A., Gardner, B., Kirby, K. R., Bulbulia, J., Gavin, M. C., & Gray, R. D. (2014). The ecology of religious beliefs. *Proceedings of the National Academy of Sciences*. doi:10.1073/pnas.1408701111.

Boyer, P., & Bergstrom, B. (2008). Evolutionary perspectives on religion. *Annual Review of Anthropology, 37*, 111–130.

Brown, D. E. (1991). *Human universals*. New York City: McGraw-Hill.

Charlton, B. (2002). What is the meaning of life? Animism, generalized anthropomorphism and social intelligence. Retrieved from <http://www.hedweb.com/bgcharlton/meaning-of-life.html>

Charlton, B. G. (2007). Alienation, recovered animism and altered states of consciousness. *Medical Hypotheses, 68*(4), 727–731.

Conard, N. J. (2010). Cultural modernity: consensus or conundrum? *Proceedings of the National Academy of Sciences of the United States of America, 107*(17), 7621–7622.

Coward, F. (2015). Scaling up: material culture as scaffold for the social brain. *Quaternary International*. doi: 10.1016/j.quaint.2015.09.064.

Cronk, L. (2005). The application of animal signaling theory to human phenomena: some thoughts and clarifications. *Social Science Information sur les Sciences Sociales, 44*(4), 603–620.

Currie, T. E., Greenhill, S. J., Gray, R. D., Hasegawa, T., & Mace, R. (2010). Rise and fall of political complexity in island South-East Asia and the Pacific. *Nature, 467*(7317), 801–804.

Currie, T. E., & Mace, R. (2014). Evolution of cultural traits occurs at similar relative rates in different world regions. *Proceedings of the Royal Society of London B: Biological Sciences, 281*(1795), 20141622.

Deacon, T., & Cashman, T. (2010). The role of symbolic capacity in the origins of religion. *Journal for the Study of Religion, Nature and Culture, 3*(4), 1–28.

Dediu, D., & Levinson, S. C. (2013). On the antiquity of language: the reinterpretation of Neandertal linguistic capacities and its consequences. *Frontiers in Psychology, 4*(397), 1–17.

Deino, A. L., & McBrearty, S. (2002). 40Ar/39Ar dating of the Kapthurin formation, Baringo, Kenya. *Journal of Human Evolution, 42*(1), 185–210.

d'Errico, F., & Stringer, C. B. (2011). Evolution, revolution or saltation scenario for the emergence of modern cultures? *Philosophical Transactions of the Royal Society B: Biological Sciences, 366*(1567), 1060–1069.

d'Errico, F., Henshilwood, C., Vanhaeren, M., & Van Niekerk, K. (2005). *Nassarius kraussianus* shell beads from Blombos Cave: evidence for symbolic behaviour in the Middle Stone Age. *Journal of Human Evolution, 48*(1), 3–24.

Duda, P., and Zrzavý, J. 2016. Human population history revealed by a supertree approach. In review.

Dunbar, R. I. (2003). The social brain: mind, language, and society in evolutionary perspective. *Annual Review of Anthropology*, 163–181.

Durkheim, E. (1965 [1912]). *The elementary forms of the religious life*. New York: Free Press.

Eliade, M. (1964). *Shamanism*: *Archaic techniques of ecstasy* (W. Trask, trans.). Princeton: Princeton University Press.

Fischer, R., Callander, R., Reddish, P., & Bulbulia, J. (2013). How do rituals affect cooperation? *Human Nature, 24*(2), 115–125.

Fu, Q., Mittnik, A., Johnson, P. L. F., Bos, K., Lari, M., Bollongino, R., et al. (2013). A revised timescale for human evolution based on ancient mitochondrial genomes. *Current Biology, 23*(7), 553–559.

Geertz, A. W. (2014). Do big gods cause anything? *Religion, 44*(4), 609–613.

Goloboff, P. A., Farris, J. S., & Nixon, K. C. (2008). TNT, a free program for phylogenetic analysis. *Cladistics, 24*(5), 774–786.

Goodall, J. (1986). *The chimpanzees of Gombe: Patterns of behavior*. Cambridge: Harvard University Press.

Greenberg, J. H. (1987). *Language in the Americas*. Stanford: Stanford University Press.

Greenberg, J. H., & Ruhlen, M. (2007). *An Amerind etymological dictionary*. Stanford: Stanford University Press.

Guglielmino, C. R., Viganotti, C., Hewlett, B., & Cavalii-Sforza, L. L. (1995). Cultural variation in Africa: role of mechanisms of transmission and adaptation. *Proceedings of the National Academy of Sciences of the United States of America, 92*(16), 7585–7589.

Hall, B. K. (2013). Homology, homoplasy, novelty, and behavior. *Developmental Psychobiology, 55*(1), 4–12.

Harris, E., & McNamara, P. (2008). Is religiousness a biocultural adaptation? In J. Bulbulia & R. Sosis et al. (Eds.), *The evolution of religion: Studies, theories & critiques* (pp. 79–85). Santa Margarita, CA: Collins Foundation Press.

Hayden, B. (1987). Alliances and ritual ecstasy: human responses to resource stress. *Journal for the Scientific Study of Religion, 26*(1), 81–91.

Henn, B. M., Cavalli-Sforza, L. L., & Feldman, M. W. (2012). The great human expansion. *Proceedings of the National Academy of Sciences, 109*(44), 17758–17764.

Henshilwood, C. S., d'Errico, F., & Watts, I. (2009). Engraved ochres from the Middle Stone Age levels at Blombos Cave, South Africa. *Journal of Human Evolution, 57*(1), 27–47.

Irons, W. (1998). Adaptively relevant environments versus the environment of evolutionary adaptedness. *Evolutionary Anthropology: Issues, News, and Reviews, 6*(6), 194–204.

Irons, W. (2001). Religion as a hard-to-fake sign of commitment. In R. Nesse (Ed.), *Evolution and the capacity for commitment* (pp. 292–309). New York: Russell Sage Foundation.

Johansson, S. (2011). Constraining the time when language evolved. *Linguistic and Philosophical Investigations, 10*, 45–59.

Johnson, D. D. P. (2005). God's punishment and public goods: a test of the supernatural punishment hypothesis in 186 world cultures. *Human Nature, 16*(4), 410–446.

Klingensmith, S. W. (1953). Child animism: what the child means by "alive." *Child Development, 24*(1), 51–61.

Krause, J., Lalueza-Fox, C., Orlando, L., Enard, W., Green, R. E., Burbano, H. A., et al. (2007). The derived FOXP2 variant of modern humans was shared with Neandertals. *Current Biology, 17*(21), 1908–1912.

Lee, D. (1989). Religious perspectives in anthropology. In A. C. Lehmann & J. E. Myers (Eds.), *Magic, witchcraft, and religion: An anthropological study of the supernatural* (2nd ed., pp. 16–23). Mountain View: Mayfield.

Lee, R. B., & DeVore, I. (1968). Problems in the study of hunters and gatherers. In R. B. Lee & I. DeVore (Eds.) *Man the hunter* (pp. 1–12). Chicago: Aldine.

Lewis, M., Simons, G., & Fennig, C. (2013). *Ethnologue*: *Languages of the world*. (17th ed.). Dallas: SIL International. Available online at http://www.ethnologue.com

Lewis, P. O. (2001). A likelihood approach to estimating phylogeny from discrete morphological character data. *Systematic Biology, 50*(6), 913–925.

Lind, J., Lindenfors, P., Ghirlanda, S., Liden, K., Enquist, M., & Patin, E. (2013). Dating human cultural capacity using phylogenetic principles. *Scientific Reports, 3*, 1–5.

Macaulay, V., Hill, C., Achilli, A., Rengo, C., Clarke, D., Meehan, W., et al. (2005). Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Science, 308*(5724), 1034–1036.

Mace, R., & Holden, C. J. (2005). A phylogenetic approach to cultural evolution. *Trends in Ecology & Evolution, 20*(3), 116–121.

Mace, R., & Jordan, F. M. (2011). Macro-evolutionary studies of cultural diversity: a review of empirical studies of cultural transmission and cultural adaptation. *Philosophical Transactions of the Royal Society B: Biological Sciences, 366*(1563), 402–411.

Maddison, W. P., & Maddison, D. R. (2015). *Mesquite*: *A modular system for evolutionary analysis*, version 3.02. Available online at mesquiteproject.org

Marean, C. W., Bar-Matthews, M., Bernatchez, J., Fisher, E., Goldberg, P., Herries, A. I. R., et al. (2007). Early human use of marine resources and pigment in South Africa during the Middle Pleistocene. *Nature, 449*(7164), 905–908.

Marlowe, F. W. (2005). Hunter-gatherers and human evolution. *Evolutionary Anthropology, 14*, 54–67.

Marlowe, F. W. (2010). *The Hadza: Hunter-gatherers of Tanzania*. Berkeley: University of California Press.

Marshall, L. (1962). !Kung bushman religious beliefs. *Africa: Journal of the International African Institute, 32*(3), 221–252.

Martin, L. H., & Wiebe, D. (2012). Religious studies as a scientific discipline: the persistence of a delusion. *Journal of the American Academy of Religion, 80*(3), 587–597.

Matthews, L. (2012). The recognition signal hypothesis for the adaptive evolution of religion. *Human Nature, 23*(2), 218–249.

McBrearty, S., & Brooks, A. S. (2000). The revolution that wasn't: a new interpretation of the origin of modern human behavior. *Journal of Human Evolution, 39*(5), 453–563.

McDougall, I., Brown, F. H., & Fleagle, J. G. (2005). Stratigraphic placement and age of modern humans from Kibish, Ethiopia. *Nature, 433*(7027), 733–736.

Mithen, S. (1998). *The prehistory of the mind*. London: Phoenix.

Murdock, G. P. (1965). *Culture and society: Twenty-four essays*. Pittsburgh: University of Pittsburgh Press.

Murdock, G. P. (1967). Ethnographic atlas: a summary. *Ethnology, 6*(2), 109–236.

Murdock, G. P., & Morrow, D. O. (1970). Subsistence economy and supportive practices: cross-cultural codes, 1. *Ethnology, 9*(3), 302–330.

Murdock, G. P., & White, D. R. (1980). Standard cross-cultural sample. In H. Barry & A. Schlegel (Eds.), *Cross-cultural samples and codes* (pp. 3–43). Pittsburgh: University of Pittsburgh Press.

Nishida, T., Kano, T., Goodall, J., McGrew, W. C., & Nakamura, M. (1999). Ethogram and ethnography of Mahale chimpanzees. *Anthropological Science, 107*(2), 141–188.

Norenzayan, A. (2013). *Big gods*: *How religion transformed cooperation and conflict.* Princeton, NJ: Princeton University Press.

Norenzayan, A., & Shariff, A. F. (2008). The origin and evolution of religious prosociality. *Science, 322*, 58–62.

Norenzayan, A., Shariff, A. F., Gervais, W. M., Willard, A. K., McNamara, R. A., Slingerland, E., et al. (2016). The cultural evolution of prosocial religions. Behavioral and Brain Sciences. doi:10.1017/S0140525X14001356,e0

Nunn, C. L. (2011). *The comparative approach in evolutionary anthropology and biology*. Chicago: University of Chicago Press.

Opie, C., Shultz, S., Atkinson, Q. D., Currie, T., & Mace, R. (2014). Phylogenetic reconstruction of Bantu kinship challenges main sequence theory of human social evolution. *Proceedings of the National Academy of Sciences, 111*(49), 17414–17419.

Pagel, M. (1994). Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. *Proceedings of the Royal Society B:Biological Sciences, 255*(1342), 37–45.

Pagel, M. (1999a). Inferring the historical patterns of biological evolution. *Nature, 401*(6756), 877–884.

Pagel, M. (1999b). The maximum likelihood approach to reconstructing ancestral character states of discrete characters on phylogenies. *Systematic Biology, 48*(3), 612–622.

Paige, K. E., & Paige, J. M. (1981). *The politics of reproductive ritual*. Berkeley: University of California Press.

Peoples, H. C., & Marlowe, F. W. (2012). Subsistence and the evolution of religion. *Human Nature, 23*(3), 253–269.

Piaget, J. (1929). *The child's conception of the world*. New York: Harcourt Brace.

Powell, R., & Shea, N. (2014). Homology across inheritance systems. *Biology & Philosophy, 29*(6), 781–806.

Radin, P. (1937). *Primitive religion: Its nature and origin*. New York: Viking Press.

Ragan, M. A. (1992). Phylogenetic inference based on matrix representation of trees. *Molecular Phylogenetics and Evolution, 1*(1), 53–58.

Rendall, D., & Di Fiore, A. (2007). Homoplasy, homology, and the perceived special status of behavior in evolution. *Journal of Human Evolution, 52*(5), 504–521.

Rossano, M. J. (2006). The religious mind and the evolution of religion. *Review of General Psychology, 10*(4), 346–364.

Rossano, M. J. (2007). Supernaturalizing social life: religion and the evolution of human cooperation. *Human Nature, 18*(3), 272–294.

Rossano, M. J. (2009a). The African Interregnum: The "where," "when," and "why" of the evolution of religion. In E. Voland & W. Schiefenhövel (Eds.) *The biological evolution of religious mind and behavior* (pp. 127–141). Berlin and Heidelberg: Springer-Verlag.

Rossano, M. J. (2009b). Ritual behaviour and the origins of modern cognition. *Cambridge Archaeological Journal, 19*(02), 243–256.

Ruhlen, M. (1991). *A guide to the world's languages: Classification* (Vol. 1). Stanford: Stanford University Press.

Sanderson, S. K., & Roberts, W. W. (2008). The evolutionary forms of the religious life: a cross-cultural, quantitative analysis. *American Anthropologist, 110*(4), 454–466.

Schluter, D., Price, T., Mooers, A., & Ludwig, D. (1997). Likelihood of ancestor states in adaptive radiation. *Evolution, 51*, 1699–1711.

Sheils, D. (1975). Toward a unified theory of ancestor worship: a cross-cultural study. *Social Forces, 54*(2), 427–440.

Sidky, H. (2010). Ethnographic perspectives on differentiating shamans from other ritual intercessors. *Asian Ethnology, 69*(2), 213–240.

Sosis, R., & Ruffle, B. J. (2003). Religious ritual and cooperation: testing for a relationship on Israeli religious and secular kibbutzim. *Current Anthropology, 44*(5), 713–722.

Spencer, H. (1870). On ancestor worship and other peculiar beliefs. *Fortnightly Review, 13*(7), 535–550.

Steadman, L. B., & Palmer, C. T. (2008). *The supernatural and natural selection*. Boulder: Paradigm.

Steadman, L. B., Palmer, C. T., & Tilley, C. F. (1996). The universality of ancestor worship. *Ethnology, 35*(1), 63–76.

Swanson, G. E. (1960). *The birth of the gods: The origin of primitive belief.* Ann Arbor: University of Michigan Press.

Tishkoff, S. A., Gonder, M. K., Henn, B. M., Mortensen, H., Knight, A., Gignoux, C., et al. (2007). History of click-speaking populations of Africa inferred from mtDNA and Y chromosome genetic variation. *Molecular Biology and Evolution, 24*(10), 2180–2195.

Tremlin, T. (2006). *Minds and gods: The cognitive foundations of religion*. New York: Oxford University Press.

Tylor, E. B. ([1871] 1958). Religion in primitive culture. New York: Harper.

Vitebsky, P. (2000). Shamanism. In G. Harvey (Ed.), *Indigenous religions* (pp. 55–67). London: Cassell.

Walker, R. S., Hill, K. R., Flinn, M. V., & Ellsworth, R. M. (2011). Evolutionary history of hunter-gatherer marriage practices. *PloS One, 6*(4), e19066.

Walker, R. S., Wichmann, S., Mailund, T., & Atkisson, C. J. (2012). Cultural phylogenetics of the Tupi language family in lowland South America. *PloS One, 7*(4), 1–9.

Watts, J., Greenhill, S. J., Atkinson, Q. D., Currie, T. E., Bulbulia, J., & Gray, R. D. (2015). Broad supernatural punishment but not moralizing high gods precede the evolution of political complexity in Austronesia. *Proceedings of the Royal Society of London B: Biological Sciences, 282*(1804), 20142556.

White, D. R. (1989). Focused bibliography for the standard cross-cultural sample. *Behavior Science Research, 23*(1–4), 1–145.

White, D. R. (2009). Pinpointing sheets for the standard cross-cultural sample: complete edition. *World Cultures eJournal, 17*(1), 2–223.

Wiebe, D. (2008). Does talk about the evolution of religion make sense? In J. Bulbulia, R. Sosis, R. Genet, E. Harris, K. Wyman, & C. Genet (Eds.), *The evolution of religion: Studies, theories, and critiques* (pp. 339–346). Margarita, CA: Collins Foundation Press.

Winkelman, M. (1990). Shamans and other "magico-religious" healers: a cross-cultural study of their origins, nature, and social transformations. *Ethos, 18*(3), 308–352.

Winkelman, M. (2010). *Shamanism: A biopsychosocial paradigm of consciousness and healing*. Santa Barbara: Praeger.

Woodburn, J. (1982). Egalitarian societies. *Man, 17*(3), 431–451.

Woodburn, J. (1997). Indigenous discrimination: the ideological basis for local discrimination against hunter-gatherer minorities in sub-Saharan Africa. *Ethnic and Racial Studies, 20*(2), 345–361.

Woodburn, J. (2005). Egalitarian societies revisited. In T. Widlok & W. G. Tadesse (Eds.), *Property and equality: Ritualisation, sharing, egalitarianism* (Vol. 1, pp. 18–31). New York: Berghahn.

Xygalatas, D., Mitkidis, P., Fischer, R., Reddish, P., Skewes, J., Geertz, A. W., et al. (2013). Extreme rituals promote prosociality. *Psychological Science, 24*(8), 1602–1605.

Zilhão, J. (2007). The emergence of ornaments and art: an archaeological perspective on the origins of "behavioral modernity." *Journal of Archaeological Research, 15*(1), 1–54.

Hervey Peoples (http://humanquestion.com/) is a PhD student in the Division of Biological Anthropology, Department of Archaeology and Anthropology at the University of Cambridge, UK. Her research interests include the evolution of religious behavior and the behavioral ecology of contemporary religious beliefs. She is the author of *The Human Question: What People Believe about Evolution, Human Origins, and the Beginning of Life* (Red Lion, 2003)

Pavel Duda is a PhD student in the Department of Zoology, Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic. His research interests include behavioral evolution of primates, including humans, cultural evolution, and human population history

Frank W. Marlowe is a retired lecturer in the Biological Anthropology Division, Department of Archaeology and Anthropology at the University of Cambridge, UK. His research focuses on the behavioral ecology of mating systems and cooperation, especially among foragers. He has worked with the Hadza since 1995 and is the author of *The Hadza: Hunter-Gatherers of Tanzania* (University of California Press, 2010)

Chapter III

Pygmejové pohledem evoluční biologie

Duda, P.

In: Hroníková, L., Schierová, Z. (Eds.)

Pygmejové: nejmenší lidé pohledem antropologie & Šebestova sbírka v Hrdličkově muzeu člověka PřF UK (pp. 18-48). Praha: Academia.

Pygmejové pohledem evoluční biologie

Pavel Duda

Pygmejové vždy přitahovali pozornost biologů, snad pro to, že jsou jednou z mála lidských populací, které lze definovat pomocí fyzických znaků spíše než na základě sdíleného jazyka či kultury. Navíc jsou pygmejové tradičně vnímáni jako velmi stará skupina lidstva, u které se z velké části zachoval původní způsob života našich lovecko-sběračských předků. Většina biologických studií zaměřených na pygmeje se tak věnuje buď příčinám jejich unikátního fenotypu, nebo jejich populační historii.

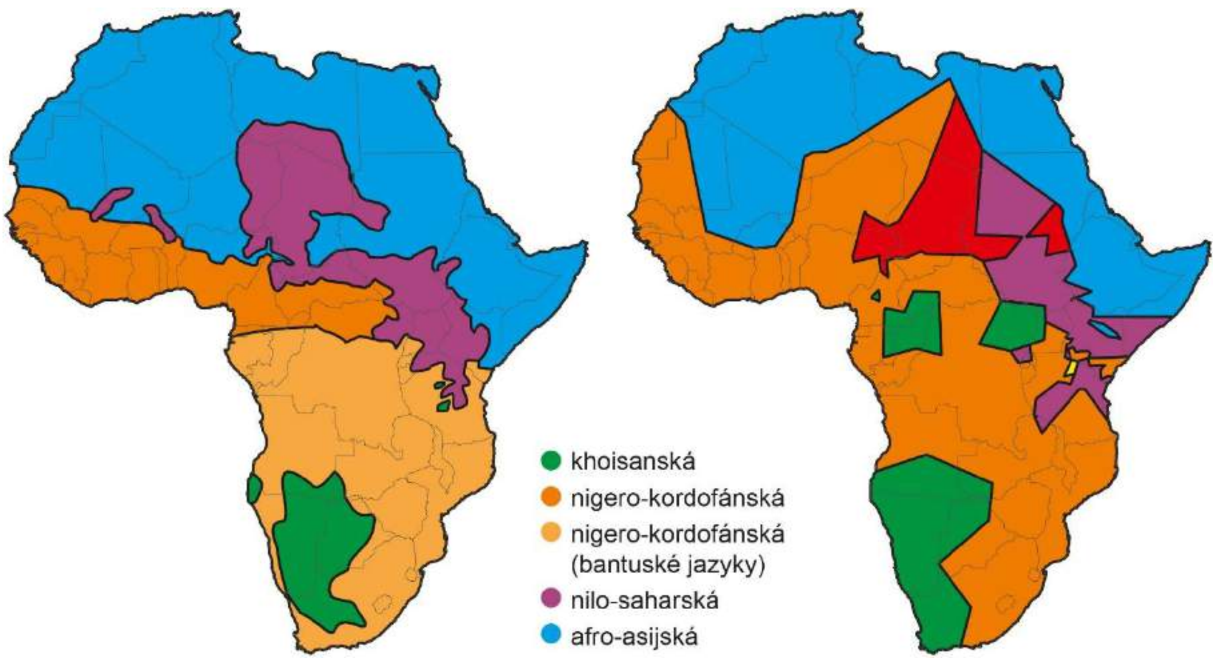
Tato kapitola se zabývá populační historií středoafrických pygmejů a asijských negritů a také záhadou původu vymřelých trpasličích lidí z ostrova Flores v Indonésii. Otázka, do jaké míry sdílejí trpasličí populace společnou historii, je důležitá pro zodpovězení dalších otázek. Co stálo za vznikem unikátních fyzických znaků a způsobem života, které pygmeje definují? Je pygmejský fenotyp adaptací? A pokud ano, k čemu slouží? Jak je tato adaptace podmíněna geneticky?

Kdo jsou pygmejové

Jak paleoantropologické výzkumy, tak populačně-genetické analýzy jednoznačně potvrzují roli afrického kontinentu coby „kolébky lidstva“. Afrika je místem, kde vývoj lidské kmenové linie (hominini) vyústil ve vznik moderního člověka (*Homo sapiens*) před přibližně 200 tisíci lety a odkud moderní lidé expandovali do zbytku světa během posledních zhruba 100 tisíc let. Lidský druh existoval na africkém kontinentě dlouho před tím, než se rozšířil na ostatní kontinenty, což vyústilo ve vysokou genetickou diverzitu dnešních Afričanů ve srovnání s ostatními lidskými populacemi.[1] Tato genetická rozmanitost koresponduje s vysokou kulturní rozmanitostí. V Africe dnes žije více než 2000 etnolingvistických skupin, bezmála třetina jejich celkového počtu. Africká etnika lze rozdělit do pěti hlavních skupin, z nichž čtyři spadají do jednotlivých jazykových rodin. Jsou to jihoafričtí mluvčí khoisanských jazyků, mluvčí nigero-kordofánských jazyků (mezi které patří také Bantuové) tvořící většinu černošského obyvatelstva subsaharské Afriky, východoafričtí mluvčí nilo-saharských jazyků a mluvčí afro-asijských jazyků obývající severní Afriku a část Arabského poloostrova. Poslední skupinu představují pygmejové. Geografické rozšíření jazykových skupin překvapivě dobře koresponduje s populačně genetickou strukturou populace kontinentální Afriky. Nejnápadnější výjimku v tomto směru představují právě pygmejové, kteří jsou jazykově nesourodí, ale přitom jasně vymezení geneticky (obr. 4).

---

[1] Campbell, M. C. – Tishkoff, S. A.: The evolution of human genetic and phenotypic variation in Africa. *Current Biology* 20, 2010, s. R166–R173.



Obr. 4 Srovnání geografické distribuce čtyř afrických jazykových rodin na základě Greenbergovy klasifikace (vlevo) a pěti ancestrálních, populačně-genetických komponent v rámci afrického kontinentu (vpravo). Komponenty v oblasti střední Afriky, narušující celkovou shodu mezi jazyky a geny, odpovídají východním a západním pygmejům a tvoří genetický celek s komponentou odpovídající jihoafrickým Khoisanům.

Afričtí pygmejové obývající Konžskou pánev se primárně dělí na dvě skupiny, východní a západní. Západní pygmejové žijí na západě Konžské pánve (od Kamerunu po západní část Demokratické republiky Kongo, DRK) v počtu zhruba 55 tisíc a zahrnují řadu skupin, mezi něž patří například Aká, Baka, Bedzan či Bongo. Východní pygmejové, souhrnně označovaní Mbuti, žijí v oblasti Ituri na severovýchodě Konžské DR v počtu zhruba 30 tisíc.[2] Dále existují ještě periferní, izolované populace pygmejů nazývané Twa nebo Cwa v centrální části DRK, ve Rwandě a Burundi.

Pygmejové jsou morfologicky, jazykově i kulturně poměrně různorodou skupinou. Variabilita průměrné výšky mezi populacemi je značná, východní pygmejové jsou menší (průměrná mužská výška je 144–145 centimetrů u východních a 152–155 centimetrů u západních pygmejů). Mbuti mají také o něco světlejší pleť a užší nos.[3] Jednotlivé skupiny východních a západních pygmejů mluví jazyky náležícími k nigero-kordofánské nebo nilo-saharské jazykové rodině, přičemž jejich genetická příbuznost a příbuznost jazyků, kterými hovoří, spolu často nekoresponduje. Zdaleka ne všichni afričtí pygmejové se živí pouze lovem a sběrem, někteří se věnují též farmaření a rybolovu. Navzdory propojení jejich lovecko-

---

[2] Batini, C. et al.: Insights into the demographic history of African Pygmies from complete mitochondrial genomes. *Molecular Biology and Evolution* 28, 2011, s. 1099–1110.

[3] Destro-Bisol, G. et al.: The analysis of variation of mtDNA hypervariable region 1 suggests that Eastern and Western Pygmies diverged before the Bantu expansion. *American Naturalist* 163, 2004, s. 212–226.

sběračské ekonomiky s ekonomikou nově příchozích pastevců a zemědělců si mnohá pygmejská etnika zachovala svou kulturní identitu.[4]

Dnešní ostrůvkovité rozšíření pygmejů je patrně reliktní, silně redukované v důsledku expanzí nigero-kordofánských a nilo-saharských pastevců a zemědělců v oblasti subsaharské Afriky a také arabské expanze v oblasti Sahary. Důkazy přítomnosti pygmejů mimo oblast jejich současného rozšíření poskytují staroegyptské dokumenty z 23. století př. n. l. Svědčí o tom, že staří Egypťané byli v kontaktu s trpasličím kmenem Aká (pozoruhodná je shoda názvu s dnes existující populací západních pygmejů) žijícím v zemi zvané Punt, v oblasti horního toku Nilu. Lingvistická a etnografická data svědčí o tom, že pygmejové žili v oblasti jižního Súdánu ještě před arabskou invazí, na konci prvního tisíciletí n. l.[5]

S využitím informací založených na kombinaci genetických, etnografických, archeologických a paleoekologických dat dokážeme vytvořit nástin populační historie afrických pygmejů, samozřejmě s vědomím, že spoustu věcí dosud nevíme.

Populační genetika a molekulární fylogenetika sehrály klíčovou roli v pochopení mnoha aspektů dávné populační historie lidstva, přičemž oba tyto obory prošly v uplynulých desetiletích zásadní proměnou. Populačně-genetický výzkum začal v 70. letech studiem tzv. klasických polymorfismů (např. krevních skupin), jejichž analýza umožňovala stanovit příbuznost studovaných populací. S nástupem automatického sekvenování a PCR[6] koncem 80. let nastoupila éra výzkumu tzv. uniparentálně dědičných markerů – mitochondriální (mtDNA) a Y-chromozomální DNA.[7] Technologický pokrok v posledních zhruba deseti letech pak umožnil takřka rutinně využívat data založená na skenování celých genomů a stovek tisíc markerů, zejména tzv. jednonukleotidových polymorfismů (*single-nucleotide polymorphisms*; SNP) a mikrosatelitů (*short tandem repeats*; STR). Získaná genomická data jsou dnes k dispozici pro zvyšující se počet lidských populací (etnik), přičemž mnohá z těchto dat jsou veřejně přístupná (např. projekty *Human Genome Diversity Project* nebo *1000 Genomes Project*).

Dnes dostupné analýzy umožňují na základě výpočtu fixačních koeficientů rozdělit studovaný soubor populací na předem definovaný počet populačně-genetických komponent. Takovému programu můžeme položit otázku: „Pokud by se soubor zahrnující stovky jedinců z desítek lidských populací reprezentujících lidstvo dělil na dvě skupiny, které by to byly? A co kdyby se dělil na pět, na 10 nebo na 20?“ Program provede ve studovaném souboru populací pomyslné řezy a efektně a přehledně popíše jejich hierarchickou strukturu. Z výsledků těchto analýz plyne, že lidstvo se primárně dělí na africké a mimoafrické. Afrika se z populačně-genetického hlediska dělí na subsaharskou a severní, která náleží k západní Eurasii. V rámci subsaharské Afriky pak pygmejové představují, spolu s mluvčími khoisanských jazyků, jednu

---

[4] Verdu, P. – Destro-Bisol, G.: African pygmies, what’s behind a name? *Human Biology* 84, 2012, s. 1–10.

[5] Cavalli-Sforza, L. L. et al.: *The History and Geography of Human Genes*. Princeton University Press, Princeton 1994, s. 158–194.

[6] Polymerázová řetězová reakce, PCR, je metoda rychlého a snadného zmnožení úseku DNA založená na principu replikace nukleových kyselin, pozn. autora.

[7] Mitochondriální DNA se dědí po mateřské linii, zatímco tzv. nerekombinující se úsek chromozomu Y se dědí po otcovské linii. V důsledku poskytuje, v některých případech, každý z těchto markerů poněkud odlišný obrázek populační historie.

ze dvou hlavních populačně-genetických komponent. To naznačuje, že pygmejové skutečně patří mezi nejstarší skupiny současného lidstva.[8,9]

Mluvčí khoisanských jazyků, především lovecko-sběračské kmeny z jihozápadní Afriky jako namibijští Ju|'hoansi (dříve nazývaní Křováci a !Kungové), vycházejí ve fylogenomických analýzách jako bazální lidská populace. K jejich oddělení od kmenové linie současného lidstva došlo podle většiny studií před více než 100 tisíci lety.[10] Středoafričtí pygmejové jsou skupinou, která se od kmenové linie lidstva oddělila po Khoisanech, případně přímo jako sesterská skupina Khoisanů, přičemž k této divergenci došlo v každém případě dávno, před více než 60 tisíci lety.[11]

Populační historie pygmejů

Existují tři hlavní teorie o vzniku středoafrických pygmejů. První teorie, jejímž autorem je francouzský antropolog Jean Hiernaux, předpokládá, že „pygmejský“ fenotyp je výsledkem dlouhodobého a nezávislého přizpůsobování se prostředí tropických deštných lesů u několika nepříbuzných skupin středoafrických lovců-sběračů.[12] Druhá teorie, jejímž autorem je italský populační genetik Luigi Luca Cavalli-Sforza, předpokládá, že pygmejové dříve tvořili v rámci střední Afriky víceméně spojitou populaci. K rozdělení na východní a západní pygmeje došlo až po příchodu zemědělců. Navzájem izolované populace se pak následkem dlouhodobých styků s nově příchozími etniky rozrůznily jazykově, geneticky i fenotypově, přičemž východní pygmejové Mbuti se nejméně odchýlili od původního stavu.[13]

Poslední teorie klade vznik pygmejů do relativně nedávné minulosti, před méně než 4000 lety, přičemž předpokládá, že jsou potomky specializovaných lovecko-sběračských „kast“ bantusky a ubangijsky mluvících zemědělců.[14] Tato teorie vlastně říká, že kamerunští a konžští pygmejové se vůči sobě mají asi jako čeští a němečtí myslivci, což ale není příliš pravděpodobné, protože zatímco populace zemědělců je poměrně geneticky homogenní, pygmejové vykazují značnou genetickou diverzitu. Navíc geneticky nejpříbuznější populací libovolného pygmejského kmene není téměř nikdy sousedící kmen zemědělců (výjimkou jsou

---

[8] Rosenberg, N. A. et al.: Genetic structure of human populations. *Science* 298, 2002, s. 2381–2385.

[9] Tishkoff, S. A. et al.: The genetic structure and history of Africans and African Americans. *Science* 324, 2009, s. 1035–1044.

[10] Schlebusch, C.M. et al.: Genomic variation in seven Khoe-San groups reveals adaptation and complex African history. *Science* 338, 2012, s. 374–379.

[11] Verdu, P. et al.: Origins and genetic diversity of pygmy hunter-gatherers from western central Africa. *Current Biology* 19, 2009, s. 312–318.

[12] Hiernaux, J.: Long-term biological effects of human migration from the African savanna to the equatorial forest: a case study of human adaptation to a hot and wet climate. In: Harrison, G. A. (ed.): *Population structure and human variation.* Cambridge University Press, Cambridge 1977, s. 187–217.

[13] Cavalli-Sforza, L. L.: African pygmies: an evaluation of the state of research. In: Cavalli-Sforza, L. L. (ed.): *African pygmies*. Orlando Academic Press, Orlando 1986, s. 361–426.

[14] Blench, R.: Are the African pygmies an ethnographic fiction? In: Biesbrouck, K. – Elders, S. – Rossel, G. (eds.): *Central African hunter-gatherers in a multidisciplinary perspective: challenging elusiveness*. CNWS, Universiteit Leiden, Leiden 1999, s. 41–60.

některé populace západoafrických pygmejů Bongo blízce příbuzné bantuským kmenům Nzebi a Teke).

Současné znalosti svědčí pro kombinaci prvního a druhého scénáře (obr. 5). Po rozdělení středoafrické populace na předky pygmejů a předky nigero-kordofánských farmářů před zhruba 60 tisíci lety se obě populace vyvíjely více méně nezávisle. K omezenému toku genů mezi nimi docházelo v posledních 40 tisících let. Populace předků bantuských farmářů přitom začala pozvolna růst už před vynálezem zemědělství, zatímco populace předků pygmejů zůstávala až do doby svého rozdělení víceméně konstantní.[15]

Populace východních a západních pygmejů sdílejí pouze staré alely[16], z čehož vyplývá, že k jejich rozdělení došlo zřejmě dávno před bantuskou expanzí. Na základě archeologických nálezů a paleobotanických dat, svědčících o existenci pralesních lovců-sběračů v oblasti Ituri před 18 tisíci lety, bylo stanoveno minimální stáří této divergence.[17] Molekulární datování (na základě mitochondriální DNA i genomických dat) poskytuje ještě starší odhady, okolo 27 tisíc let, které se překvapivě dobře kryjí s nástupem vrcholné fáze poslední doby ledové (tzv. glaciálního maxima) před 26,5–19 tisíci lety, která byla spojena s dramatickou redukcí rozlohy afrického deštného lesa.[18] Dnešní areály západních a východních pygmejů spadají do oblastí, kde se v tomto období nacházela pralesní refugia, byť o jejich přesné poloze a rozloze se vedou spory. To by znamenalo, že kontinuita jejich způsobu života nebyla přerušená po desítky tisíc let, navzdory klimatickým změnám.

Později byla pygmejská populace silně ovlivněna expanzí bantusky mluvících zemědělců. Bantuská expanze představovala jednu z nejvýznamnějších migračních událostí v historii lidstva spojenou s šířením jazyka, kultury, technologie a v neposlední řadě genů. Bantuská expanze, podnícená vznikem zemědělství a ve svých pozdějších fázích podpořená znalostí železné metalurgie, započala před zhruba 5000 lety v oblasti na hranicích jižní Nigérie a západního Kamerunu. Odtud se mluvčí bantuských jazyků šířili sub-saharskou Afrikou, až nakonec dosáhli jihu afrického kontinentu někdy v pátém století našeho letopočtu. Zemědělství a s ním spojená produkce potravin umožňuje život v mnohem vyšších populačních hustotách než lov a sběr. Zemědělci tak nad lovci a sběrači rychle získali demografickou převahu. Populace západních i východních pygmejů se postupně zmenšovala již od poslední doby ledové, ale příchod bantuských zemědělců vedl v obou těchto populacích k dramatickému „*bottlenecku*“.[19,20]

Hlavní trasa bantuské expanze vedla skrze pralesy Konžské pánve. Před asi 2000 lety dospěli mluvčí bantuských jazyků k jižnímu okraji konžského pralesa, odkud pak v rychlém sledu osídlili lesostepní oblasti jižní a východní oblasti subsaharské Afriky, od Keni po

---

[15] Verdu, P. et al.: Origins and genetic diversity of pygmy hunter-gatherers from western central Africa. *Current Biology* 19, 2009, s. 312–318.

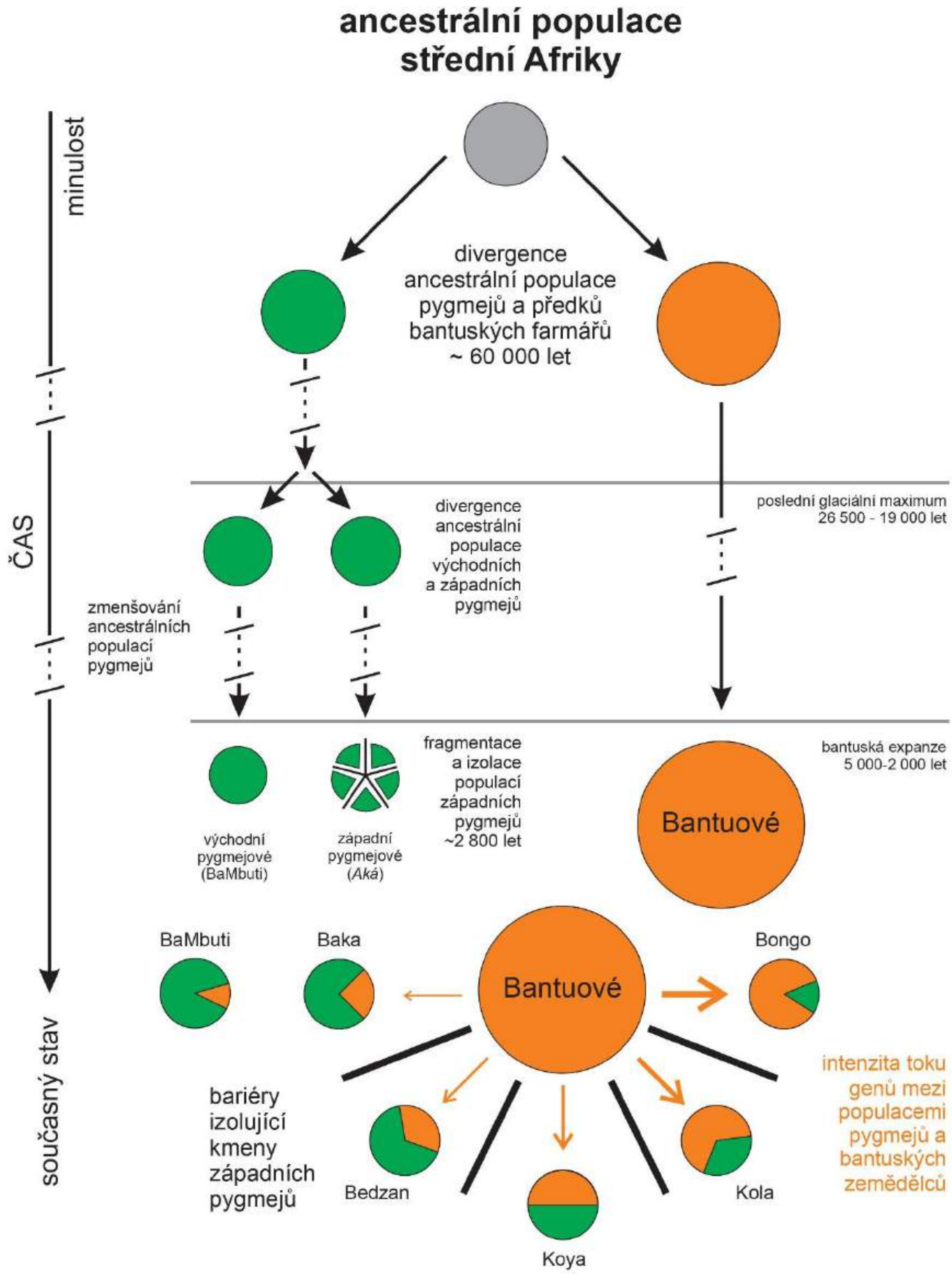
[16] Formy genů. Každý gen se může vyskytovat v jedné nebo více formách (alelách), pozn. autora.

[17] Mercader, J.: Forest people: the role of African rainforests in human evolution and dispersal. *Evolutionary Anthropology* 11, 2002, s. 117–124.

[18] Destro-Bisol, G. et al.: The analysis of variation of mtDNA hypervariable region 1 suggests that Eastern and Western Pygmies diverged before the Bantu expansion. *American Naturalist* 163, 2004, s. 212–226.

[19] Tzv. efekt hrdla láhve nastávající při prudkém zmenšení populace, které nevyhnutelně vede k vymizení některých alel, a tedy k poklesu genetické diverzity, pozn. autora.

[20] Batini, C. et al.: Insights into the demographic history of African Pygmies from complete mitochondrial genomes. *Molecular Biology and Evolution* 28, 2011, s. 1099–1110.



Obr. 5 Schematické znázornění populační historie středoafrických pygmejů a bantuských zemědělců.

Jihoafrickou republiku. Do západní oblasti rovníkové Afriky tedy zemědělství dospělo podstatně dříve než do východní části. Proto se lišil také dopad na původní lovce-sběrače žijící na východě a na západě, především míra jejich kulturní a genetické asimilace.

Východní pygmejové nebyli bantuskou expanzí zasaženi tak silně jako západní pygmejové. V genomických analýzách vycházejí Mbuti jako skupina pygmejů nejbližší Khoisanům, což je dáno tím, že u nich nedošlo k silné introgresi[21] genů nigero-kordofánských a nilo-saharských etnik. Genomy pygmejů Mbuti vykazují nejnižší podíl bantuských genů (13 %), nejvyšší naopak západní pygmejové Bedzan (54 %).[22] Podobné výsledky plynou i z analýz mitochondriální a Y-chromozomální DNA.[23,24] Tento fakt odráží i vnější morfologie východních pygmejů, která v některých aspektech (světlejší pleť, užší nos) připomíná Khoisany, v jiných (mimořádně drobná postava) je zase specificky pygmejská.

Oblast rozšíření západních pygmejů sousedí s místem vzniku bantuských jazyků a postupovala tudy bantuská expanze. Její dopad na místní pygmeje byl tudíž velmi významný. Expanze vedla před zhruba 2800 lety k dodatečné fragmentaci a izolaci pygmejských populací žijících v oblasti západní rovníkové Afriky.[25] Místní populace je dnes geneticky poměrně homogenní. Převažují v ní mladé Y-chromozomální haploskupiny spojené s bantuskou expanzí (zejména E1b1a a B2a).[26] Přesto mají západní pygmejové vyšší genetickou diverzitu než místní bantuské populace a jejich Y-chromozomální genomy stále obsahují bazální haploskupiny jako A a B2b typické pro Khoisany. Jednotlivé kmeny západních pygmejů také vykazují vyšší vzájemnou příbuznost na základě Y-chromozomálních genů než na základě jaderných genů.[27]

Y-chromozomální genomy západních i východních pygmejů obsahují vysoké frekvence starých haploskupin, které sdílí s jihoafrickými Khoisany (haploskupina B2b, přítomná v nejvyšších frekvencích u namibijských Ju|'hoansi). Přes významný dopad bantuské expanze tedy dokážeme přinejmenším v některých částech genomů pygmejů nalézt stopy dávných příbuzenských vztahů k lovcům a sběračům z jižní a východní Afriky. Naproti tomu jejich mitochondriální genomy[28] se od sebe značně liší. Sdílení starých Y-chromozomálních haploskupin mezi západními a východními pygmeji a Khoisany kontrastuje se sdílením starých mitochondriálních haploskupin mezi západními pygmeji a bantuskými zemědělci. Odlišné výsledky plynoucí z analýz Y-chromozomálních a mitochondriálních genomů jsou způsobeny

---

[21] Tok genů mezi již dříve izolovanými populacemi v důsledku hybridizace (zpětného křížení), pozn. autora.

[22] Tishkoff, S. A. et al.: The genetic structure and history of Africans and African Americans. *Science* 324, 2009, s. 1035–1044.

[23] Destro-Bisol, G. et al.: The analysis of variation of mtDNA hypervariable region 1 suggests that Eastern and Western Pygmies diverged before the Bantu expansion. *American Naturalist* 163, 2004, s. 212–226.

[24] Quintana-Murci, L. et al.: Maternal traces of deep common ancestry and asymmetric gene flow between Pygmy hunter-gatherers and Bantu-speaking farmers. *Proceedings of the National Academy of Sciences of the USA* 105, 2008, s. 1596–1601.

[25] Verdu, P. et al.: Origins and genetic diversity of pygmy hunter-gatherers from western central Africa. *Current Biology* 19/4, 2009, s. 312–318.

[26] Berniell-Lee, G. et al.: Genetic and demographic implications of the Bantu expansion: insights from human paternal lineages. *Molecular Biology and Evolution* 26, 2009, s. 1581–1589.

[27] Coia, V. et al.: Binary and microsatellite polymorphisms of the Y-chromosome in the Mbenzele pygmies from the Central African Republic. *American Journal of Human Biology* 16, 2004, s. 57–67.

[28] Haploskupina je skupina haplotypů, kombinace alel odkazující na specifické místo na chromozomu, pozn. autora.

demografickými procesy, které vedou k asymetrickému toku genů mezi sousedícími bantuskými zemědělci a západopygmejskými lovci-sběrači.

Zatímco smíšené sňatky mezi bantuskými muži a pygmejskými ženami jsou poměrně běžné, k opačným situacím dochází jen zřídka. Proto se staré, typicky pygmejské mitochondriální haploskupiny (např. L1c1) vyskytují i v genomech bantuských zemědělců, zatímco typicky bantuské mitochondriální haploskupiny v genomech pygmejů téměř nenacházíme.[29] K introgresi bantuské Y-chromozomální DNA docházelo skrze nemanželské styky s pygmejskými ženami i skrze uzavírání legitimních smíšených sňatků. Pygmejské ženy jsou mezi bantuskými muži žádané pro svou vysokou plodnost a nízkou cenu za nevěstu. Míra introgrese bantuských genů byla zjevně ovlivněna kulturními zvyklostmi a ekonomikou jednotlivých pygmejských etnik, jakož i jejich postojem vůči smíšeným sňatkům.

O různé míře sexuálně asymetrického toku genů mezi bantuskými a pygmejskými populacemi svědčí to, že ačkoli se „bantuská“ Y-chromozomální haploskupina E1b1a vyskytuje v genomech pygmejů poměrně hojně, její frekvence se mezi různými skupinami západních pygmejů značně liší (25–55 %).[30] Nejméně geneticky asimilovaní mezi západními pygmeji jsou Baka, známí svou endogamií a odmítavým postojem vůči smíšeným sňatkům. Zatímco sňatky pygmejských žen a bantuských mužů jsou obvykle přípustné, sňatky pygmejských mužů a bantuských žen jsou zpravidla tabu. Existují však výjimky, například u gabonských pygmejů kmene Bongo se vyskytují oba typy sňatků.

Po stopách zaniklého jazyka pygmejů

Více o populační historii pygmejských etnik by mohly napovědět jejich jazyky, zde ovšem narážíme na problém akulturace. Pygmejové svůj původní jazyk ztratili v důsledku kontaktu s nově příchozími zemědělskými etniky. Každá z jejich skupin dnes hovoří odlišným jazykem patřícím buďto do nigero-kordofánské jazykové rodiny (jazyky bantuské a ubangijské), nebo nilo-saharské jazykové rodiny (jazyky centrální súdánské větve). Většina pygmejů hovoří jazykem nebo dialektem jazyka některé ze sousedících skupin farmářů.

Absence příbuznosti mezi jazyky, jimiž pygmejové mluví, je vskutku zarážející. Nejnápadnější jsou v tomto směru východní pygmejovéMbuti.Mbuti zahrnují tři sousedící skupiny – Asua, Efé a Sua geneticky stěží odlišitelné a sdílející řadu kulturních znaků, nikoli však jazyk. Zatímco Efé hovoří lese – jazykem sousedních farmářů, který patří do centrální súdánské větve nilo-saharských jazyků, Asuové mají vlastní jazyk, který je lese příbuzný. Suové hovoří dialektem bantuského jazyka lega. Mbuti tedy hovoří třemi rozdílnými jazyky patřícími do dvou jazykových rodin. Je vlastně s podivem, že si udrželi mnohé ze své původní kultury i genetickou příbuznost k jihoafrickým lovcům-sběračům.

Navzdory této situaci se některé studie pokusily úspěšně identifikovat v dnešních jazycích pozůstatky původního pygmejského jazyka. Západopygmejská etnika Aká a Baka mluví

---

[29] Batini, C. et al.: Phylogeography of the human mitochondrial L1c haplogroup: genetic signatures of the prehistory of Central Africa. *Molecular Phylogenetics and Evolution* 43, 2007, s. 635–644.

[30] Berniell-Lee, G. et al.: Genetic and demographic implications of the Bantu expansion: insights from human paternal lineages. *Molecular Biology and Evolution* 26, 2009, s. 1581–1589.

vzájemně nesrozumitelnými jazyky. Aká mluví bantuskými jazyky skupiny ngondi-ngiri a Baka mluví ubangijským jazykem skupiny Gbanzili. Kdybychom se snažili o přiměr, řekli bychom, že jsou si příbuzné asi jako čeština a hindština. Navzdory minimální interakci jejich mluvčích sdílejí tyto jazyky více než 20 % svých slovníků. Téměř 90 % sdílených slov se týká lovecko-sběračských aktivit. Jde o slova specializovaného slovníku popisující lesní produkty, flóru a faunu (slova jako *krev* či *včelí vosk* a názvy pralesních zvířat, jako jsou *křečkomyš*, *štětkoun* či *poto*) a dále chování zvířat a lovecké techniky. Zbylých 10 % slov se vztahuje ke společenským aktivitám, hudbě, rituálům a náboženství.

Společný specializovaný slovník je nepochybně pozůstatkem jazyka, kterým hovořili předkové pygmejů Aká a Baka předtím, než byl jejich areál rozdělen, a každá skupina adoptovala odlišný jazyk nově příchozích zemědělců. Existence pozůstatků společného jazyka ukazuje na nepřerušenou existenci tradičního lovecko-sběračského způsobu života opírajícího se o znalost tropického deštného lesa. Původní způsob života společných předků pygmejů Aká a Baka byl zřejmě v mnohém podobný tomu současnému.[31]

Otázka, jaký byl původní jazyk pygmejů, zůstane zřejmě navždy nevyřešená. Integrace nigero-kordofánských a nilo-saharských jazyků byla příliš důkladná. Na základě genetické příbuznosti pygmejů a jihoafrických lovců-sběračů a studia hudby, coby „druhého jazyka“, můžeme pouze spekulovat o příbuznosti zaniklých pygmejských jazyků s dnešními khoisanskými jazyky.[32]

Na dnešní populaci afrických pygmejů se tedy podepsaly postupně geografie afrického kontinentu, později klimatické změny a v poslední fázi demografické a socio-kulturní procesy, které vedly k setření mnohých stop jejich společné populační historie. Vznik a populační diferenciace afrických pygmejů jsou však zjevně staré a bantuská expanze vedla pouze k větší fragmentaci již rozdělené populace. Vzhledem k dlouhé nezávislé historii jednotlivých populací středoafrických pygmejů byl na vznik paralelních adaptací dostatek času.

Asijští „pygmejové“

Současná populačně-genetická struktura obyvatel afrického kontinentu se začala formovat dlouho předtím, než se první moderní lidé vydali do Eurasie. Když se tak stalo, jejich cesta vedla nejprve na východ po jižním pobřeží Asie. Snáze se migruje v rámci jednoho podnebného pásu a cesta na východ představovala cestu nejmenšího odporu. V období před 60–40 tisíci lety lovci-sběrači kolonizovali v poměrně rychlém sledu jižní a jihovýchodní Asii, Indonésii, západní Melanésii a Austrálii, která toho času tvořila spolu s Papuou-Novou Guineou a Tasmánií jediný kontinent nazývaný Sahul.[33] Tato raná migrace moderních lidí ubírající se takzvanou „jižní cestou“ vedla ke vzniku trpasličích etnik, jejichž potomci dodnes přežívají na Andamanských ostrovech, v Malajsii, Thajsku, Indonésii, na Filipínách, Papui-Nové Guineji a

---

[31] Bahuchet, S.: Changing language, remaining pygmy. *Human Biology* 84, 2011, s. 11–43.

[32] Tishkoff, S. A. et al.: The genetic structure and history of Africans and African Americans. *Science* 324, 2009, s. 1035–1044.

[33] Macaulay, V. et al.: Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Science* 308, 2005, s. 1034–1036.

v severní Austrálii. Jejich souhrnné označení negriti pochází ze španělštiny (*negrito* je zdrobnělina slova *negro*) a odkazuje k jejich drobné postavě, tmavé pleti a kudrnatým vlasům, díky kterým působili tito lidé na evropské kolonisty v jihovýchodní Asii poněkud nepatřičným dojmem. Jejich populační historie i způsob života a soužití s nově příchozími zemědělskými etniky představují působivou paralelu afrických pygmejů.

Exemplárním reliktem původního obyvatelstva jižní Asie se zdají být obyvatelé Andamanských ostrovů, souostroví v Bengálském zálivu, které je pozůstatkem někdejšího pevninského mostu mezi Myanmarem (Barmou) a Sumatrou. Andamanci jsou pobřežní a lesní lovci-sběrači drobné postavy (nejmenší mezi negrity) a velmi tmavé pleti.

Izolované haploskupiny v mitochondriální DNA Andamanců naznačují jejich příbuznost s původními (kmenovými) společnostmi jižní Indie a také novoguinejskými a australskými domorodci.[34] Fylogenetické analýzy založené na jaderných genech a SNP ukazují, že Andamanci jsou příbuznější dnešním Asiatům než Afričanům, o jejich starobylosti a výlučnosti v rámci jižní Asie však není pochyb.[35,36] Divergence čistě andamanských mitochondriálních haploskupin je poměrně nedávná (zhruba 10 tisíc let), což odpovídá i mladému archeologickému záznamu na ostrovech (maximálně 2000 let). Dnešní Andamanci tak mohou být potomky staré, reliktní populace, která však Andamanské ostrovy osídlila relativně nedávno odněkud z oblasti dnešního Myanmaru.[37] Na Andamanských ostrovech žijí dnes čtyři populace negritů lišící se svým vzhledem i kulturou a hovořící izolovanými jazyky, snad vzdáleně příbuznými indopacifickým jazykům, kterými se dnes mluví především ve vnitrozemí Papuy-Nové Guineje.[38] Není ovšem jasné, jestli jsou si všechny andamanské jazyky vůbec příbuzné.[39]

Další populace asijských negritů žije ve vnitrozemských pralesích nedalekého Malajského poloostrova v počtu méně než 3000 jedinců. Tito lidé, označováni exonymem Semangové (slovo *semang* označuje v malajštině člověka, který upadl do otroctví kvůli dluhům[40]), tvoří jednu ze tří skupin původních obyvatel poloostrova souhrnně nazývaných *Orang Asli*. Dalšími dvěma jsou Senoi a Malajci, *Melayu Asli*.[41] Tři skupiny *Orang Asli* se od sebe liší vzhledem i způsobem života. Semangové jsou drobní, černí a kudrnatí. Jsou to kočovní, pralesní lovci-sběrači, vyskytují se především v severní vnitrozemské části poloostrova. Senoi jsou o něco vyšší postavy, mají světlejší pleť a spíše vlnité vlasy. Jsou to víceméně usedle žijící

---

[34] Endicott, P. et al.: The genetic origins of the Andaman islanders. *The American Journal of Human Genetics* 72, 2003, s. 178–184.

[35] Thangaraj, K. et al.: Genetic affinities of the Andaman islanders, a vanishing human population. *Current Biology* 13, 2003, s. 86–93.

[36] Chaubey, G. – Endicott, P.: The Andaman Islanders in a regional genetic context: reexamining the evidence for an early peopling of the archipelago from South Asia. *Human Biology* 85, 2013, s. 153–171.

[37] Barik, S. S. et al.: Detailed mtDNA genotypes permit a reassessment of the settlement and population structure of the Andaman Islands. *American Journal of Physical Anthropology* 136, 2008, s. 19–27.

[38] Whitehouse, P. et al.: Kusunda: an Indo- Pacific language in Nepal. *Proceedings of the National Academy of Sciences of the USA* 101, 2004, s. 5692–5695.

[39] Abbi, A.: Is Great Andamanese genealogically and typologically distinct from Onge and Jarawa? *Language Sciences* 31, 2009, s. 791–812.

[40] Lewis, M. P. (ed.): *Ethnologue: Languages of the World.* 16th Edition. SIL International, New York 2009.

[41] Fix, A. G.: Malayan paleosociology: implications for patterns of genetic variation amongst the Orang Asli. *American Anthropology* 97, 1995, s. 313–323.

pralesní zemědělci. *Melayu Asli* mají světlou pleť a rovné vlasy a živí se farmařením, rybařením a obchodem. Většinu obyvatel Malajského poloostrova přitom dnes tvoří potomci ekonomických migrantů z Číny, Indie a dalších oblastí, kteří se smísili s původními obyvateli.

Klasická teorie předpokládá nezávislý původ a postupný příchod těchto tří skupin na Malajský poloostrov, přičemž každá migrace vedla k částečnému nahrazení původního obyvatelstva. Podle této teorie jsou Semangové, spolu s ostatními negrity, potomky dávné migrace lovců-sběračů trpasličího vzrůstu, kteří doputovali z Afriky před více než 40 tisíci lety. Senoi odvození od indických kmenových společností přišli na Malajský poloostrov později, před zhruba 20 tisíci lety. *Melayu Asli* přišli z východní Asie jako poslední před asi 4000 lety. Alternativní teorie předpokládá vznik těchto skupin jako důsledek lokální diferenciace. Hranice mezi nimi totiž není nijak ostrá a stejně jako v případě afrických pygmejů nenacházíme jasnou korespondenci mezi morfologií, jazykem a způsobem života.

Z genetických analýz je patrné, že jihovýchodní Asie je tvořena komplikovanou a mnohaúrovňovou mozaikou různě starých genetických komponent jak lokálního, tak východoasijského původu. Ve fylogenetických analýzách založených na mitochondriální DNA a později SNP vycházejí Semangové jako bazální skupina sesterská ostatním malajským etnikům, která si v sobě uchovala největší podíl genů původních obyvatel.[42] Na druhou stranu Semangové, stejně jako Andamanci, nejsou blíže příbuzní ostatním negritům nebo dokonce africkým pygmejům než Malajcům. Myšlenku provázanosti populační historie jednotlivých vrstev obyvatel Malajského poloostrova podporuje fakt, že bezmála všechna místní etnika mluví vzájemně příbuznými jazyky patřícími do jižní větve austroasijské jazykové rodiny.

Další skupiny negritů žijí na ostrovech filipínského souostroví (například Agta a Aeta na Luzonu, Ati na ostrově Panay a Mamamwa na Mindanau) a všechny mluví austronéskými jazyky. Výsledky genetických studií ukazují, že jde o skupiny volně distribuované mezi jednotlivými větvemi místního obyvatelstva. Negriti z Andamanských ostrovů, Malajského poloostrova a Filipín si nejsou blízce příbuzní.[43,44] Liší se rovněž svou morfologií. Zatímco Andamanci a Semangové jsou svými obličejovými rysy skutečně podobní Afričanům a sobě navzájem, filipínští negriti jsou navzdory tmavé pleti a kudrnatým vlasům více podobní východním Asiatům.[45] Může jít o důsledek různě intenzivního mísení s nově příchozími skupinami obyvatelstva z východní Asie. O společném původu a někdejší provázanosti těchto populací však svědčí jejich kulturní podobnosti přetrvávající navzdory tomu, že každá z těchto skupin hovoří dnes jazykem odlišné jazykové rodiny. Především negriti z Malajského

---

[42] Hatin, W. I. et al.: Population genetic structure of peninsular Malaysia Malay sub-ethnic groups. *PLoS One* 6, e18312, 2011.

[43] The HUGO Pan Asian SNP Consortium, Mapping human genetic diversity in Asia. *Science 326*, 2009, s. 1541–1545.

[44] Migliano, A. B. et al.: Evolution of the pygmy phenotype: evidence of positive selection from genome-wide scans in African, Asian, and Melanesian pygmies. *Human Biology* 85, 2013, s. 251–284.

[45] Bulbeck, D.: Craniodental affinities of Southeast Asia's "Negritos" and the concordance with their genetic affinities. *Human Biology* 85, 2013, s. 95–133.

poloostrova a Filipín vykazují nápadnou podobnost svých náboženských představ a náboženských praktik.[46,47]

Negrity najdeme také v horách Papuy-Nové Guineje (například kmen Ok) a v tropických lesích severního Queenslandu (kmen Barrinean). V rozporu s tradiční teorií, která předpokládá, že jde o přímé potomky první vlny kolonizátorů Sahulu příbuzné asijským negritům, jsou tito lidé příbuznější „vysokým“ papuánským, respektive australským domorodcům.[48,49] Všechny doklady svědčí pro to, že Sahul byl v pleistocénu kolonizován pouze jednou a že trpasličí vzrůst některých populací je spíše výsledkem lokálních adaptací. To však nic nemění na tom, že jejich populační historie je velmi dávná, sahající do dob nejstaršího osídlení této oblasti.

Záhada floreských „hobitů“

Dnešní trpasličí populace se zmenšily nezávisle na sobě. To, že v lidské vývojové linii existuje tendence ke vzniku trpasličích forem, působivě ilustruje kauza floreských „hobitů“, *Homo floresiensis*. Objev malých, zjevně archaických homininů v jeskyni Liang Bua na indonéském ostrově Flores roku 2003 vzbudil senzaci a rozvířil neutuchající debatu mezi paleoantropology (obr. 6).[50,51]

„Hobiti“ byli vysocí asi 110 centimetrů a vážili okolo 26 kilogramů, tedy podstatně méně než kterákoli ze současných lidských trpasličích populací. Nadto měli velmi malý mozek. Objem jejich mozkovny, původními metodami stanovený na 380 cm^3, později upravený na 417 cm^3, měl velikostí mnohem blíže k mozku šimpanzů (cca 350 cm^3) než moderních lidí (cca 1300 cm^3).[52]

Nálezy z jeskyně Liang Bua jsou staré pouhých 95–13 tisíc let. *Homo floresiensis* tedy přežíval na Floresu ještě dlouho poté, co *Homo sapiens* kolonizoval Indonésii a Austrálii. O tom, že se moderní lidé museli s „hobity“ setkat, svědčí nejen stáří kosterních pozůstatků a kamenných nástrojů v jeskyni, ale rovněž v Malajsii a Indonésii rozšířené zkazky o pralesních trpaslících. Na Sumatře jsou tito tvorové, dobře známí kryptozoologům, nazýváni *Orang Pendek*. Na samotném Floresu jim místní říkají *Ebu Gogo* (což lze do češtiny přeložit jako „žravá stařena“). Tyto údajně věčně hladové bytosti žijící ve vnitrozemí ostrova chodí po setmění do vesnic krást potraviny. Někdy také unášejí místní děti, aby se od nich naučily vařit.[53]

---

[46] Tyto populace sdílejí víru v boha hromu a soubor zdánlivě arbitrárních tabu, při jejichž porušení sešle tento bůh bouři na provinilce a jeho blízké, pozn. autora.

[47] Blust, R.: Terror from the Sky: Unconventional Linguistic Clues to the Negrito Past. *Human Biology* 85, 2013, s. 401–416.

[48] Tommaseo-Ponzetta, M. et al.: Mountain pygmies of western New Guinea: a morphological and molecular approach. *Human Biology* 85, 2013, s. 285–307.

[49] McAllister, P. et al.: The Australian Barrineans and their relationship to southeast Asian negritos: an investigation using mitochondrial genomics. *Human Biology* 85, 2013, s. 485–502.

[50] Brown, P. et al.: A new small-bodied hominin from the Late Pleistocene of Flores, Indonesia. *Nature* 431, 2004, s. 1055–1061.

[51] Lahr, M. – Foley, R.: Human evolution writ small. *Nature* 431, 2004, s. 1043–1044.

[52] Falk, D. et al.: The brain of LB1, Homo floresiensis. *Science* 308, 2005, s. 242–245.

[53] Wong, K.: The littlest human. *Scientific American* 2, 2005, s. 56–65.

Obr. 6 Floreský „hobit“, Homo floresiensis, podle rekonstrukce Johna Gurcheho.

Nemalou záhadou je už sama přítomnost „hobitů“ na ostrově. Flores není součástí asijského šelfu a zřejmě nikdy nebyl dosažitelný po souši. Nejbližší spojení s pevninou představuje dnes 300 kilometrů široký průliv mezi Floresem a jižním cípem Sulawesi.[54,55]

Homo floresiensis byl původně považován za miniaturizovanou ostrovní formu *Homo erectus*. Z nálezů nástrojů víme, že populace homininů existovala na Floresu před více než 880 tisíci lety. Přestože se kosterní pozůstatky jejich autorů nedochovaly, předpokládá se, že šlo právě o tento druh, který na sousední Jávě přežíval do relativně nedávné doby, před zhruba 150 tisíci lety.[56]

Kombinace stáří nálezu, velikosti těla a malé mozkovny spolu s absencí podobných fosilií v kontinentální Asii však vzbudila podezření, že *Homo floresiensis* není samostatným lidským druhem, ale populací moderních lidí trpících růstovou patologií či genetickým

---

[54] V pleistocénu, kdy „velké Sulawesi“ zahrnovalo i ostrov Selayar, byla šířka průlivu 80 kilometru. Fosilní záznam však dokazuje, že i velká zvířata, jako například sloni, se příležitostně dokázala na Flores dostat, pozn. autora.

[55] Morwood, M. J. – Jungers, W. L.: Conclusions: implications of the Liang Bua excavations for hominin evolution and biogeography. *Journal of Human Evolution* 57, 2009, s. 640–648.

[56] Indriati, E. et al.: The Age of the 20 meter Solo River terrace, Java, Indonesia and the survival of Homo erectus in Asia. *PLoS ONE* 6, e21562, 2011.

onemocněním. Konkrétní podezření padlo na mikrocefalii.[57] „Spor o hobita“ byl, zejména v prvních letech po jeho objevu, velmi horký. Patologická hypotéza byla opakovaně vyvracena a znovu předkládána na základě srovnání morfologie „hobita“ s jinými archaickými lidmi, dnešními lidoopy, pygmeji i mikrocefaliky.[58],[59]

Ve světle současných poznatků se však tato hypotéza nezdá příliš pravděpodobná. „Hobiti“ se totiž vyznačují řadou pozoruhodných morfologických znaků, kterými se zásadně liší od *Homo sapiens*. Patří mezi ně nezvyklý tvar mozkovny, obličejová morfologie, obličej bez výrazné brady, tvar stoliček, stavba zápěstí a tělní proporce připomínající australopitéky a rané africké zástupce rodu *Homo*.[60] Přitom ne všechny tyto znaky musí být primitivní, část z nich, zejména pozměněné tělní proporce, může být naopak odvozená.[61] U dnešních mikrocefaliků nicméně nic podobného nenacházíme a ani u pygmejů se miniaturizace nepojí s podobnými evolučními reverzemi. Představa populace složené výhradně z postižených jedinců se specifickou nástrojovou kulturou a přežívající na izolovaném ostrově po desítky či stovky tisíc let je navíc poněkud přitažená za vlasy. Kosterní pozůstatky „hobitů“ svědčí o tom, že vedli aktivní život nepoznamenaný fyzickým postižením.[62]

Mimořádně drobný vzrůst „hobitů“ je spíše příkladem tzv. ostrovního nanismu. U zvířat žijících na ostrovech dochází často ke změnám tělesné velikosti oproti jejich na pevnině žijícím protějškům. Směr této změny není snadno předvídatelný (vedle ostrovního nanismu existuje i ostrovní gigantismus), efekt závisí na velikosti ostrova a jeho odlehlosti. Vesměs platí, že spíše malá zvířata se zvětšují a spíše velká se zmenšují. Výsledkem těchto změn jsou například obří želvy (*Geochelone*) ze souostroví Galapág a Seychel nebo velké ještěrky (*Gallotia*) z Kanárských ostrovů. Z nedávné minulosti známe třeba velkého nelétavého holuba dronteho (*Raphus*) z ostrovaMauricius nebo velké labutě (*Sygnus*) a trpasličí slony (*Elephas*) ze středomořských ostrovů. Nyní se k nim přidal i trpasličí (pra)člověk z ostrova Flores. Někdejší floreská fauna představovala působivou ukázku ostrovního efektu v praxi. „Hobiti“ sdíleli ostrov s trpasličími slony (*Stegodon*), velkými „krysami“ (*Papagomys*), čápy marabu (*Leptoptilos*) a obřími varany (*Varanus*).[63] Potravní vztahy uvnitř tohoto bizarního ekosystému byly vskutku kontraintuitivní. Zatímco „hobiti“ lovili zejména velké hlodavce a příležitostně též mláďata slonů, sami se občas stávali kořistí „komodských draků“ a obřích marabu. Homérova představa pygmejů jako trpaslíků válčících s velkými tažnými ptáky tak v případě „hobitů“ došla svého naplnění.

---

[57] Případně endemitní hypothyreoidální kretenismus, Laronův syndrom nebo Downův syndrom, pozn. autora.

[58] Aiello, L. C.: Five years of Homo floresiensis. *American Journal of Physical Anthropology* 142, 2010, s. 167–179.

[59] Jacob, T. et al.: Pygmoid Australomelanesian Homo sapiens skeletal remains from Liang Bua, Flores: population affinities and pathological abnormalities. *Proceedings of the National Academy of Sciences of the USA* 103, 2006, s. 13421–13426.

[60] Morwood, M. J. – Jungers, W. L.: Conclusions: implications of the Liang Bua excavations for hominin evolution and biogeography. *Journal of Human Evolution*, 2009, s. 640–648.

[61] Lyras, G. A. et al.: The origin of Homo floresiensis and its relation to evolutionary processes under isolation. *Anthropological Science* 117, 2009, s. 33–43.

[62] Kaifu, Y. et al.: Posterior deformational plagiocephaly properly explains the cranial asymmetries in LB1: a reply to Eckhardt and Henneberg. *American Journal of Physical Anthropology* 143, 2010, s. 335–336.

[63] Varani, jako jediní zástupci někdejší floreské megafauny, přežili do dnešních dnů na nedalekém ostrově Komodo a přilehlých ostrůvcích, pozn. autora.

Dvě zbývající teorie o původu *Homo floresiensis* předpokládají, že byl samostatným druhem člověka, buďto extrémně miniaturizovaným, odvozeným od *Homo erectus*, nebo ne tak dramaticky miniaturizovaným, zato však velmi bazálním odděleným kdesi na úrovni raných afrických zástupců rodu *Homo* nebo dokonce australopitéků. Obě tyto možnosti jsou svým způsobem senzační.

Pro první z těchto hypotéz svědčí zejména fylogenetická analýza morfologických znaků, která řadí „hobita“ do příbuzenství druhů *Homo habilis* a *Homo rudolfensis*. To by znamenalo, že k oddělení jeho předků od linie homininů došlo před zhruba dvěma miliony let.[64] To s sebou však nese nutnost velké migrace raných afrických lidí přes Asii, která po sobě nezanechala stopy.

Hlavním argumentem proti druhé z hypotéz bylo to, že mozek *Homo floresiensis* byl příliš malý, než aby mohl vzniknout druhotnou miniaturizací mozku *Homo erectus*. Nové měření objemu mozkovny „hobita“ s využitím počítačové tomografie však ukázalo, že předchozí odhady byly podhodnocené a objem jeho mozkovny byl ve skutečnosti větší, asi 426 cm^3. Navíc *Homo erectus* se v rámci svého areálu rozšíření vyznačoval značnou velikostní variabilitou a „hobita“ je třeba srovnávat s jeho hypotetickým předkem, tedy raným jávským *Homo erectus*, který byl o něco menší, než byl velikostní průměr tohoto druhu, a měl také menší mozek. Rozdíl ve velikosti mozku jávského a floreského člověka je zčásti důsledkem celkového zmenšení. Zbylou část rozdílu je nutné vysvětlit jiným způsobem.[65]

Byť to nebývá pravidlem, k relativnímu zmenšení mozku vůči velikosti těla u ostrovních trpaslíků občas dochází. Prokazatelně se tak stalo u vymřelého madagaskarského hrocha *Hippopotamus lemerlei* a u „kamzíka“ *Myotragus balearicus*, který žil na ostrově Mallorca.[66],[67] Zmenšení mozku, orgánu s vysokými energetickými nároky, mohlo představovat dodatečnou adaptaci k životu na ostrovech vyznačujících se vzhledem ke své omezené rozloze často menším množstvím potravy. Někdy mohlo být zmenšení mozku také důsledkem absence predátorů, díky čemuž bylo možné „oželet“ bystré smysly. V případě „hobitů“ není jasné, kterou část svých kognitivních schopností „obětovali“ výměnou za nižší energetickou náročnost svých mozků. Na Floresu byli přítomní velcí predátoři a kamenné nástroje spojované s *Homo floresiensis* nejsou oproti nástrojům *Homo erectus* nijak nápadně zjednodušené.[68]

Přestože možnost, že florešti „hobiti“ trpěli genetickým onemocněním, dnes nelze zcela vyloučit, množství důkazů naznačuje, že *Homo floresiensis* byl samostatným druhem člověka, u něhož redukce tělesné velikosti (a velikosti mozku) představovala adaptaci k životu v ostrovním ekosystému. Hobiti tak s největší pravděpodobností nemají nic přímo společného se současnými trpasličími populacemi. Případy afrických pygmejů, asijských negritů a floreských

---

[64] Argue, D. et al.: Homo floresiensis: a cladistic analysis. *Journal of Human Evolution* 57, 2009, s. 623–639.

[65] Kubo, D. et al.: Brain size of Homo floresiensis and its evolutionary implications. *Proceedings of the Royal Society B* 280, 20130338, 2013.

[66] Weston, E. M. – Lister, A. M.: Insular dwarfism in hippos and a model for brain size reduction in Homo floresiensis. *Nature* 459, 2009, s. 85–88.

[67] Köhler, M. – Moyá-Solá, S.: Reduction of brain and sense organs in the fossil insular bovid Myotragus. *Brain, Behavior and Evolution* 63, 2004, s. 125–140.

[68] Niven, J. E.: Brains, islands and evolution: breaking all the rules. *Trends in Ecology & Evolution* 22, 2007, s. 57–59.

„hobitů“ dokládají, že u lidí (homininů) se drobný vzrůst vyvinul minimálně třikrát nezávisle na sobě. Jaké jsou ale příčiny jeho vzniku?

Proč jsou pygmejové malí

Je samozřejmě možné, že postava pygmejů není žádnou adaptací, ale pouhým důsledkem dětské podvýživy u pralesních lovecko-sběračských populací. Nezdá se však, že by se na růstu pygmejů nějak podepisoval potravní stres. Předně pygmejové rostou v dětství, v období od tří do zhruba deseti let, překvapivě rychle. Ze srovnání růstových křivek žen trpasličích etnik (Aká, Agta a Aeta) a žen v západních společnostech (Američanky), u kterých podvýživa v dětství vede k podprůměrné výšce v dospělosti odpovídající výšce pygmejů, vyplývají zjevné rozdíly. Zatímco podvyživené dívky rostou pomalu a dlouho, pygmejské rostou víceméně normální rychlostí, ale přestanou růst dříve, ve 12–13 letech. V tomto věku dorostou do výšky okolo 140 centimetrů, které podvyživené Američanky dosáhnou až v 15 letech. Věk ukončení růstu odpovídá těm nejlépe živeným, nejrychleji rostoucím Američankám, které je však ve výsledku přerostou o víc než 30 centimetrů. Podobné růstové křivky vykazují také chlapci. Příčinou drobné postavy pygmejů je tudíž pomalý růst v průběhu dospívání, nikoli v dětství.[69]

Tato specifická růstová trajektorie může být podmíněna geneticky, narušením dráhy růstových hormonů. Trpasličí populace mají obvyklou hladinu růstového hormonu (GH1). Ve srovnání s ostatními lidskými populacemi však na něj jejich metabolismus reaguje v omezené míře. Rezistence vůči GH1 bývá způsobena buď sníženou hladinou jeho receptoru (GHR), nebo vazebného proteinu (GHBP), což je právě případ trpasličích etnik. Navíc mají mnohá z nich (afričtí pygmejové, malajští a filipínští negriti, ale nikoli kmen Ok z Papuy-Nové Guineje) dlouhodobě sníženou hladinu inzulinu podobného růstového faktoru 1 (IGF1), jehož hladina u dětí v ostatních lidských populacích vzrůstá v pubertě zhruba trojnásobně.[70,71]

Nadto se ukázalo, že děti pygmejských žen a bantuských mužů jsou v průměru vyšší než děti, jejichž rodiče jsou oba pygmejové. Závěry původních studií, založených na porovnávání výšky dětí z pygmejských a smíšených manželství byly později potvrzeny také geneticky. Vyšší podíl bantuských genů se projeví na vyšším vzrůstu, což dokládá, že rozdílná výška těchto etnik v sobě skutečně nese genetickou komponentu.[72,73] Lze tedy předpokládat, že trpasličí vzrůst je skutečnou adaptací na určitý aspekt lovecko-sběračského způsobu života v prostředí tropických deštných lesů.[74]

---

[69] Migliano, A. B. – Vinicius, L. – Lahr, M. M.: Life history trade-offs explain the evolution of human pygmies. *Proceedings of the National Academy of Sciences of the USA* 104, 2007, s. 20216–20219.

[70] Perry, G. H. – Dominy, N. J.: Evolution of the human pygmy phenotype. *Trends in Ecology & Evolution 24*, 2009, s. 218–225.

[71] Baumann, G. et al.: Short stature and decreased serum growth hormone- binding protein in the Mountain Ok people of Papua New Guinea. *The Journal of Clinical Endocrinology & Metabolism* 72, 1991, s. 1346–1349.

[72] Cavalli-Sforza, L. L.: African pygmies: an evaluation of the state of research. In: Cavalli-Sforza, L. L. (ed.): *African pygmies.* Orlando Academic Press, Orlando 1986, s. 361–426.

[73] Becker, N. S. et al.: Indirect evidence for the genetic determination of short stature in African Pygmies. *American Journal of Physical Anthropology* 145, 2011, s. 390–401.

[74] Diamond, J. M.: Why are pygmies small? *Nature* 354, 1991, s. 111–112.

Obr. 7 Ukázka typické proporcionality postav pygmejů.

První z klasických teorií předpokládá, že drobná postava představuje adaptaci na setrvalý potravní stres. Tropické deštné lesy se vyznačují sezónně omezenou potravní nabídkou. Například v oblasti Ituri se ovoce po určitou část roku téměř nevyskytuje a pygmejové Efé se v tomto období živí převážně housenkami. Tropické deštné lesy jsou vlastně „ekologickými ostrovy“ a pralesní formy druhů zvířat bývají oproti svým protějškům žijícím v otevřené krajině menší. Totéž platí i pro lidské populace. Hypotézu potravního stresu však zpochybňují antropologická pozorování. Pygmejové stráví lovem, sběrem a přípravou jídla zhruba půl dne,
ostrovy“ a pralesní formy druhů zvířat bývají oproti svým protějškům žijícím v otevřené krajině

menší. Totéž platí i pro lidské populace. Hypotézu potravního stresu však zpochybňují antropologická pozorování. Pygmejové stráví lovem, sběrem a přípravou jídla zhruba půl dne, nezdá se tedy, že by trpěli hladem. Na druhou stranu současný stav věcí určitě neodpovídá podmínkám, v nichž pygmejové vznikali. Mnozí z nich dnes pravidelně směňují svůj úlovek za zemědělské produkty a někteří se sami věnují zemědělství.

Druhá teorie předpokládá, že drobná postava je adaptací k životu v teplém a vlhkém prostředí. Drobná postava znamená vyšší poměr povrchu těla vůči jeho objemu, což umožňuje efektivnější ochlazování těla pomocí pocení. Menší tělo navíc produkuje méně vlastního tepla. Neexistuje však jednoznačná negativní korelace mezi teplotou a vlhkostí na jedné straně a velikostí těla na straně druhé, která by platnost této teorie potvrdila. Třetí teorie považuje drobnou postavu za adaptaci k pohybu v hustém podrostu tropických deštných lesů, která může být výhodná také při šplhání. Aby získali med, šplhají pygmejové do korun stromů, jejichž výška někdy přesahuje 50 metrů a pády tvoří nezanedbatelnou složku jejich mortality (u mužů kmene Aká stojí za sedmi procenty všech úmrtí). Je jisté, že drobná těla dokáží úspěšněji čelit krátkodobému nedostatku potravy, účinněji uvolňovat teplo a snáze se pohybovat v hustém pralesním podrostu. Žádná z těchto teorií však nedokáže vysvětlit geografické rozšíření pygmejů, tedy to, že některé populace žijí mimo oblasti tropických deštných lesů a některé v relativně chladnějším a sušším prostředí. Sezónní podvýživa nevede nutně k drobné postavě, což lze vidět u některých lidských populací, které také zažívají pravidelná období nedostatku potravy, a přesto patří mezi společnostmi s tradičním způsobem hospodaření k těm nejvyšším na světě (nilo-saharští pastevci, např. tanzanští Masajové nebo keňští Turkanové).

Vedle těchto tradičních teorií, které předpokládají, že drobná postava je adaptací sama o sobě, existuje alternativní teorie, podle které je trpasličí vzrůst pouhým důsledkem jiné adaptace. Tou je podle ní urychlení životní historie a časný nástup reprodukce v prostředí s vysokou úmrtností.[75]

Zřejmě nejvíce zarážející biologickou charakteristikou pygmejů je, spolu s jejich drobným vzrůstem, extrémně vysoká úmrtnost, jejíž hodnoty se spíše blíží těm, které zaznamenáváme u volně žijících šimpanzů. Pravděpodobnost dožití se dospělosti je u afrických pygmejů (Aká, Efé) pouhých 30–40 %, u východoasijských negritů (Agta, Aeta, Batek) nanejvýš 50 %. Očekávaná délka života po dosažení věku 15 let u všech těchto etnik je okolo 30 let, u kmene Efé pouhopouhých 20 let. Naproti tomu afričtí lovci-sběrači žijící mimo pralesy (např. Ju|'hoansi) a rovněž východoafričtí pastevci se dospělosti dožívají s pravděpodobností přesahující 60 % a očekávaná délka života po jejím dosažení je vždy více než 40 let. V prostředí s extrémně vysokou úmrtností je výhodné dospět a reprodukovat se co nejdříve. Pygmejské ženy nemají důvod s reprodukcí otálet. Vzhledem k očekávané délce života se jen malá část z nich (pouhých 13–31 %) dožije menopauzy. Pro platnost této teorie svědčí jak zmíněné růstové křivky, tak věkově specifická distribuce porodů. Plodnost pygmejských žen dosahuje vrcholu v porovnání s jinými etniky v mladém věku. Nejvíce potomků, kteří se dožijí dospělosti, mají ženy, které měly první dítě již ve věku 15 let, což je u jiných lidských populací doba nástupu reprodukce, která se pojí spíše se snížením celkového reprodukčního výstupu. Pygmejské ženy

---

[75] Migliano, A. B. et al.: Life history trade-offs explain the evolution of human pygmies. *Proceedings of the National Academy of Sciences of the USA* 104, 2007, s. 20216–20219.

jsou svou vysokou plodností známé i mezi bantuskými muži a zřejmě z tohoto důvodu tutsijští králové ve Rwandě pojímali za manželky pygmejky.

Skutečnou adaptací pygmejů tak může být modifikace životní historie vyplývající z nevyhnutelného kompromisu mezi výhodami vyššího vzrůstu a nevýhodami pozdního nástupu pohlavní dospělosti a reprodukce.

Vůbec nejčastější příčinou úmrtí u pygmejů jsou infekční choroby, což nepřekvapí vzhledem k množství patogenů v prostředí tropických deštných lesů. Proto bychom u nich mohli čekat řadu adaptací na posílení imunity. Nedávná studie hledala stopy selekce v genomu kamerunských pygmejů Baka, Bakola a Bedzan. Mezi zkoumanými variantami jednonukleotidových polymorfismů objevila řadu těch, které podmiňují fungování metabolismu, produkci hormonů a imunitu, potenciálně zodpovědných za specifické pygmejské adaptace. Nejzajímavější úsek genomu byl nalezen na chromozomu 3. Tento úsek obsahuje gen DOCK3, který ovlivňuje výšku postavy u Evropanů a také gen CISH, který posiluje imunitu vůči bakteriémiím, malárii a tuberkulóze, zároveň však inhibuje receptor klíčového růstového hormonu.[76,77] Výška je komplexní znak a tento objev rozhodně nepředstavuje úplné vysvětlení genetického pozadí pygmejského fenotypu, je spíše první vlaštovkou. Zjištění, že tělesný vzrůst a imunita mohou být vzájemně negativně korelovány, však představuje nečekanou podporu hypotézy, že trpasličí vzrůst pygmejů není důsledkem pozitivní selekce na velikost postavy, ale vedlejším produktem jiné adaptace.

Studie, které srovnávají růstové křivky, zkoumají hormonální řízení růstu a hledají geny zodpovědné za výšku postavy, ukazují, že mechanismy, které vedou k malému vzrůstu, se u jednotlivých trpasličích etnik liší. Výše uvedené hypotézy se navzájem nevylučují. Ačkoli potravní stres, vysoká vlhkost a teplota, hustý podrost a prevalence tropických chorob nevysvětlují vzrůst pygmejů samy o sobě, přispívají v různé míře k vysoké úmrtnosti. Ta mohla stát u zrodu nezávislých, konvergentních adaptací, které se pojí s drobnou postavou. Evoluce metabolismu, růstu a imunity byla vzájemně provázaná. Dnešní geografická distribuce a biologické adaptace trpasličích etnik jsou tedy důsledkem komplexních evolučních a historických procesů, kterým dnes teprve začínáme rozumět.

---

[76] Bílkovinná molekula, která se váže na specifický hormon a spouští buněčnou odezvu na něj, pozn. autora.

[77] Jarvis, J. P. et al.: Patterns of ancestry, signatures of natural selection, and genetic association with stature in Western African pygmies. *PLoS Genetics* 8, e1002641, 2012.

Chapter IV

Kulturní fylogenetika - Využití fylogenetických metod ke studiu evoluční historie jazyka a kultury

Duda, P.

In: Ovčáčková, L. (Ed.)

Biologické a evoluční teorie kultury. Praha: Academia.

(*in press*)

Kulturní fylogenetika - Využití fylogenetických metod ke studiu evoluční historie jazyka a kultury

Pavel Duda

„*…je nezpochybnitelné, že jazykové oblasti nám podávají důkazy o vzniku druhů jejich postupnou diferenciací a zachování vyšších organismů v boji o život. Dva hlavní body Darwinovy teorie mají toto společné s mnoha jinými významnými objevy, že jsou potvrzovány i v oblastech, kde zůstaly zprvu nepovšimnuty.*“[78]

(Schleicher 1863, 33)

Cílem této kapitoly je představit kulturní fylogenetiku, její historické pozadí, teoretická východiska a rozmanitá témata, kterými se zabývá.

Kulturní fylogenetika je obor, který studuje kulturu s využitím fylogenetických metod, používaných v evoluční biologii k rekonstrukci příbuzenských vztahů mezi druhy nebo populacemi. Aplikované na kulturu umožňují tyto metody rekonstruovat evoluci a fylogenezi jazyků nebo kulturních artefaktů a odpovědět tak na některé ze zásadních otázek historické lingvistiky, antropologie a archeologie. S jejich pomocí dokážeme rekonstruovat dávné lidské migrace, stanovit stáří kulturních fenoménů nebo popsat, jak jsou tyto fenomény podmíněné vnějším prostředím nebo sebou navzájem. Součástí kulturní fylogenetiky je obor, nazývaný evoluční archeologie, zabývající se rekonstrukcí příbuzenských vztahů archeologických kultur. Vedle kulturní fylogenetiky je coby příbuzný, ale nikoli totožný, vyčleňován také obor nazývaný koevoluce genů a kultury (*gene-culture co-evolution*) nebo

---

[78] „*…unbestreitbarer ist aber auf sprachlichem Gebiete die Entstehung der Arten durch allmähliche Differenzierung und die Erhaltung der höher entwickelten Organismen im Kampfe ums Dasein. Die beiden Hauptpunkte der Darwinschen Lehre theilen also mit mancher andern wichtigen Erkenntniss die Eigenschaft, dass sie auch in solchen Kreisen sich bewähren, welche anfänglich nicht in Betracht gezogen wurden.*“ (Schleicher 1863, 33)

teorie podvojné dědičnosti (*dual inheritance theory*) (viz kapitolu Jany Švorcové a Karla Kleisnera).

Uvažovat o lidských kulturách jako o analogii biologických druhů může být velmi přínosné. Fylogenetické metody jsou totiž velmi užitečné nástroje k testování starých hypotéz a vytváření nových. Evoluční biologové a humanitní vědci, zabývající se kulturní historií (historičtí lingvisté, kulturní antropologové a archeologové), si často kladou podobné otázky a přitom čelí podobným metodologických problémům, jen si to po většinu 20. století neuvědomovali (Atkinson & Gray 2005). Příběh kulturní fylogenetiky tak není jen příběhem pozoruhodných paralel mezi biologickou a kulturní evolucí, ale také paralelního vývoje vědních oborů, které se jimi zabývají - evoluční biologie a kulturní antropologie, fylogenetiky a historické lingvistiky – a toho, jak k sobě tyto obory nacházely cesty.

Pozoruhodné paralely - biologická evoluce a evoluce kultury

Americký filozof Daniel Dennett ve své knize *Darwin's Dangerous Idea* přirovnal Darwinovu myšlenku evoluce k univerzální kyselině, která naleptává bezmála každý filozofický koncept a zásadně přetváří náš pohled na svět (Dennett 1996). Není proto divu, že již od vydání knihy *O vzniku druhů přírodním výběrem* (Darwin 1859) vedli vědci a filozofové debaty o tom, zda procesy, jimž podléhá vývoj biologických druhů, mohou být aplikovatelné také na jiné fenomény.

Sám Darwin byl přesvědčen, že proces fylogeneze není omezen pouze na živé organismy. Ve své knize *O původu člověka* (1871) popsal paralely mezi evolucí biologických druhů a evolucí jazyků: „*Utváření rozličných jazyků a jednotlivých živočišných druhů, stejně jako důkazy jejich postupného vývoje jsou obdivuhodně analogické… V různých jazycích nalézáme podivuhodné shody, způsobené společným původem slov, a analogie, vyvolané podobným způsobem utváření.*“[79] (Darwin 2006, 108–109). Ve svém stěžejním díle *O vzniku druhů* (1859) obhajoval Darwin přirozenou biologickou klasifikaci založenou na příbuzenských vztazích mezi organismy. Tu přirovnával ke klasifikaci lidských etnik založené na příbuzenských vztazích mezi jazyky: „*Kdybychom měli dokonalý rodokmen lidstva, genealogické uspořádání lidských ras by umožňovalo nejlepší klasifikaci různých jazyků,*

---

[79] “*The formation of different languages and of distinct species, and the proofs that both have been developed through a gradual process, are curiously parallel… We find in distinct languages striking homologies due to community of descent, and analogies due to a similar process of formation.*” (Darwin 1871, 89–90)

kterými se nyní mluví na celém světě. Kdyby měly být zahrnuty všechny vyhynulé jazyky a všechna přechodná a pozvolně se měnící nářečí, bylo by takové uspořádání, jak myslím, jediné možné.“[80] (Darwin 2007, 292).

Je ironií, že kniha *O vzniku druhů* pojednává převážně o procesech probíhajících v rámci jednotlivých evolučních linií (anagenezi), nikoli o jejich štěpení, tedy o tom, jak nové druhy ve skutečnosti vznikají (kladogenezi). Darwin nicméně považoval kladogenezi, která je důsledkem jím popsaných anagenetických procesů, za velmi důležitou. Zřejmě proto je jedinou ilustrací v jeho knize schématický „strom života“ (dendrogram), znázorňující původ štěpících se příbuzenských linií ze společného předka (Obr. 1a).

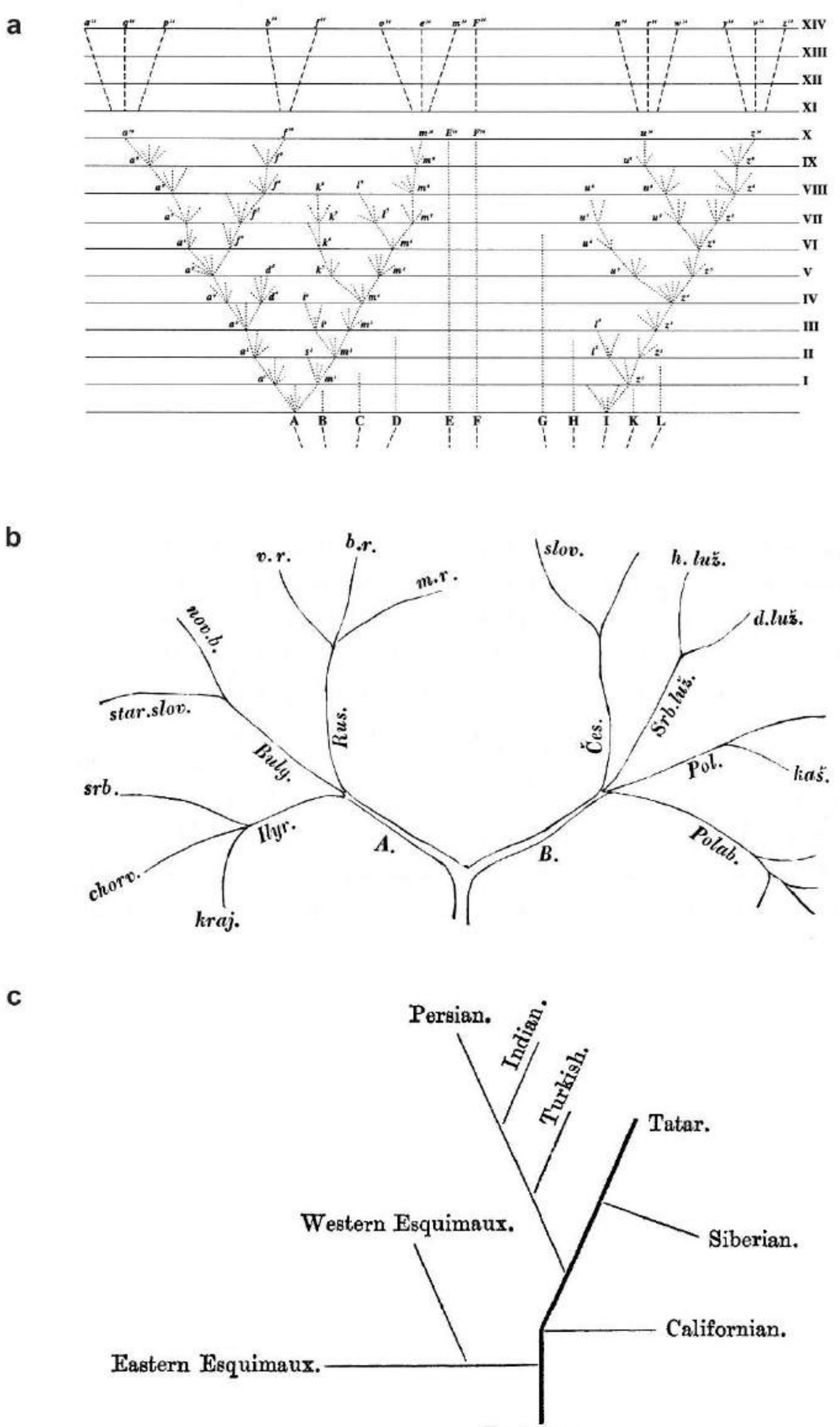
Myšlenka původu jazyků ze společného předka je ve skutečnosti starší než Darwinova evoluční teorie a rodokmeny kulturních fenoménů (jazyků a kulturních artefaktů) jsou starší než nejstarší příbuzenské stromy biologických druhů. Ernst Haeckel, velký popularizátor darwinismu, zaslal roku 1863 výtisk *Vzniku druhů* jazykovědci Augustu Schleicherovi, svému příteli a kolegovi na univerzitě v Jeně. Schleicher si po Haeckelově naléhání Darwinovu knihu přečetl a v otevřeném dopise Haeckelovi odpověděl, že jde sice o myšlenku nesmírně zajímavou, ale ne úplně novou, protože s předpoklady, které Darwin aplikuje na živý svět, jazykovědci běžně pracují, když studují historii jazyků. V letech 1850–1857, před tím, než nastoupil na Univerzitu v Jeně, působil Schleicher na pražské Karlově univerzitě. Během svého pobytu v Čechách se naučil velmi dobře česky, je mimo jiné autorem prvních českých překladů ze sanskrtu. Jeho kniha z roku 1853, obsahuje také rodokmen indoevropských jazyků. Je možné, že Schleicher tuto myšlenku přejal od svého kolegy, Františka Ladislava Čelakovského, který na Karlově univerzitě působil od roku 1849. Součástí Čelakovského posmrtně vydané knihy *Čtení o srovnávací mluvnici slovanské na universitě pražské* (1853) je také rodokmen slovanských jazyků (Obr. 1b).

Darwin sám Schleichera neznal, alespoň ne v době, kdy pracoval na *Vzniku druhů*[81]. Byl však ovlivněn názory jiných filologů, například Johna Herschela, který spekuloval o původu jazyků ze společného předka a jejich rozrůznění procesem postupných změn (Cannon 1961). V dopise geologovi Charlesi Lyellovi, který byl pro Darwina rovněž významným inspiračním

---

[80] “*If we possessed a perfect pedigree of mankind, a genealogical arrangement of the races of man would afford the best classification of the various languages now spoken throughout the world; and if all extinct languages, and all intermediate and slowly changing dialects, had to be included, such an arrangement would be the only possible one.*“ (Darwin 1859, 498)

[81] Darwin se o Schleicherově práci dozvěděl prostřednictvím jeho otevřeného dopisu Haeckelovi. V anglickém překladu vyšel v roce 1869 pod názvem *Darwinism Tested by the Science of Language*.



Obr. 1 (a) Schématický „strom života“ (dendrogram) je jedinou ilustrací v knize *O vzniku druhů* (upraveno podle Darwin 1859, 514–515), (b) dendrogram slovanských jazyků v knize *Čtení o srovnávací mluvnici slovanské na universitě pražské* (Čelakovský 1853, 3) (c) dendrogram středověkých kompozitních luků (Balfour 1889, 244).

zdrojem, Herschel argumentoval, že čas nutný k nahromadění rozdílů mezi současnými jazyky musel být podstatně delší, než kolik poskytuje biblická chronologie a svět tudíž musí být mnohem starší, než obecně předpokládá[82]. Darwin tento dopis četl a s nadšením o něm referoval své sestře Caroline. To bylo v roce 1837, více než 20 let před vydáním *Vzniku druhů* (Darwin Correspondence Project, "Letter no. 346"). Herschel nebyl zdaleka jediný, kdo věřil, že studium jazyků nám umožní nahlédnout do lidské prehistorie. William Jones, britský soudce působící v Kalkatě, si již ve druhé polovině 18. století povšiml podobností mezi sanskrtem, perštinou, klasickou řečtinou a latinou, z nichž usoudil, že tyto jazyky pocházejí ze společného předka (dnes nazývaného proto-indoevropský jazyk). Všestranný Thomas Jefferson ve své knize *Notes on the State of Virginia* (1785) spekuloval o možnosti využití lingvistických dat k rekonstrukci historických vztahů mezi kmeny severoamerických indiánů a také k odhadu stáří těchto jazyků[83].

Nejen filologové, ale také stemmatologové pracovali s klíčovou myšlenkou procesu postupného vývoje a štěpení příbuzenských linií dlouho před tím, než Darwin tyto principy aplikoval na živý svět. Vůbec nejstarší publikované příbuzenské stromy představují rodokmeny rukopisů literárních děl. Tím zřejmě vůbec nejstarším je rodokmen rukopisů Vestgótského práva, švédského provinčního zákoníku ze 13. století, který publikoval v roce 1827 Carl Johan Schlyter.

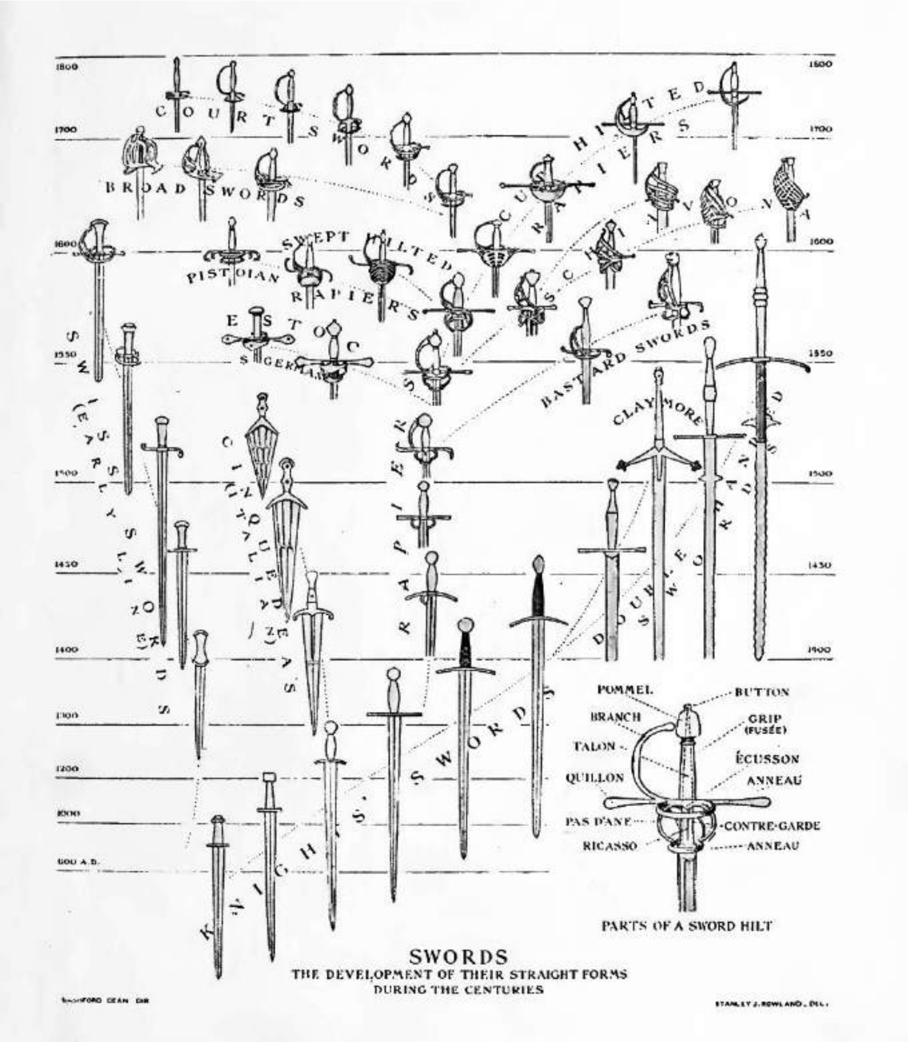
Ve druhé polovině 19. století se evolucionismus stal nesmírně vlivným názorovým proudem, který zasahoval do humanitních věd a v této době byly publikovány fylogeneze nejen jazyků ale i kulturních artefaktů (Obr. 1c, d). Přírodní a humanitní obory nebyly v té době tak zřetelně oddělené, jako je tomu dnes, a docházelo tak mezi nimi k vzájemně přínosné mezioborové cross-fertilizaci. Řada významných jazykovědců 19. století, například Jacob Grimm[84], Franz Bopp nebo samotný August Schleicher, byli zároveň nadšenými amatérskými

---

[82] "*[W]hen we see what amount of change 2000 years has been able to produce in the languages of Greece and Italy or 1000 in those of Germany, France and Spain we naturally begin to ask how long a period must have lapsed since the Chinese, the Hebrew, the Delaware and the Malesass [Malagasy] had a point in common with the German and Italian and each other.—Time! Time! Time!—we must not impugn the Scripture Chronology, but we must interpret it in accordance with whatever shall appear on fair enquiry to be the truth for there cannot be two truths.*" (Hershel 1837)

[83] "*A separation into dialects may be the work of a few ages only, but for two dialects to recede from one another till they have lost all vestiges of their common origin, must require an immense course of time; perhaps not less than many people give to the age of the earth. Agreater number of those radical changes of language having taken place among the red men of America, proves them of greater antiquity than those of Asia.*" (Jefferson 1785, 110)

[84] Bratr Wilhelma Grimma. Společně se proslavili především jako sběratelé pohádek. Jeho hlavním příspěvkem historické lingvistice je tzv. Grimmův zákon, popisující hláskové změny v germánských jazycích.



Obr. 2 Evoluce středověkých a novověkých mečů podle Bashforda Deana (1915) (nepublikováno; upraveno podle Lipo et al. 2006, 84).

botaniky. Někteří významní archeologové a etnografové té doby, například Oscar Montelius, Augustus Pitt Rivers nebo Henry Balfour, studovali evoluci a fylogenezi kulturních artefaktů (Obr. 1c). Za zmínku stojí osobnost Bashforda Deana, amerického zoologa, který se zabýval systematikou devonských paryb (žraloků, chimér a pancířnatců) a zároveň historií vojenské výzbroje. Dean publikoval řadu rodokmenů zachycujících anagenetické a kladogenetické procesy v evoluci středověkých a novověkých zbraní a brnění (Obr. 2). Dodnes zůstává jediným člověkem, který byl zároveň zaměstnancem Amerického přírodovědného muzea a Metropolitního muzea umění (obě instituce sídlí na newyorském Manhattanu, naproti sobě přes Central Park).

Evolucionismus byl tím prvním myšlenkovým proudem v antropologii, který stál u zrodu antropologie jako vědního oboru. Jeho představiteli byli například Herbert Spencer, Lewis Henry Morgan, Edward Burnett Tylor či James George Frazer (viz kapitolu Lindy Hroníkové). Zásadním problémem evolucionistické antropologie bylo, že její představitelé, přestože se na Darwina rádi odvolávali, chápali evoluci špatně, nikoli jako proces postupných změn v příbuzenských liniích, odvozených od společného předka, ale jako pokrok. Nezadržitelný, všeobjímající pokrok. Na základě předpokladů, že všechny lidské společnosti procházejí předem danými stadií (od matriarchátu k patriarchátu, od animismu přes polyteismus k monoteismu apod.) v procesu vývoje směřujícího od divošství přes barbarství k civilizaci, pak tito antropologové klasifikovaly etnika podle domnělé „kulturní vyspělosti". V podstatě pohlíželi na současná etnika jako na reprezentace různých vývojových stádií lidské kultury. Tyto unilinealistické představy o kulturní evoluci byly ve 20. století nahrazeny difuzionalismem a důsledným kulturním relativismem, představovanými Franzem Boasem a jeho žáky. Přesvědčení, že lidské kultury se vyvíjejí odděleně, nezávisle na ostatních a mohou tak nabývat nesrovnatelných podob, se stalo základním paradigmatem moderní antropologie.

Paralelní vývoj evoluční biologie a historické lingvistiky nicméně pokračoval dále do 20. století. Koncem 19. století Paul Brugmann zdokonalil metody historické lingvistiky, když zavedl rozlišování slov na inovace a retence a správně rozpoznal, že sdílené inovace jsou pro odvození příbuzenských vztahů mezi jazyky mnohem cennější než sdílené retence. V podstatě tak popsal klíčové koncepty fylogenetické systematiky (kladistiky), pouze s odlišnými názvy pro *apomorfie* (inovace) a *plesiomorfie* (retence), *synapomorfie* (sdílené inovace) a *symplesiomorfie* (sdílené retence), popsané o více než půl století později německým zoologem Willi Hennigem. Koncem padesátých a počátkem šedesátých let 20. století došlo k významným metodologickým posunům v evoluční biologii díky zavedení metody numerické taxonomie (fenetiky), jejímiž autory jsou Robert Sokal a Peter Sneath a popsání principu molekulárních hodin, jehož autory jsou Émile Zuckerkandl a Linus Pauling. O něco dříve, v 50. letech, však byly Morrisem Swadeshem popsány analogické metody v historické lingvistice, nazvané lexikostatistika a glottochronologie. Tyto metody umožňovaly odvodit příbuzenské vztahy mezi jazyky z distanční matice založené na procentech sdílené slovní zásoby a stanovit stáří divergencí jazykových linií na základě předpokladu o rychlosti jazykových změn. Swadeshovy metody byly přes počáteční nadšení později tvrdě kritizovány a v současnosti jsou minimálně využívány[85]. Zatímco biologové přijali fylogenetické metody za své, historičtí lingvisté a kulturní antropologové vůči nim zůstávali po větší část 20. století značně skeptičtí (Atkinson & Gray 2005).

Přes nedůvěru humanitních vědců k evolučním přístupům docházelo pravidelně k pokusům o překlenutí obou oborů. Zřejmě nejslavnějším z těchto pokusů je memetika (viz kapitolu Jaroslava Flegra). Přes slibné začátky a ohlas u odborné i laické veřejnosti však memetika nebyla coby vědní obor příliš úspěšná. Přinejmenším nikdy nedala vzniknout empirické linii výzkumu a v posledních deseti letech se vědecké literatury pozvolna vytratila.

Zatímco memetika začala chytlavou metaforou v jedné v nejvlivnějších populárně-naučných knih 20. století, začátky kulturní fylogenetiky byly mnohem skromnější. Teoretické základy tohoto oboru položili v 70. letech biologové Luigi Luca Cavalli-Sforza a Marcus Feldman, působící na Stanfordově univerzitě. Cavalli-Sforza a Feldman popsali principy kulturní evoluce s využitím matematických modelů vycházejících z populační genetiky a epidemiologie. O další rozvoj oboru se v 80. letech zasloužili antropologové Roberd Boyd a Peter Richerson, působící na Kalifornské univerzitě v Berkeley, kteří obor později zpopularizovali v knize *Not by genes alone* (v českém překladu vyšla pod názvem *V genech není všechno*). V 70. a 80. letech docházelo zároveň k prudkému rozvoji populačně-genetických metod a k jejich využívání ke studiu lidské prehistorie. První fylogeneze lidských populací byly založené na analýze tzv. klasických polymorfismů (např. krevních skupin). S nástupem automatického sekvenování DNA a PCR koncem 80. let započala éra výzkumu matrilineárně přenosné mitochondriální DNA a o později patrilineárně přenosného nerekombinujícího úseku mužského chromozomu Y. V roce 1988 publikoval Cavalli-Sforza se svými kolegy práci, která porovnávala fylogenetický strom lidských populací, získaný analýzou distanční matice založené na 120 genech, s „fylogenezí“ založenou na jazykové klasifikaci na úrovni jazykových rodin a velkorodin, jejímž autorem je Merritt Ruhlen. Ačkoli byla studie tvrdě kritizována, především kvůli použití kontroverzní jazykové klasifikace, upozornila na paralely mezi genetickou a kulturní evolucí a podnítila novou éru zájmu o rekonstrukci lidské prehistorie na základě kulturních, nejen genetických dat.

---

[85] Metodologické problémy lexikostatistiky lze rozdělit do tří kategorií. Zaprvé, převedením lexikálních dat do distanční matice dochází ke ztrátě informace, což snižuje schopnost metody rekonstruovat příbuzenské vztahy. Zadruhé, používaná shluková analýza spojuje spíše jazyky, které se vyvíjejí podobnou rychlostí, než jazyky, které jsou si příbuzné. Zatřetí, předpoklad, že rychlost jazykové změny je konstantní v prostoru a v čase a napříč sémantickými kategoriemi, je nerealistický, což v důsledku činí nerealistickou také výslednou topologii stromu a jeho datování (odhad stáří jednotlivých uzlů).

V devadesátých letech docházelo k akumulaci stále většího množství genetických dat a s ním k rozvoji evolučně-biologických metod. Distanční metody byly postupně opouštěny na úkor metod založených na kladistické (parsimonní) analýze homologických znaků, které nejsou zatíženy stejnými metodologickými problémy jako fenetika. Pokročilejší likelihoodové a bayesiánské metody fylogenetické analýzy umožnily ještě spolehlivější rekonstrukci příbuzenských vztahů, včetně jejich časové škály. Zároveň docházelo k rozvoji fylogenetických komparativních metod, umožňujících publikované fylogenetické stromy dále využívat k rozmanitým účelům, jako je rekonstrukce ancestrálních znakových stavů, měření znakových korelací se zohledněním fylogenetických vztahů nebo testování evolučních modelů. Přibývalo výzkumníků, rekrutujících se jak z přírodních, tak z humanitních oborů, kteří tyto metody aplikovali na lingvistická, etnografická i archeologická data, často v kombinaci s daty genetickými. Tito výzkumníci jako Mark Pagel, Ruth Maceová, Clare Holdenová, Stephen Lycett, Stephen Shennan, Russel Gray, Kevin Lalland a další položili základy kulturní fylogenetiky jako vědeckého oboru, postaveného na syntéze kulturních dat a evolučně-biologických metod (Lipo et al. 2006, Mace et al. 2005, Mace & Pagel 1994).

Dnes je kulturní fylogenetika dynamicky se rozvíjejícím vědním oborem, zabývajícím se celou řadou aspektů kulturní evoluce. Studuje například, jak se mechanismy kulturního přenosu mění v průběhu historie nebo jak se liší pro různé typy kulturních znaků a také do jaké míry horizontální přenos narušuje stromovitý proces kulturní evoluce. Část oboru je zaměřena na rekonstrukci fylogenetických vztahů mezi jazyky či kulturními artefakty, část na využití publikovaných fylogenezí k testování rozmanitých hypotéz o evoluci kultury.

Kultura jako evoluční systém - variabilita, dědičnost a selekce

Existuje velké množství odlišných definic kultury (viz např. Soukup 2004). Autorem jedné z nejznámějších je Edward Burnett Tylor, který kulturu definoval jako „*komplexní celek, který zahrnuje poznání, víru, umění, právo, morálku, zvyky a všechny ostatní schopnosti a obyčeje, jež si člověk osvojil jako člen společnosti.*“[86] (viz kapitolu Lindy Hroníkové). Taková amorfní definice, která charakterizuje kulturu jako „vše lidské“ není pro zkoumání procesů, jakými se kultura vyvíjí v prostoru a v čase, zvlášť přínosná.

---

[86] „…that complex whole which includes knowledge, belief, art, morals, law, custom, and any other capabilities and habits acquired by man as a member of society.“ (Tylor 1871, 1)

Vědci, zabývající se kulturní evolucí definují kulturu úsporně a pragmaticky jako *„soubor informací, které získává jedinec od ostatních příslušníků svého druhu prostřednictvím učení, napodobování a jiných forem sociálních přenosu, jež jsou schopné ovlivňovat jeho jednání*“ (Richerson & Boyd 2005, 17), přičemž touto informací může být *myšlenka, znalost, dovednost, názor, hodnota* či *postoj*. Kulturní evoluce jak pak definována jako proces selektivního přežívání úspěšných kulturních variant přenášených na jiné jedince prostřednictvím sociálního přenosu i jiných, ne-selektivních procesů jako je drift, inovace či migrace (Laland & Brown 2011).

Nelze tvrdit, že tato definice kultury je správnější než jiné, je ale třeba si ujasnit, co kulturní fylogenetici vlastně studují, když studují kulturu. Obecná definice, která umožnuje kulturu atomizovat, rozdělit ji na jednotlivé elementy, činí kulturní fenomény přístupnější formální analýze (umožňuje popsat frekvenci těchto elementů v populaci nebo úspěšnost, s jakou se dědí a šíří). Obecnější definice kultury otevírá možnost nacházet prvky kulturního či protokulturního chování u zvířat a skutečně, takových případů, včetně výroby a používání jednoduchých nástrojů, dnes známe celou řadu (viz kapitolu Jitky Lindové).

Většina z nás si pod pojmem evoluce představí evoluci biologickou ve spojení s populačně-genetickými koncepty. Můžeme se proto ptát: Není koncept „evoluce“ aplikovaný na lidskou kulturu pouhou metaforou? Aplikace formálních fylogenetických metod při studiu evoluce kultury vyžaduje, aby procesy biologické a kulturní evoluce byly plně srovnatelné.

Evoluce, jak ji definoval Darwin (1859), je důsledkem společného původu a postupných změn v příbuzenských liniích (Darwinovými slovy „(*common*) *descent with modification*“) a působení přirozeného výběru („*natural selection*“). Každý evoluční systém vyžaduje existenci variability jedinců (entit) v populaci, existenci mechanismu, který zajišťuje dědičnost alespoň části této variability a existenci selekce, která vede k rozdílné úspěšnosti, s jakou se různé vlastnosti jedinců dědí, a to v závislosti na podmínkách prostředí (varianta, která je úspěšná v určitých podmínkách, nemusí být stejně úspěšná v podmínkách, jaké panují jinde). Nic víc. Nezáleží na tom, jakým způsobem je variabilita mezi jedinci generovaná, co činí různé vlastnosti úspěšnými (co podmiňuje jejich *fitness*) nebo jaké jsou konkrétní mechanismy dědičnosti těchto vlastností. Evoluce je prostým důsledkem procesů zahrnujících variabilitu, dědičnost a selekci. Není tudíž pravda, že pojem „evoluce“, pokud není aplikován pouze na živé organismy, není ničím víc než pouhou analogií. Každý systém, řídící se těmito třemi základními principy, je evoluční.

Je nutné zdůraznit, že ačkoli je kultura evolučním systémem, procesy kulturní evoluce nejsou ve všech směrech analogické procesům evoluce biologické (viz kapitolu Jaroslava Flegra). Některé důležité rozdíly jsou popsány níže.

Variabilita je nezbytným předpokladem procesu evoluce. Rozmanitost lidské kultury je skutečně omračující. Jednotlivé lidské populace se liší svým sociálním uspořádáním, reprodukčním a rodičovským chováním, společenskými institucemi, tradicemi a normami. Praktikujeme monogamii, polygynii i polyandrii, platíme věno nebo cenu za nevěstu, naše příbuzenství odvozujeme po otcovské či mateřské linii či jinak, stejně tak náš sociální status a majetek dědíme podle různých pravidel. Budujeme si obydlí z rozličných materiálů, různých podob a velikostí. Obživu si zajišťujeme rozmanitými způsoby a s využitím široké škály technologií. Jsme lovci, sběrači, pastevci a zemědělci a konzumujeme různorodou stravu, semeny trav počínaje a velrybami konče. Hovoříme více než sedmi tisíci jazyky (Lewis et al. 2016). Ruku v ruce s rozmanitostí jazyků jde rozmanitost kulturních produktů lidské zkušenosti a představivosti, rituálů a kulturních praktik.

Lidská kultura nejenže se co do rozmanitosti vyrovná biologické diverzitě, ale je také podobně geograficky distribuovaná. Vyšší diverzita jazyků je obecně v tropických oblastech, na ostrovech a v horách. Vůbec nejvyšší diverzitu jazyků nacházíme na Nové Guinei (což je tropický ostrov s hornatým vnitrozemím). Na území státu Papua-Nová Guinea[87] s přibližně 7,6 miliony obyvatel se mluví 839 původními jazyky, což činí zhruba 12 % jazyků světa. Naproti tomu na území Číny, která má 1,4 miliardy obyvatel a jejíž rozloha je oproti Papui-Nové Guinei zhruba dvanáctinásobná, se mluví jen 274 původními jazyky, z nichž jen 56 je oficiálně uznáváno čínskou vládou.

Jazyková diverzita je takto nenáhodně distribuovaná a koreluje se stejnými ekologickými faktory jako diverzita druhová. Mezi tyto faktory patří délka vegetační sezóny, primární produktivita ekosystému, míra evapotranspirace, průměrný roční úhrn srážek, průměrná roční teplota apod. V distribuci kulturní (jazykové) diverzity pak v důsledku pozorujeme stejné trendy jako v distribuci druhové diverzity, například latitudinální gradient diverzity (snižující se početnost druhů i jazyků ve vyšších zeměpisných šířkách) nebo

---

[87] Stát, zahrnující východní polovinu ostrova Nová Guinea a Bismarckovo souostroví (melanéské ostrovy Manus, Nová Británie, Nové Irsko a Bougainville a mnoho menších ostrovů).

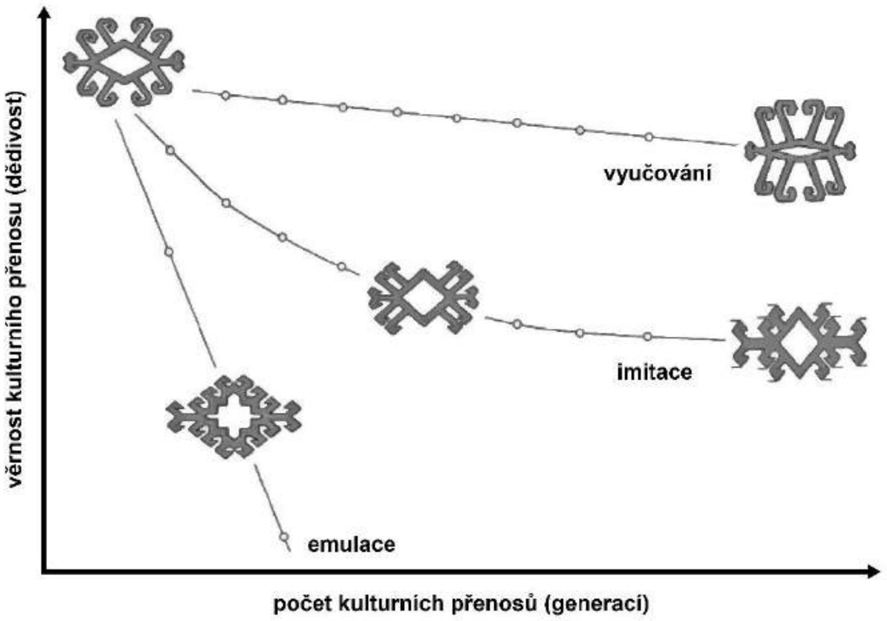
Rappaportovo pravidlo (zvětšující se areály druhů i jazyků ve vyšších zeměpisných šířkách) (Mace & Pagel 1995, Moore et al. 2002).

Rozdílná úspěšnost kulturních variant je tím, co proces kulturní evoluce pohání. Klíčové je, aby mezi sebou jednotlivé kulturní varianty reálně soupeřily. Že k takovému soupeření dochází, je patrné z distribuce kulturní diverzity, například diverzity jazyků. Z přibližně 7 100 jazyků jen asi 300 má více než 1 milion mluvčích, přičemž těmito 5 % jazyků hovoří téměř 95 % procent světové populace. Více než 90 % všech věřících jsou následovníky jednoho z pěti náboženství (z přibližně 4 300 existujících).

Kulturní varianty nesoupeří o „přežití“. Soupeření se vede mezi omezeným počtem funkčně ekvivalentních řešení určitého problému a o naši omezenou pozornost a paměťovou kapacitu. Psychologové například popsali vzájemnou interferenci při vybavování se významově podobných (konkurenčních) slov (Baddeley 1997). Archeologové popsali změny frekvencí různých typů artefaktů (např. hrotů kopí a šípů) v archeologickém záznamu, kdy zvýšení četnosti jednoho typu vede ke snížení četnosti alternativních (konkurenčních) typů (Lyman et al. 2009).

Kulturní linie mohou, stejně jako biologické linie, eventuálně vyhynout v důsledku kompetice nebo driftu. Dobře známé jsou případy vymřelých jazyků nebo jazyků ohrožených vyhynutím. Totéž se však týká i tradic a technologií. Za všechny „kulturní extinkce“ uveďme alespoň vymizení střelných zbraní v Japonsku v 17. století n. l. (Perrin 1979), kánoí na Gambierových ostrovech ve Francouzské Polynésii ve 12. století n. l. (Hiroa & Buck 1942) nebo kostěných nástrojů na Tasmánii před zhruba čtyřmi tisíci lety (Henrich 2004).

Příklad Tasmánie je mimořádně zajímavý. Tasmánie, ostrov o rozloze 67 800 km^2 (o něco menší než Česká republika), byl osídlen před nejméně 38 tisíci lety lidmi, kteří sem přišli z Austrálie suchou nohou (v pleistocénu tvořila Tasmánie s Austrálií a také s Novou Guineou jediný pevninský celek nazývaný Sahul). Na konci pleistocénu, před 10-12 tisíci lety, když stoupla hladina světového oceánu, se Tasmánci ocitli v izolaci, odříznutí od Austrálie Bassovým průlivem. Při prvním kontaktu s Evropany v roce 1642 žilo na ostrově 5–10 tisíc lidí, jejichž nástrojová kultura byla výrazně jednodušší a chudší než kultura jejich australských protějšků (údajně znali pouhých 24 typů nástrojů). Z holocénního archeologického záznamu na ostrově víme, že Tasmánci v průběhu deset tisíc let trvající izolace pozbyli mnohých kulturních vymožeností, které zahrnovali například kostěné nástroje (rybářské háčky, jehly) nebo nástroje



Obr. 3 Schéma znázorňující věrnost replikace jednotlivých způsobů kulturního přenosu – emulace, imitace a sociálního učení, zde na příkladu ornamentálního designu turkmenských koberců (upraveno podle Tehrani & Riede 2008, 326).

s rukojetí (palice, kopí s kamenným hrotem). Tasmánci tak ztratili například znalost rybolovu nebo šití. Z archeologického záznamu je patrné, že ryby dříve tvořily podstatnou část jídelníčku Tasmánců a sešívané kožešinové oděvy jim zřejmě umožnily přestát období posledního glaciálního maxima na ostrově, který se nachází poměrně hluboko na jihu. V době kontaktu s Evropany však Tasmánci nelovili ryby, pouze lachtany, škeble a korýše, a neznali oděvy. Namísto toho si potírali tělo směsí lachtaního tuku, popela a okru.

Příčiny podobných kulturních extinkcí nejsou úplně jasné. Roli hraje, vyjádřeno evolučně-biologickými pojmy, změna selekčního režimu, izolace a drift v malé, izolované populaci. Na udržení kulturních variant je potřebná dostatečná velikost a hustota populace, jinak tyto varianty z populace vymizí působením stochastických procesů, podobně jako z populace vymizí alely v důsledku genetického driftu. Dostatečná velikost populace je pro vznik a udržení kulturní komplexity velmi důležitá. Dnes se zdá, že i takzvaná svrchně paleolitická kulturní revoluce, která přinesla zásadní změnu rozmanitosti lidské nástrojové i symbolické kultury, byla spuštěna demografickou změnou (zvýšení lidské populační hustoty v Africe) a ta byla zase zpuštěna změnou klimatu (Powell et al. 2009).

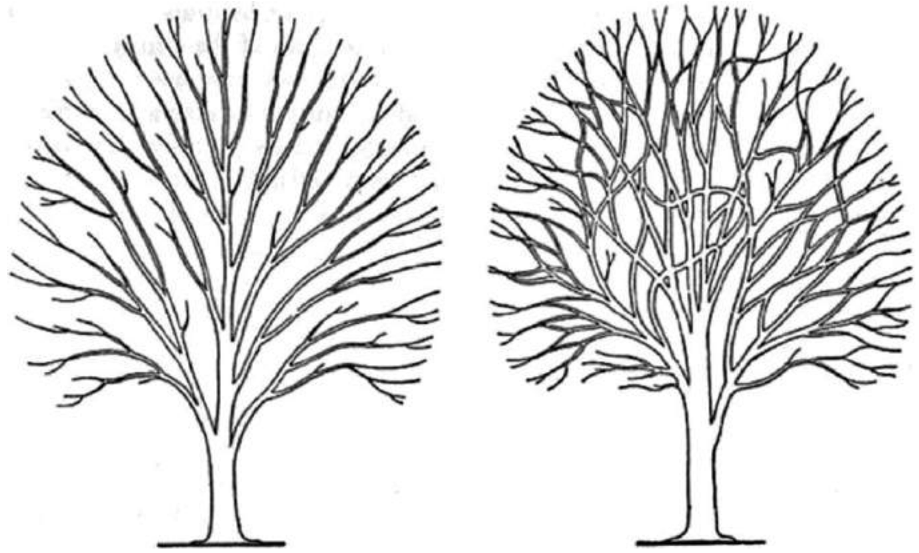
Dědičnost kultury je zajištěna kulturním přenosem - sociálním učením. Sociální učení je ve skutečnosti směsný pojem zahrnující několik odlišných způsobů kulturního přenosu, lišící se mechanismem a věrností replikace (Obr. 3). Těmito způsoby je posílení podnětu, emulace, imitace a vyučování (viz kapitolu Jitky Lindové). Posílení podnětu (*stimulus enhancement*) popisuje proces, při němž vystavení jedince určitému chování zvyšuje pravděpodobnost, že si jedinec toto chování časem sám osvojí (např. chození na koncerty klasické hudby nás inspiruje k tomu, začít se učit hrát na housle). Jedná se o poměrně jednoduchou formu učení, experimenty i počítačové simulace nicméně ukazují, že i tento způsob učení může vést ke vzniku dlouhodobě přetrvávající tradice (Lycett 2015). Emulace je proces, kdy předmětem nápodoby je výsledný produkt chování, ale postup jeho vytváření se učíme sami (např. se učíme hrát na housle skladbu, kterou známe z poslechu). Imitace je proces, který zahrnuje přesnou nápodobu určitého postupu, předváděného demonstrantem (např. pozorujeme někoho při hře na housle a snažíme se napodobit jeho techniku). Vyučování je proces, kdy vyučující doprovází demonstraci činnosti výkladem, žáka opravuje, případně mu vede ruku (tak jako při placeném vyučování hry na housle). Na uvedených příkladech si lze snadno představit, jak se všechny výše uvedené způsoby učení mohou kombinovat při osvojování si určitého chování.

V kulturní evoluci se dále uplatňují selekční faktory, takzvané selektivní biasy (*selective biases*). Ty se dělí se na takzvané obsahové (*content biases*), modelové (*context biases*) a frekvenční (*frequency biases*) (Mesoudi 2011). Obsahové biasy zahrnují tendence k přednostnímu osvojování si kulturních variant na základě funkčních a estetických kritérií. Modelové biasy zahrnují tendence k přejímání variant na základě různých společenských vlivů, přejímání častých variant nebo naopak vzácných variant (konformita a nonkonformita), napodobování úspěšných nebo vysoce postavených jedinců, sobě podobných jedinců, jedinců určitého pohlaví a podobně, které se podepíší na průběhu kulturní evoluce (viz kapitolu Petra Turečka a Jana Havlíčka).

Vertikální a horizontální kulturní přenos - fylogeneze a etnogeneze

Izolace kulturních linií je zajišťována řadou modelových biasů, etnocentrismem, jazykovými a geografickými bariérami a v neposlední řadě i násilím mezi komunitami. Přesto kultura v různé míře „difunduje“ napříč kulturními, jazykovými i geografickými oblastmi.

Odklon od evolučních přístupů v antropologii a historické lingvistice na přelomu 19. a 20. století nastal mimo jiné proto, že nastupující představitelé „difuzionistické antropologie“ si



Obr. 4 Strom života a strom lidské kultury (Kroeber 1923, 68).

uvědomily, že způsob šíření kultury nevyhovuje unilinealistickým představám evolucionistických antropologů o vývoji kultury. Antropologové a jazykovědci, např. Franz Boas nebo Roman Jakobson, začali zdůrazňovat rozdíly mezi biologickou a kulturní evolucí, spočívající v tom, že zatímco vztahy mezi organismy lze ilustrovat vhodně pomocí fylogenetického stromu, vztahy kultur či jazyků nikoli, protože kulturní elementy se přenášejí nejen vertikálně, z generace na generaci, ale také horizontálně, mezi nepříbuznými jedinci a napříč generacemi. Zatímco fylogenezi organismů lze popsat pomocí postupného vývoje a divergence příbuzenských linií (anageneze a kladogeneze), ve vývoji kultury dochází velmi často k situacím, kdy linie, které se dříve rozdělily, splývají opět dohromady (syngeneze) díky procesům jako jsou výpůjčky, akulturace a asimilace. Tyto procesy se začaly nazývat kulturní difuze nebo etnogeneze. Tyto názvy mají za cíl zdůraznit, že mechanismy vzniku a vývoje živočišných druhů a lidských kultur jsou zásadně odlišné. Důraz na tyto odlišnosti, vyjádřený ilustrací, jejímž autorem je Boasův žák Alfred Kroeber (Obr. 4), pak na dobré půlstoletí paralyzovaly snahy o empirické studium kulturní evoluce.

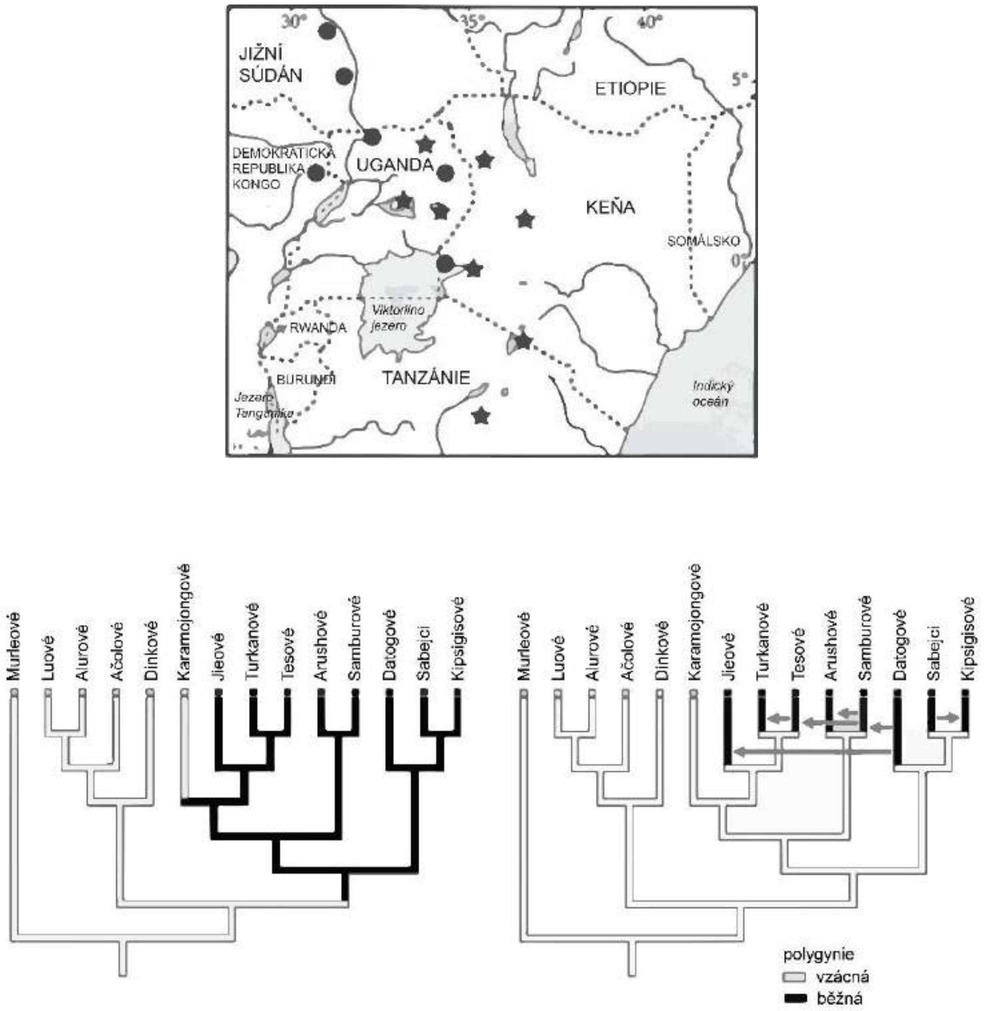
Představa, že kulturní přenos se zásadním způsobem liší od biologické dědičnosti, přetrvává u mnoha biologů i antropologů. Například paleontolog Stephen Jay Gould v jednom ze svých esejů napsal: „*Biologická evoluce je systém neustálého rozdělování bez následného spojování větví. Linie, které jsou jednou rozděleny, zůstanou rozděleny navždy. V lidské historii* *je přenos informace mezi liniemi možná tím nejvýznamnějším zdrojem kulturních změn.*“[88] (Gould 1991, 65).

Naše dnešní znalosti o povaze biologické (a kulturní) evoluce umožňují tuto představu poupravit. Nepatřičně zjednodušený model čistě stromovité (dichotomické) evoluce organismálních linií je zde postaven do kontrastu s komplexním modelem síťovité (retikulátní) evoluce kulturní. Povýšit tento kontrast na dichotomii znamená pokroutit naše znalosti biologické i kulturní evoluce.

K druhotnému splývání dříve oddělených evolučních linií ve skutečnosti často dochází i v evoluci organismů. K horizontálnímu přenosu genetického materiálu dochází prostřednictvím virů a plazmidů. Transpozony, původně parazitické úseky DNA, se mohou přenášet mezi druhy. Geny z mitochondrií a chloroplastů, organel endosymbiotického původu, jsou přenášeny do jádra buňky. Jde o kuriózní případ dávného splynutí nepříbuzných organismálních linií, který dodnes pokračuje vnitrobuněčným přenosem genů. Rozsáhlý horizontální přenos byl zjištěn u vířníků pijavenek (Bdelloidea), v jejichž genomu se nachází DNA bakterií, rostlin i hub. Vedle horizontálního přenosu mezi biologickými liniemi dochází také k jejich splývání, hybridizaci. Řada známých živočišných druhů je hybridního původu. Například zubr evropský (*Bison bonasus*) vznikl křížení pratura (*Bos primigenius*) a bizona stepního (*Bison priscus*). Jelen milu (*Elaphurus davidianus*), vyskytující se v severovýchodní Číně, je hybridem jelena wapiti (*Cervus canadensis*) a jelena lyrorohého (*Panolia eldi*). Vlk rudohnědý (*Canis rufus*), vyskytující se na východním pobřeží Spojených států amerických, je hybridem vlka (*Canis lupus*) a kojota (*Canis latrans*). Několik druhů hybridního původu lze najít mezi primáty. Částečně hybridního původu je i anatomicky moderní člověk. Malé procento genofondu populací moderního člověka, žijících mimo subsaharskou Afriku, pochází od blízkovýchodních populací neandrtálce (*Homo nenaderthalensis*) a také od vědecky dosud nepopsaných asijských příbuzných neandrtálců, nazvaných denisovci podle nálezů z Denisovy jeskyně na Altaji. Geny těchto druhů pronikly do našeho genofondu v důsledku hybridizace s expandující populací anatomicky moderního člověka. Horizontální přenos je tedy poměrně běžný i v biologické evoluci, jeho existence však nemění nic na tom, že biologická evoluce je převážně stromovitá. Podle toho ostatně horizontální přenos vůbec dokážeme rozpoznat.

---

[88] “*Biological evolution is a system of constant divergence without subsequent joining of branches. Lineages, once distinct, are separate forever. In human history, transmission across lineages is, perhaps, the major source of cultural change.*” (Gould 1991, 65)



Obr. 5 Mapa a příbuzenský strom východoafrických populací. Kruhy označují populace, ve kterých je polygynní rodinné uspořádání vzácné (<20 %), hvězdy populace, v nichž je běžné (>20 %). Na příbuzenském stromu je znázorněn hypotetický vznik polygynního rodinného uspořádání jako důsledek vertikálního přenosu (fylogeneze) a horizontálního přenosu (etnogeneze) (upraveno podle Borgerhoff Mulder et al. 2006, 57).

Ačkoli se kulturní znaky mohou šířit horizontálně, podstatná část z nich se přenáší během dětství a dospívání z rodičů na potomky. Některé kulturní znaky mají vysokou dědivost[89], která se vyrovná biologickým znakům (např. výška nebo barva očí nemají o nic vyšší dědivost než jazyk nebo náboženství). U většiny znaků navíc mechanismům jejich dědičnosti nerozumíme (je vyšší rozvodovost v některých rodinách podmíněna biologicky nebo kulturně?) a některé biologické znaky se dědí kulturně (deformace páteře budou častější

---

[89] Dědivost v tomto případě vyjadřuje, do jaké míry je (kulturní) fenotyp potomka shodný s fenotypem jeho biologických rodičů.

v rodinách s genetickými sklony k osteoporóze, ale také v rodinách úředníků). Převážně vertikální přenos a vysoká dědivost kulturních znaků se týká jak postindustriálních společností, tak „přírodních“ národů. Studenti Stanfordovi univerzity získávají své náboženské vyznání a politické přesvědčení převážně od svých rodičů a stejně jako indiáni kmene Tsimane v Bolívii své etnobotanické znalosti (Laland & Brown 2011).

Konflikt o dominantní roli vertikálního nebo horizontálního přenosu není triviální a problémy s ním spjaté nelze jednoduše smést ze stolu odmítnutím jednoho nebo druhého pohledu na věc. Mechanismy kulturního přenosu by měly být zkoumány empiricky na příkladech jednotlivých znaků (Obr. 5). Některé kulturní znaky jsou totiž výsledkem dlouhodobého vertikálního přenosu z generace na generaci, jiné jsou produktem (často obousměrné) kulturní difuze a jiné představují skutečné kulturní inovace.

Fylogenetika disponuje metodami, které umožnují význam fylogeneze a etnogeneze coby procesů generujících kulturní diverzitu explicitně posoudit. Rozsáhlá metaanalýza (Collard et al. 2006) srovnávala „stromovitost“ biologických datasetů (založených na molekulárních, morfologických i etologických datech) a kulturních datasetů (jak kulturních artefaktů, například hrotů šípů či keramiky, tak rozmanitých kulturních praktik). Studie srovnávala hodnoty konzistenčního indexu (CI) a retenčního indexu (RI) pro různé biologické a kulturní datasety. CI a RI jsou kladistické metriky, které udávají, do jaké míry lze podobnosti a rozdíly mezi skupinami či taxony vysvětlit topologií fylogenetického stromu[90]. Z této analýzy vyplývá, že kulturní datasety neobsahují signifikantně více konfliktů s čistě stromovitým modelem evoluce než biologické datasety a že procesy vzniku a vývoje biologické a kulturní diverzity nejsou fundamentálně odlišné. Spíše než a priori stanovit, jak kulturní evoluce probíhala, je potřeba zjistit, který evoluční model, případně kombinace evolučních modelů, hrála v konkrétních případech roli.

Podíl různých procesů, fylogeneze a etnogeneze, se může měnit v čase, v závislosti na socio-politických faktorech, jak ukazuje fylogenetická analýza designu tradičních turkmenských tkanin vyráběných v 18.–20. století. Autorem analýzy je antropolog Jamshid Tehrani z *Durhamské univerzity.* Po bitvě u Geok Tepe v roce 1881 se Turkmeni dostali pod nadvládu carského Ruska, které na jejich území zřídilo Ruský Turkestán. Turkmeni byli

---

[90] Konzistenční index uvádí podíl tzv. homologických znaků, které jsou v souladu s topologií kladogramu a tzv. homoplastických znaků, které jsou s kladogramem v konfliktu. Retenční index uvádí podíl synapomorfií, homologických znaků, lokalizovaných na vnitřních větvích kladogramu, a tedy udržujících jeho stromovitý tvar.

začleněni do struktury koloniálního státu a jejich tradiční klanová sociální organizace se začala rozpadat. Byli také nuceni se usadit a přejít od tradičního pasteveckého způsobu života k zemědělství. Navíc přišli o příjmy plynoucí z pravidelných nájezdů na perské a afghánské vesnice, z prodeje otroků a výkupného. Stali se tak více závislí na místním trhu a důležitým zdrojem příjmů se pro ně stal právě prodej koberců, předložek a dalšího tradičního textilu. Kladistická analýza znaků, popisujících jednotlivé prvky designu tkanin pěti turkmenských klanů, ukazuje, že před rokem 1881 byl vertikální přenos (fylogeneze) dominantním procesem v evoluci designu těchto tkanin. Design tkanin vyrobených po roce 1881 ukazuje na nárůst významu horizontálního přenosu (etnogeneze). Ve chvíli, kdy ornamentální design koberců přestal být vyjádřením příslušnosti k rodinnému klanu a stal se zbožím, přirozeně stouplo množství kulturních výpůjček a s ním i množství konfliktního signálu v datasetech, vyjádřeným hodnotou konzistenčního indexu (0,69 před a 0,61 po roce 1881). Navzdory tomu zůstal vertikální přenos dominantním procesem v evoluci designu turkmenských tkanin (Tehrani & Collard 2002).

Význam obou procesů se nejenže může měnit v průběhu času a v závislosti na místních podmínkách, ale může se také lišit pro různé druhy kulturních znaků. Zdá se rozumné předpokládat, že některé aspekty kultury se budou šířit primárně po příbuzenské linii, zatímco jiné budou závislé na místních podmínkách a jiné se budou šířit horizontálně mezi populacemi. Otázkou, jak různé typy kulturních znaků korelují s jazykovým příbuzenstvím, ekologickými podmínkami a geografií (blízkostí populací), se zabývala studie, vzniklá pod vedením Cavalli-Sforzy (Guglielmino et al. 1995). Tato studie využívala etnografická data popisující 47 kulturních znaků u 277 populací v subsaharské Africe. Výsledky ukázaly, že znaky popisující sociální organizaci (např. rodinné uspořádání, způsob odvozování příbuzenství a příbuzenská terminologie nebo pravidla dědění majetku) korelují s jazykovým příbuzenstvím. Tyto znaky jsou velmi konzervativní, přenášejí se vertikálně a jejich současná distribuce je patrně výsledkem historických migrací. Znaky popisující způsob obživy (např. míra závislosti na lovu a sběru, pastevectví či zemědělství) nepřekvapivě korelují s ekologickými podmínkami. Tyto znaky jsou patrně ekologickými adaptacemi. Některé znaky (např. iniciační rituály nebo náboženství) korelují pouze s geografickou vzdáleností populací a jejich distribuce je nejspíš z velké části výsledkem horizontálního přenosu.

Metody sloužící k rekonstrukci fylogenetických stromů mohou odhalit i to, že fylogenetický strom není vhodným popisem historie zkoumaných kulturních znaků. Ukazuje to studie břidlicových plaket pocházejících z pozdního neolitu a rané doby bronzové (3500–2750 let př. n. l.), nalézaných na jihozápadně Pyrenejského poloostrova (Rivero & O'Brien 2014). Tyto plakety, vysoké 5–10 cm, jsou zdobené geometrickými, antropomorfními nebo zoomorfními motivy. Každá z nich má v sobě vyvrtaný jeden nebo dva otvory, o kterých se předpokládá se, že sloužily k provlečení šňůrky, takže se nosily jako náhrdelníky. Podle populární hypotézy se jedná o heraldické objekty, udávající příslušnost k rodinnému klanu a genealogickou vzdálenost od významného předka. Kladistická analýza motivů na břidlicových plaketách však poskytla strom s velmi nízkým rozlišením, což je v rozporu s předpokladem, že motivy na plaketách značí příslušnost k příbuzenské linii. Optimizace („namapování“) znaků na tento strom odhalila velké množství homoplasií (konvergentních a reverzních změn). To naznačuje, že autoři těchto plaket měli při jejich vytváření značnou tvůrčí svobodu a že motivy na plaketách byly zřejmě často přejímány i tvůrci z nepříbuzných klanů.

Fylogenetické metody určené k vytváření stromů pochopitelně vytvoří strom na základě analýzy libovolného znaku. Publikované stromy jsou však vždy určitým typem konsensu mezi velkým počtem stromů založených na jednotlivých znacích. Pokud je strom pro každý jednotlivý znak v konfliktu s ostatními, bude mít konsensus těchto stromů velmi nízké rozlišení. Takový polytomický strom značí, že evoluce studovaných kulturních fenoménů není stromovitá a ve velké míře se v ní projevuje horizontální přenos[91]. Dnešní metody, vycházející z populační genetiky, umožnují rekonstruovat nejen fylogenetické stromy, ale také fylogenetické sítě, které umožňují odhalit konflikty v kulturních datech (např. metoda *NeighborNet*) nebo horizontální přenos a hybridizaci (např. metoda *Galled networks*) (Willems et al. 2016).

Fylogeneze kultury

Jak už bylo řečeno, nejstarší publikované rodokmeny kulturních artefaktů popisují příbuzenské vztahy rukopisů literárních děl. Tradiční metody, které používali stemmatologové ke stanovení příbuzenských vztahů rukopisů, lze aplikovat pouze na krátké texty. Fylogenetická analýza si poradí s libovolně dlouhými texty, časově náročným úkolem zůstává pouze převod textu do podoby matice znaků. Fylogenetické analýzy se tak dočkaly například *Canterburské povídky* od Geoffreyho Chaucera (Barbrook et al. 1998). *Canterburské povídky* se dochovaly ve zhruba 80 verzích, samotný Chaucerův originál však ne. Fylogenetická analýza 58 dochovaných verzí prologu *Povídky ženy z Bathu* z 15. století umožnila stanovit, které verze jsou nejbližší ztracenému originálu ze 14. století. Také naznačila, že původní Chaucerův

---

[91] Alternativním vysvětlením může být velmi rychlá kladogeneze nebo prostě nedostatek dat.

rukopis byl nejspíš pracovní, opoznámkovanou verzí. Proto se i verze vycházející přímo z něj mezi sebou značně liší jak textem, tak pořadím jednotlivých povídek.

Fylogenetické analýzy se neomezují pouze na psané verze příběhů. I vyprávění (pohádky, pověsti, legendy, mýty) je možné popsat souborem znaků a ty pak podrobit fylogenetické analýze. Za zmínku stojí například fylogenetická analýza pohádky O Červené karkulce, jejímž autorem je již zmíněný Jamshid Tehrani (2013). Nám známá verze této pohádky pochází z roku 1857 a jejími autory jsou Jacob a Wilhelm Grimmové. Autorem nejstarší tištěné verze z roku 1697 je Charles Perrault. Perraultova verze je oproti verzi bratří Grimmů temnější a více moralizující. Historie Červené karkulky však podle všeho sahá do mnohem hlubší minulosti. Různé verze tohoto příběhu nalézáme v ústní tradici celé řady evropských národů a také na některých místech Afriky. Další, odlišnější verze, vycházející z čínské pohádky O tygří babičce, nalézáme ve východní Asii. V Aarne-Thompson-Utherově katalogu lidových vyprávění jsou různé verze příběhu o Červené karkulce uváděny pod zkratkou ATU 333. Některé verze se svým obsahem blíží příběhu o vlku a dětech (ATU 123), rozšířeném po značné části Eurasie a Afriky, o kterém někteří folkloristé předpokládají, že je Červené karkulce vzdáleně příbuzný. Do této skupiny spadá například pohádka O neposlušných kůzlátkách. 58 verzí příběhů ATU 333 a ATU 123 rozšířených v Eurasii a Africe bylo popsáno pomocí matice 72 znaků[92]. Parsimonní a bayesiánská fylogenetická analýza ukázala, že Červená karkulka a Neposlušná kůzlátka skutečně pochází ze společného předka. Příběh o vlku a dětech (ATU 123) je antického původu. Jeho nejstarší psané verze jsou součástí sbírek Ezopových bajek. Červená karkulka (ATU 333) se od této linie příběhů odštěpila ve středověku. Africké verze jsou odvozeny od blízkovýchodních verzí ATU 123 a jejich podobnosti s Červenou karkulkou jsou výsledkem konvergence[93]. Čínské verze pohádky O tygří babičce jsou podle všeho hybridního původu, jak ukázala analýza metodou *NeighborNet*. Vznikly zřejmě smísením prvků některé z evropských verzí ATU 333 a východoasijských verzí ATU 123 v 17. století Analýza tak vyvrátila hypotézu některých folkloristů a sinologů, že Červená karkulka pochází z čínského folkloru.

Podobné fylogenetické analýzy slibují rekonstruovat archetypální verze pohádek a mýtů a rekonstruovat dávné migrace lidské migrace. Distribuce některých kulturních znaků zřejmě

---

[92] Tyto znaky popisují například druhovou příslušnost a pohlaví oběti, druhovou příslušnost útočníka, příbuzenský vztah k osobě, kterou oběť přichází navštívit, důvod návštěvy, místo, kde dojde k útoku, zda je oběť nakonec zachráněna, kdo je jejím zachráncem, jakým způsobem zachránce zabije útočníka atd. Nechybí ani znak „Červený čepeček oběti nepřítomen [0], přítomen [1]“.

[93] Výjimkou je nigerijská verze Červené karkulky v igboštině, která je odvozená přímo od evropských verzí.

reflektuje velmi staré příbuzenské vztahy. Negriti, lovecko-sběračská etnika tmavé pleti a drobného vzrůstu žijící na Andamanských ostrovech, na Malajském poloostrově a na Filipínách, sdílejí víru v boha hromu a soubor tabu, jejichž porušení přivolá jeho hněv[94]. Tyto kulturní shody svědčí o příbuzenském vztahu mezi těmito etniky, jejichž předkové přišli do oblasti jihovýchodní Asie a Indonésie dlouho před příchodem austroasijských a austronéských zemědělců (Blust 2013). Australští domorodci žijící v pobřežních oblastech Austrálie sdílejí mýty o potopě, které zřejmě reflektují zvednutí hladin oceánů na konci pleistocénu a zánik Sahulu (Nunn & Reid 2016). Etnika východní Sibiře a severu Severní Ameriky sdílejí specifickou verzi mýtu, podle které je souhvězdí velké medvědice vyobrazením lovu[95]. Její rozšíření může souviset s jednou z migračních vln z Asie do Severní Ameriky (Berezkin 2005).

Analýzy literárních a folklórních památek nicméně tvoří jen malou část kulturně-fylogenetických studií. Větší část tvoří analýzy materiální kultury, jako jsou výše zmíněné studie designu vlněných tkanin nebo břidlicových plaket. Část studií se zabývá designem kamenných nástrojů, jehož analýzy využívá k rekonstrukci lidských migrací v pleistocénu, nebo k odhalení nezávislých (konvergentních) vzniků obdobných technologických postupů při jejich výrobě. *Jiné studie se zabývají designem keramiky, proutěných košů, výtvarného umění, hudebních nástrojů nebo* architektury *(Lycett 2015).*

Tyto analýzy se neomezují pouze na lidskou kulturu. Antropologové Stephen Lycett a Mark Collard a primatolog William MacGrew provedli kladistickou analýzu datasetu, popisujícího 65 znaků chování u sedmi šimpanzích populací ze západní, střední a východní Afriky (Lycett et al. 2007). Fylogeneze založená na těchto znacích neodráží fylogenezi šimpanzích populací, založenou na genetických datech, což je v rozporu *s představou, že rozdíly v chování těchto populací jsou podmíněné geneticky. Jelikož behaviorální odlišnosti mezi šimpanzími populacemi nejsou závislé na místní ekologii ani na fylogenezi, nezbývá než předpokládat, že tyto rozdíly v chování, zahrnující používání nástrojů, hygienické návyky a sociální normy jsou podmíněny kulturně (viz kapitolu Jitky Lindové). Hodnoty CI a RI pro tento dataset jsou srovnatelné jak s hodnotami pro datasety popisující lidskou kulturu, tak pro mezidruhové datasety založené na* molekulárních, morfologických a etologických datech. *Vnitroskupinový vertikální přenos hrál zřejmě dominantní roli i v evoluci šimpanzí kultury.*

---

[94] Tato bizarní tabu zahrnují např. pálení včelího vosku, hvízdání si za zpěvu cikád, pozorování kopulujících psů, používání hrnce k nabírání vody nebo souložení za bílého dne. Boha hromu je možné si udobřit tím, že se člověk řízne do prstu a mrští několik kapek své krve proti větru ve směru, z kterého přichází bouře.

[95] Jedná se o mýtus, jehož různé verze jsou rozšířeny po celé Eurasii, v Severní a v Jižní Americe. Jednou z jeho verzí je i mýtus o Kallistó.

Biologická evoluce	**Evoluce jazyka**
jednotky dědičnosti (např. nukleotidy, aminokyseliny a geny)	jednotky dědičnosti (např. fonémy, slova a syntax)
přírodní selekce	kulturní selekce a trendy
replikace DNA	sociální učení (např. emulace, imitace a vyučování)
mutace	inovace (např. chyby, hláskové změny a slovní výpůjčky)
homologické znaky	kognáty
drift	drift
anageneze (průběžné změny v rámci evolučních linií)	průběžné změny jazyka (např. vznik češtiny ze staročeštiny)
kladogeneze (speciace)	štěpení jazyků (prostřednictvím migrace, sociální nebo politické separace)
horizontální přenos genetické informace	slovní výpůjčky
hybridní druhy (např. zubr, jelen milu nebo vlk rudohnědý)	kreolštiny (např. haitská kreolština, surinamština nebo unserdeutsch)
klinální variabilita	nářeční kontinua
fosilie	staré texty
extinkce	vymření jazyka

Tabulka 1 Některé paralely mezi biologickou evolucí a evolucí jazyka (upraveno podle Pagel 2009, 406).

Evoluce a fylogeneze jazyka

Většinu kulturně-fylogenetických analýz tvoří analýzy jazyka, který je „modelovým organismem“ kulturní fylogenetiky (Tabulka 1). Jazyk splňuje kritéria selekčně neutrálního znaku. Slova jsou arbitrárně znějící zvuky nesoucí význam[112]. Z toho, jak slova znějí, by neměla plynout žádná selekční výhoda. Jazyky, přinejmenším části slovní zásoby, jsou velmi konzervativní. To platí jak pro slova, tak pro gramatickou strukturu (pořadí větných členů apod.). Jazyk se učíme primárně od rodičů, sourozenců a dalších pokrevních příbuzných. To, že jazyk se přenáší především vertikálně, se v důsledku projevuje tím, že rozšíření příbuzných jazyků zpravidla koreluje s rozšířením určitých genetických variant. Fylogeneze založené na genetických datech a fylogeneze jazyků jsou si často podobné, a to na různých geografických a jazykových škálách. Rozdělení jedné populace na dvě fyzicky oddělené populace povede ke snížení kontaktu mezi jedinci z těchto populací, což přispěje k jejich genetickému i jazykovému odlišení. Geografická vzdálenost dvou populací je dobrým prediktorem intenzity migrace mezi nimi, což se projeví jak na genetické, tak na jazykové diverzitě těchto populací (Creanza et al. 2015).

Tak jako ve fylogenetice rekonstruujeme fylogeneze organismů na základě sdílených homologických znaků, ať už jde o kosti nebo nukleotidy, v historické lingvistice musíme nejprve identifikovat homologická slova. Tato slova, která podobně znějí (fonologická podobnost) a mají podobný význam (sémantická podobnost), se nazývají kognáty[113]. Například slova *bratr* (česky), *brat* (polsky), брат [brat] (rusky), *brother* (anglicky), *Bruder* (německy), *bróðir* (islandsky), *bror* (švédsky), *brālis* (litevsky), *brawd* (velšsky), *frère* (francouzsky), *fratello* (italsky), *frate* (rumunsky), *frāter* (latinsky), φράτηρ [f*rátēr*] (starořecky), *bʰrā́tar* (v sanskrtu) jsou navzájem homologická a všechna pocházejí ze společného předka, kterým je proto-indoevropské **bʰréh₂tēr*[114].

Jestliže jsou kognáty analogií homologických znaků, pak určení, která slova jsou kognáty a která ne, je analogické postupu, který fylogenetici nazývají alignment. To může být

---

[112] To platí do jisté míry a pro většinu sémantických kategorií. Například onomatopoická slova jako *bum* a *kuku* nebo ideofony (expresiva), například *cik cak* nebo *halabala*, náhodně neznějí a lze očekávat, že budou znít podobně i v nepříbuzných jazycích.

[113] Kognát (z latinského *cognatus*, společného původu) je antropologický a právní pojem pro příbuzné v přímé linii, a to po otci i po matce.

[114] Francouzský renesanční učenec Joseph Justus Scaliger již v 16. století klasifikoval evropské jazyky do čtyř skupin na základě způsobu, jakým nazývají boha - *deus*, *god*, *bog* a θεός [*theós*]. Úspěšně tak identifikoval románské, germánské a slovanské jazyky a izolovanou řečtinu.

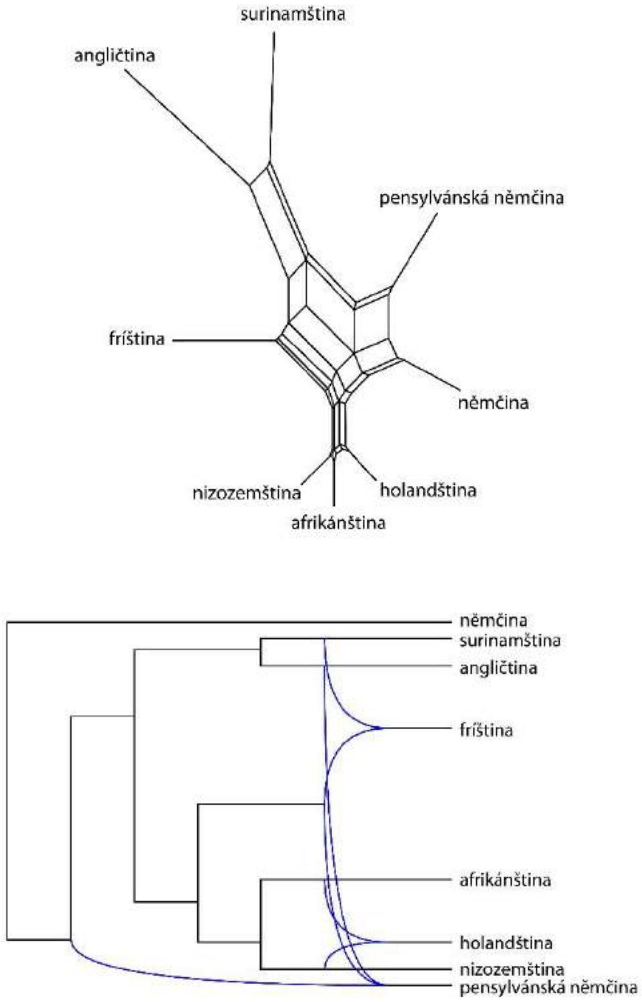
docela oříšek, protože homologie neznamená nutně podobnost (zatímco například českého *bratr* a anglické *brother* jsou zjevně kognáty, u francouzského *frère* už si natolik jistí nejsme). Homologie značí odkaz k původnímu stavu, ke společnému předkovi, ale jak se evoluční změny v průběhu času hromadí, tak se homologické prvky postupně odlišují. České *bratr* je homologické slovenskému *brat* a také arménskému *ełbair*, podobně jako je lidská klíční kost (klavikula) homologická klíční kosti šimpanze a také sáňkám (furcule) holuba. Po určitém čase už kognáty nelze spolehlivě rozeznat a slova ve vzdáleně příbuzných jazycích mohou znít podobně čistě náhodou. Tyto takzvané falešné kognáty jsou analogií konvergencí (homoplasií). Například v arabštině i v mongolštině se bratr řekne *akh*, což ale nevypovídá nic o příbuzenském vztahu těchto jazyků. Kognát může chybět a může být nahrazen nehomologickým slovem. Například španělský výraz pro bratra *hermano* a podobně znějící výrazy v iberorománských jazycích, jako je portugalské *irmão*, nejsou odvozena od latinského *frāter*, ale od latinského adjektiva *germanus*. Kognáty mohou být také odlišná slova jednoho jazyka, například anglická slova *skirt* a *shirt* jsou kognáty. Tyto takzvané dublety jsou analogické homologickým molekulárním sekvencím vzniklým nikoli rozdělením evolučních linií (speciací), ale duplikací uvnitř evoluční linie (takzvané paralogní sekvence). Ve většině jazyků pak nalézáme nemalé množství slovních výpůjček, důsledků horizontálního přenosu[115]. Například maltský výraz pro bratra *brother* je výpůjčka, nikoli doklad o příbuzenském vztahu k angličtině. Maltština není indoevropský, ale afroasijský (semitský) jazyk. V evoluci jazyků nalézáme také analogie mezidruhové hybridizace - kreolské a smíšené jazyky.

Podobných anomálií je v jazyce spousta a každá znamená problém pro fylogenetickou analýzu. Řešení spočívá v analýze většího souboru dat (slov), čímž se jednotlivé anomálie navzájem vyruší. Slova sloužící k rekonstrukci dávných příbuzenských vztahů ovšem nemohou být vybrána náhodně. Mnohá slova se pro tento účel nehodí, protože budou v mnoha jazycích chybět nebo půjde o slovní výpůjčky z jiných jazyků[116]. Jazykovědci proto pracují se souboryslov takzvaného základního slovníku, která se vyvíjejí pomalu, přenášejí se vertikálně

---

[115] Různé jazyky jsou v různé míře inkluzivní. Nejméně jazykových výpůjček (2 %) obsahuje mandarínská čínština, nejvíce (63 %) selická romština.

[116] Angličtina obsahuje nepřeberné množství jazykových výpůjček z rozmanitých jazyků. Tak například slovo *rajče* (*tomato*) pochází z jazyka Aztéků (nahuatl), zatímco slovo *kečup* (*ketchup*) pochází z kantonštiny. Slovo *čaj* (*tea*) pochází z mandarínské čínštiny, *káva* (*cofee*) z arabštiny (přes turečtinu), *kavárna* (*café*) z francouzštiny, cukr (*sugar*) ze sanskrtu (přes perštinu a arabštinu). Slovo *kamarád* (*pal*) je přejaté z romštiny a slovo *brebentit* (*jabber*) z australského jazyka Woiwurrung. Slova *tabu* (*taboo*) a *tetování* (*tatoo*) pochází z austronéských (malajsko-polynéských) jazyků. Výpůjčky z češtiny do angličtiny zahrnují slova *pistole* (*pistol*), polka (*polka*) a *robot* (*robot*).



Obr. 6 Fylogenetická síť a fylogenetický strom germánských jazyků, znázorňující konflikty se stromovitým modelem evoluce a hybridizační události, vytvořené metodami *NeighborNet* a *Galled networks* (upraveno podle Willems et al. 2016, 9).

a málo se vypůjčují mezi nepříbuznými jazyky (na rozdíl od slov jako *banán* nebo *software*). Slova základního slovníku zahrnují tázací zájmena, základní číslovky, základní slovesa, označení částí lidského těla, základní úkony, přírodní substance, živly a další slova, která jsou víceméně nezávislá na místních podmínkách a na technologické vyspělosti příslušného etnika. Nejznámějším takovým souborem slov je takzvaný Swadeshův seznam, pojmenovaný podle Morrise Swadeshe, zakladatele lexikostatistiky. Tři verze tohoto seznamu čítají 200, 100 a 35 slov (*já*, *ty*, *to*, *kdo*, *co*, *jeden*, *dva*, *ryba*, *pes*, *veš*, *krev*, *kost*, *vejce*, *roh*, *ocas*, *ucho*, *oko*, *nos*, *zub*, *jazyk*, *ruka*, *vědět*, *umřít*, *dát*, *slunce*, *měsíc*, *voda*, *sůl*, *kámen*, *vítr*, *oheň*, *rok*, *plný*, *nový*, *slovo*).

Matici znaků, zakládající se zpravidla na přítomnosti/nepřítomnosti kognátu pro dané slovo v daném jazyce (např. *bratr* [0], *brat* [0], *hermano* [1], *irmão* [1]), je pak možné analyzovat některou z fylogenetických metod, která rekonstruuje příbuzenské vztahy mezi jazyky, které jsou minimálně ovlivněné horizontálním přenosem. Například angličtina je pestrou směsicí slov, pocházejících z jazyků Keltů, Římanů, Anglů, Sasů, Jutů a Normanů, s četnými příspěvky z latiny a řečtiny, francouzštiny a italštiny a také z turečtiny, arabštiny, perštiny a hindštiny. Uvádí se, že 99 % anglických slov uvedených v *Oxford English Dictionary* jsou výpůjčky z jiných jazyků, přičemž více než polovina z nich jsou výpůjčky z normanštiny, které do anglického slovníku pronikly po bitvě u Hastingsu v roce 1066. Ve Swadeshově seznamu je nicméně z 200 slov pouhých 12 výpůjček a fylogenetická analýza základního slovníku tak spolehlivě odhalí angličtinu jako severogermánský jazyk příbuzný fríštině. Novější metody rekonstrukce umožnují znázornit také horizontální přenos (Obr. 6).

Fylogenetická analýza lexikálních dat umožňuje rekonstruovat nejen příbuzenské vztahy, ale také délky větví, které odpovídají počtu změn, které se udály v jednotlivých příbuzenských liniích. Po dosazení kalibračních bodů, založených na historických pramenech nebo archeologickém záznamu, je možné odvodit stáří jednotlivých uzlů a kořene fylogenetického stromu (tato metoda nepředpokládá konstantní rychlost lexikální změny jako glottochronologie). Tyto metody pak můžeme využít při studiu historie jednotlivých jazykových skupin.

Rekonstrukce dávných lidských migrací - kulturní fylogeografie

Kde ležela indoevropská pravlast (*Urheimat*) a jak staré jsou indoevropské jazyky, jsou jedny z největších otázek historické lingvistiky. Tradiční, takzvaná kurganská hypotéza litevsko-americké archeoložky Marije Gimbutasové předpokládá, že indoevropská pravlast ležela v oblasti *ponticko-kaspických stepí* na sever od Černého a Kaspického moře (dnešní jihozápadní Rusko a Ukrajina). Mluvčí proto-indoevropského jazyka byli kočovní pastevci skotu, ovcí, koz a koní, kteří přitáhli do Evropy v době bronzové, před 4–6 tisíci lety. Alternativní hypotéza britského archeologa Colina Renfrewa předpokládá, že proto-indoevropané byli zemědělci, kteří do Evropy přišli z oblasti Anatolie (dnešní Turecko), a to podstatně dříve, před 8–10 tisíci lety.

První bayesianská fylogenetická analýza indoevropských jazyků stanovila stáří indoevropské jazykové rodiny na 7 800–9 800 let (Gray & Atkinson 2003). Navazující studie určila jako zdroj expanze Anatolii (Bouckaert et al. 2012). Novější analýza téhož datasetu, založená na odlišné metodě časové kalibrace fylogenetického stromu stanovila stáří společného předka indoevropských jazyků na šest tisíc let[117] (Chang et al. 2015). Stepní hypotéza byla podpořena také rozsáhlou archeogenetickou analýzou, která odhalila, že lidé kultury jamnaja, žijící v klíčové oblasti ponticko-kaspických stepí, expandovali před zhruba 4 500 lety do Evropy, kde významnou měrou přispěli ke genetickému profilu lidí Kultury šňůrové keramiky. Tato kultura se následně rozšířila přes značnou část střední a východní Evropy. Autoři studie spekulují, že lidé kultury jamnaja by mohli být oněmi záhadnými mluvčími proto-indoevropského jazyka (Haak et al. 2015).

Rekonstrukce fylogenetického stromu jazyků umožnuje také rekonstrukci slovníku společného předka těchto jazyků (proto-jazyka). Jde o mapování změn jednotlivých slov na topologii stromu od koncových větví ke společným předkům. Tento postup je analogický tomu, co evoluční biologové nazývají fylogenetická rekonstrukce ancestrálních znakových stavů. Zatímco biologové mohou rekonstruované morfologické znaky porovnat s fosilním záznamem, jazykovědci, kteří rekonstruují lexikální znaky, porovnávají své závěry s archeologickým záznamem, historickými prameny a starými texty.

Rekonstruovaný proto-jazyk nám umožňuje nahlédnout do způsobu života, sociálního uspořádání a prostředí v jakém žili jeho mluvčí a v kombinaci s fylogenetickým stromem nám může pomoci lokalizovat pravlast příslušné jazykové rodiny. Například proto-indoevropský jazyk obsahoval slova jako *kůň* (**éḱwos*), dobytek *(*g^{w}ōus*), *jařmo (*yugóm*), *kolo (*k^{w}ek^{w}lo*-) a „*cestovat ve voze*“ *(*weǵh*-), což dále podporuje hypotézu o stepním původu indoevropských jazyků. Rekonstrukce, prováděné autoritami v oboru historické lingvistiky, se tradičně neobešly bez jisté dávky intuice a spekulací. Evoluční biolog Ward Wheeler z Amerického muzea přírodní historie, vyvinul k rekonstrukci proto-jazyka algoritmus, využívaný v molekulární fylogenetice k rekonstrukci ancestrálních sekvencí DNA (slova, stejně jako geny, koneckonců také zapisujeme pomocí sekvencí písmen). Tento algoritmus pak uplatnil při rekonstrukci evoluce a fylogeneze juto-*aztéckých jazyků* (Wheeler & Whiteley 2015).

Jak hluboko do minulosti je možné vztahy mezi jazyky rekonstruovat? Většina jazykovědců je přesvědčena, že vzhledem k rychlosti, jakou se jazyk vyvíjí, přestávají být po

---

[117] Tato analýza se od obou předchozích lišila jedním podstatným detailem. Středověké a starověké jazyky, považované za předky dnešních skupin indoevropských jazyků (např. latina jako předek románských jazyků a sanskrt jako předek indoíránských jazyků) vystupovaly v této analýze nikoli na pozici příbuzných, jako koncové větve, ale na skutečně pozici předků, tj. jako vnitřní větve fylogenetického stromu.

zhruba osmi tisících letech kognáty rozeznatelné a všechny zdánlivé fonologické a sémantické podobnosti mezi slovy nejsou ničím jiným než dílem náhody (jde o falešné kognáty). Z toho vyplývá, že na základě lexikálních dat není možné rekonstruovat fylogenetické vztahy starší než zhruba osm tisíc let. Fylogenetik Mark Pagel, působící na Univerzitě v Readingu, studoval rychlost, jakou se vyvíjejí slova v indoevropských jazycích (Pagel et al. 2007). Mezi jednotlivými slovy existují velké rozdíly v „mutační“ rychlosti a to i v rámci základního slovníku (Swadeshův seznam o 200 slovech). Zatímco pro slovo *bratr* existují v rámci indoevropských jazyků dva kognáty, pro slovo *špinavý* jich existuje 46. „Mutační rychlost“ slov lze vyjádřit „poločasem rozpadu“ slova, což je doba, po jejímž uplynutí existuje 50 % šance, že kognát bude nahrazen ne-kognátem (ať už skrze mutaci nebo slovní výpůjčku). S využitím fylogeneze indoevropských jazyků dospěl Pagel a jeho tým k závěru, že průměrný (mediánový) poločas rozpadu slova v indoevropském jazyce je 2 000–2 500 let. To se může zdát jako poměrně krátká doba, nicméně je to stále víc, než je průměrná délka existence indoevropského jazyka (500–1 000 let). Homologická slova tedy přežívají štěpení jazyků, podobně jako homologické geny přežívají speciaci druhů. Přitom existují i mimořádně konzervativní slova, k nimž patří zájmena *já* a *kdo* nebo základní číslovky *dva, tři* a *pět*, která mají napříč indoevropskými jazyky jen jeden kognát, a jejich poločas rozpadu dalece přesahuje 10 tisíc let. Nejkonzervativnější slova jsou ta, která používáme v každodenní mluvě s nejvyšší frekvencí. Sem patří tázací zájmena (*co, kdy, kde, jak*), ukazovací zájmena (*tady, tam*), záporná zájmena (*nic*) a základní číslovky (*jedna, dvě*), tedy obecně slova, která slouží hlavně k informování se o sociálních vztazích. Vyjadřujeme jimi, *kdo, kdy, kde* a *co komu* udělal, případně *jak často* a *kolikrát*. Frekvence používání slov v současných jazycích tak predikuje rychlost jejich evoluce v průběhu tisíců let. Tato konzervativní slova je možné využít k rekonstrukci dávných příbuzenských vztahů mezi jazykovými rodinami. Další možnost posunutí hranice rekonstruovatelných vztahů hlouběji do minulosti nabízí využití strukturních (gramatických a syntaktických) znaků, jako je ohýbání slov nebo pořadí větných členů, které se vyvíjejí pomaleji a jsou méně náchylné k horizontálnímu přenosu než lexikální znaky (Dunn et al. 2005). Tyto přístupy tak otevírají možnost rekonstrukce velmi starých příbuzenských vztahů mezi jazykovými rodinami a izolovanými jazyky.

Většina dosavadních fylogenetických analýz jazyka se nicméně týká jednotlivých jazykových rodin, jako je rodina indoevropská, austronéská nebo bantuská, které jsou holocenního stáří a jejichž expanze byla podnícena nějakou kulturní inovací, nejčastěji vynálezem zemědělství.

Jestliže je jazyk modelovým organismem kulturní fylogenetiky, pak Oceánie je její přírodní laboratoří. Osídlení této oblasti světa, zejména Vzdálené Oceánie[118], je relativně mladé a ostrovní kultury představují poměrně dobře vymezené jednotky. Množství fylogenetických analýz jazyka i jiné, materiální i nemateriální kultury, se proto týká právě této oblasti světa.

Otázka, jak lidé dokázali kolonizovat odlehlé tichomořské ostrovy, fascinovala Evropany po staletí. Na těchto ostrovech se mluví polynéskými jazyky, které jsou součástí Austronéské jazykové rodiny. Tato jazyková rodina zahrnuje více než 1 200 jazyků, rozšířených v oblasti Indického a Tichého oceánu o rozloze přes 26 000 km^2. Většina učenců, počínaje Jamesem Cookem, předpokládala, že Oceánie byla kolonizována od západu, odněkud z jižní nebo jihovýchodní Asie. V 50. letech dvacátého století zájem o toto téma vyvrcholil díky úsilí Thora Heyerdahla, který podnikl v roce 1947 na balzovém voru Kon-Tiki plavbu od pobřeží Peru k souostroví Tuamotu ve Francouzské Polynésii ve snaze dokázat, že Oceánie byla kolonizována od východu, z Jižní Ameriky. Většina antropologů, archeologů i jazykovědců však zůstávala přesvědčena o asijském původu Austronésanů.

Na přelomu století se proti sobě stály dvě hypotézy. První z těchto hypotéz byla v 80. letech zpopularizována biologem a antropologem Jaredem Diamondem pod názvem *Express Train to Polynesia* (Diamond 1988) a později dále rozvedena archeology a lingvisty. Tato hypotéza předpokládá, že austronéské jazyky mají původ na Taiwanu, odkud se před zhruba pěti tisíci lety začaly šířit do Indonésie a Oceánie. Jejich expanze byla spojena s šířením tzv. lapitské kultury spojené se zemědělstvím a chovem domácích zvířat, pokročilými rybářskými a mořeplaveckými technikami a charakterizované obsidiánovými nástroji a typickou keramikou. Během rychle postupující expanze přes Filipíny, Indonésii a Melanésii se mluvčí austronéských jazyků minimálně mísili s původními obyvateli, příbuznými Papuáncům, kteří tyto oblasti světa kolonizovali dlouho před nimi. K mísení genů a kultury mělo dojít až v následujících obdobích. Druhá hypotéza, která byla zpopularizovaná genetiky Stephenem Oppenheimerem a Martinem Richardsem pod názvem *Slow Boat* model (Oppenheimer & Richards 2001), předpokládá starší, pleistocénní původ austronéských jazyků v oblasti Indonésie na východ od Wallaceovy linie (Sulawesi, Moluky a Malé Sundy) před zhruba 15 tisíci lety a jejich postupné šíření ve všech směrech, spojený s intenzivním genetickým a kulturním mísením. Pozdější verze této hypotézy se z velké části zakládají na genetice,

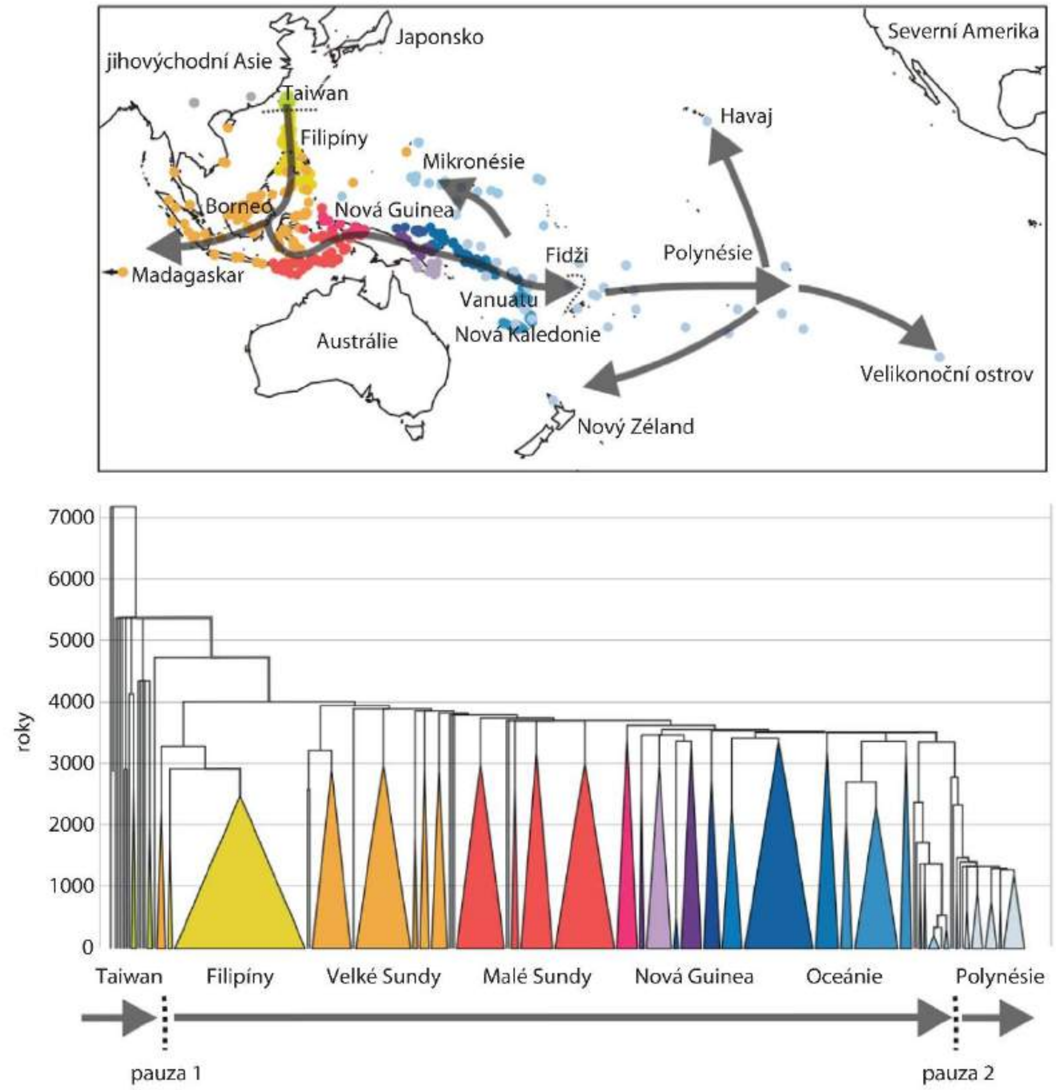
---

[118] Oblast Tichomoří na východ od Šalamounových ostrovů.

především analýzách mitochondriální DNA a chromozomu Y, které dokládají, že obyvatelé západní Polynésie jsou směsí východoasijských a papuánských genetických komponent.

Fylogenetická analýza je mimořádně vhodným nástrojem pro testování těchto hypotéz. První hypotéza implikuje topologii stromu s bazální pozicí taiwanských (formoských) jazyků a hřebínkovitou strukturu, odrážející postupnou kolonizace ostrovů, v souladu s postupem expanze navrženým hypotézou *Express Train*. Stáří kořene tohoto stromu a délka větví by navíc měli odrážet holocénní stáří jazykové rodiny a rychlost expanze. Druhá hypotéza implikuje strom, zakořeněný v oblasti jihovýchodní Indonésie, sestávající ze dvou kladů - východního a západního. Tento strom by měl mít nižší rozlišení, vzhledem k předpokládanému pleistocennímu stáří jazykové rodiny a intenzivnímu horizontálnímu přenosu v průběhu postupného šíření austronéských jazyků.

První fylogenetická analýza, založená na lexikálních datech, jejímiž autory jsou antropologové Russel Gray a Fiona Jordanová z Aucklandské univerzity, podpořila hypotézu o rychlé expanzi austronéských jazyků z Taiwanu (Gray & Jordan 2000). Navazující analýza stanovila stáří austronéské jazykové rodiny na zhruba pět tisíc let a rekonstruovala chronologii expanze (Obr. 7). Délka větví umožnila stanovit, jak se rychlost expanze měnila v průběhu času (Gray et al. 2009). Fylogenetické analýzy jazyků byly později doplněny o komplementární analýzy kulturních artefaktů. Analýza designu lapitské keramiky z oblasti východní Melanésie odhalila množství konfliktů se stromovitým modelem evoluce, svědčícím o intenzivním kulturním mísení v oblasti (Cochrane & Lipo 2010). Naopak analýza designu polynéských kánoí určila jako zdroj expanze do Polynésie Fidži a rekonstruovala postup expanze, který je v souladu s archeologickými daty (Rogers et al. 2009). Analýzy genetických dat svědčily spíše pro hypotézu *Slow Boat*, tedy pleistocénní původ austronéských jazyků se zdrojem v Indonésii a jejich pomalejší šíření zahrnující intenzivní mísení. Obyvatelé této oblasti jsou totiž výsledkem dlouhodobého mísení původních obyvatel Indonésie a Melanésie s nově příchozími Austronésany. Potvrzení hypotézy *Express Train* přinesla až archeogenetická studie, která prokázala, že lidé kultury lapita, žijící před 2–3 tisíci lety na ostrovech Vanuatu a Tonga, byli geneticky příbuzní dnešním Taiwancům, ale nikoli Papuáncům, a k mísení tedy došlo dodatečně (Skoglund et al. 2016).



Obr. 7 Mapa a fylogenetický strom 400 austronéských jazyků, zachycující průběh austronéské expanze. Na mapě a na vodorovné škále jsou vyznačeny dvě delší pauzy v průběhu jinak plynule postupující expanze. První před kolonizací Filipín (oddělení malajsko-polynéských jazyků od východoformoských jazyků, jedné z devíti skupin austronéských jazyků na Taiwanu), druhá před kolonizací Polynésie (oddělení polynéských jazyků od jazyků souostroví Fidži) (upraveno podle Gray et al. 2009, 480).

Archeologie ve spojení s historickou lingvistikou, kulturní fylogenetikou a populační genetikou tak nakonec poskytly ucelený obraz kolonizace Oceánie mluvčími austronéských jazyků. Zdrojem austronéské expanze byl před zhruba pěti tisíci lety Taiwan. Austronéské jazyky se dělí na deset skupin, z nichž devět (zahrnujících 20 jazyků) se vyskytuje pouze na Taiwanu, zatímco desátá, malajsko-polynéská (zahrnující 1 237 jazyků) je rozšířena od Filipín po Madagaskar na západě a Velikonoční ostrov a Havaj na východě[119]. Expanze byla zřejmě podnícena vynálezem vahadlové kánoe, který umožnil překonání 350 km širokého *Luzonského průlivu a kolonizaci Filipín*. Tuto teorii podporuje mimo jiné přítomnost terminologie spojené s vahadlovými kánoemi v proto-malajsko-polynéském, ale nikoli proto-austronéském slovníku. Z Filipín se mluvčí austronéských jazyků šířili na jihovýchod, do oblasti Velkých Sund a do Indočíny a eventuálně také na Madagaskar[120]. Expanze zároveň postupovala na východ přes Malé sundy a podél severního pobřeží Nové Guineje do Melanésie a Vzdálené Oceánie, aniž by se významně mísili s původními obyvateli Nové Guineje a okolních souostroví. Souostroví Fidži pak sloužilo jako odrazový můstek pro kolonizaci vzdálených ostrovů v Polynésii, až nakonec v 10.–13. století n. l. byly kolonizovány souostroví Havaj, Velikonoční ostrov a Nový Zéland[121].

Další provedené fylogenetické analýzy jazyka zahrnují jazyky bantuské, semitské, turkické, severokavkazské, uralské, pama-nyunganské, mon-khmerské, japonsko-rjúkjúské, dene-jenisejské, juto-aztécké, aravacké, jazyky tupí-guaraní a další. Nadto se objevují první formální fylogenetické analýzy zahrnující více jazykových rodin, které dávají naději, že jednou budeme schopní rekonstruovat globální fylogenezi všech sedmi tisíc světových jazyků (Jäger 2015).

Fylogenetické komparativní analýzy kulturních fenoménů

V roce 1888 měl Edward Burnett Tylor přednášku na Královském antropologickém Institutu v Londýně. Tylor, který shromáždil etnografická data o 350 lidských společnostech, zjistil, že společnosti s méně komplexním společenským uspořádáním (např. lovecko-sběračské společnosti s kmenovým uspořádáním) jsou zpravidla matriarchální, zatímco společnosti

---

[119] Je to jako kdyby se v oblasti západní Eurasie sahající od Velké Británie po Srí Lanku a Asám hovořilo pouze slovanskými jazyky, zatímco všechny ostatní indoevropské jazyky, včetně románských, germánských, keltských a dalších se vyskytovaly pouze v Irsku.

[120] Malgaština je austronéský jazyk, blízce příbuzný bornejskému jazyku Ma'anyan. Prvky materiální kultury spojující Malgaše s obyvateli Indonésie zahrnují vahadlové čluny, hudební nástroje (např. xylofon), metalurgické techniky či pěstování původem asijských zemědělských plodin (rýže, banány, yamy a taro). Genetické analýzy svědčí o osídlení Madagaskaru skupinou čítající zřejmě pouhých několik desítek lidí indonéského původu v 9. stol. n. l.

[121] Austronéští mořeplavci, kteří ve 13. stol. n. l. kolonizovali Velikonoční ostrov, nakonec dosáhli také západního pobřeží Jižní Ameriky. O předkolumbovských kontaktech mezi obyvateli Polynésie a Jižní Ameriky svědčí genetické i kulturní důkazy, mimo jiné přítomnost původem jihoamerické plodiny povijnice batátové (*Ipomoea batatas*) v Oceánii, archeologické nálezy slepičích kostí v Chile datované do 14. stol. n. l. a v neposlední řadě jihoamerická genetická komponenta v genomech obyvatel Velikonočního ostrova. Oceánie sice byla kolonizována od západu, ale Thor Heyerdahl měl pravdu v tom, že technologie, kterou austronéští mořeplavci disponovali, umožňovala plavbu z Jižní Ameriky do Polynésie.

s komplexnějším společenským uspořádáním (např. zemědělské společnosti se státním zřízením) jsou zpravidla patriarchální. Z toho usoudil, že evoluce lidského rodinného uspořádání spěje od matriarchálního k patriarchálnímu. V publiku přítomný Francis Galton namítl, že z korelace mezi společenským a rodinným uspořádáním nelze vyvozovat nic o evoluční sekvenci. Je nutné vzít v úvahu příbuzenské vztahy mezi společnostmi i možné kulturní výpůjčky. Galtonova kritika se v budoucnu stala známá pod názvem *Galtonův problém* nebo také problém *statistické autokorelace*. Lidské společnosti nejsou navzájem nezávislými jednotkami statistické analýzy jak díky svým příbuzenským vztahům, tak díky své geografické blízkosti, která usnadňuje horizontální přenos. Tento problém, společný antropologii i evoluční biologii, si antropologové uvědomili jako první a v průběhu dvacátého století se s ním snažili vypořádat různými způsoby, zpravidla vytvářením datasetů pro komparativní analýzy, sestávajících z „nepříbuzných“ společností (např. *Standardní kross-kulturní vzorek*). Takové metody mohou Galtonův problém umenšit, ale nikoli eliminovat. V 80. letech vyvinuli evoluční biologové statistické metody umožňující měřit znakové korelace se zohledněním fylogenetických vztahů a také rekonstruovat evoluční historii. Od 90. let začali tyto metody využívat i sami antropologové. Zakládají se na kombinaci etnografického záznamu a fylogenetického stromu lidských etnik, založeného na analýze genetických nebo lingvistických dat, a umožňují nejen testovat evoluční sekvence navržené antropology, ale také rekonstruovat aspekty života dávných společností, které nezanechávají stopy v antropologickém záznamu.

Studie antropologů Ruth Maceové a Toma Currieho z University College London mapovala vzestupy a pády socio-politické organizace v průběhu kolonizace Oceánie mluvčími austronéských jazyků (Currie & Mace 2011). Austronéská expanze, na jejímž počátku stáli taiwanští zemědělci, dala postupně vzniknout vyspělým náčelnictvím i několika státním útvarům, ovládajícím rozsáhlá území. Vyspělá náčelnictví vznikla například ve vzdálené Oceánii na souostrovích Samoa a Tonga, státní útvary vznikly na Sumatře (Šrívidžaja) a na Jávě (Madžapahit) a později také na Madagaskaru (Merinská říše) a na Havaji (Havajského království). Některá etnika, například Maorové na novém Zélandu, naopak dodatečně přešla k lovecko-sběračskému způsobu života a k jednoduchému náčelnictví. V některých případech, například na Gambierových ostrovech nebo na Velikonočním ostrově, došlo k rapidnímu zhroucení socio-politické komplexity (Diamond 2008). Oceánie tak představuje unikátní přírodní laboratoř pro testování navržených modelů evoluce kultury. Currie a Maceová využili fylogenezi austronéských jazyků v kombinaci s etnografickým záznamem, proti kterým testovali několik modelů evoluce socio-politické organizace. Model, který nejlépe odpovídá distribuci znakových stavů na fylogenezi, předpokládá postupný nárůst socio-politické komplexity (od kmenového uspořádání přes jednoduché a vyspělé náčelnictví po státní zřízení) a její postupný, případně rapidní pokles. Studie také zjistila, že ancestrální austronéská společnost byla kmenová, bez jurisdikční hierarchie a že vznik státního uspořádání je podmíněn přítomností sociální stratifikace. Tyto závěry jsou v pozoruhodném souladu s předpoklady evolucionistických antropologů Herberta Spencera a Edwarda Burnetta Tylora. Související studie, jejímiž autorkami jsou Fiona Jordanová a Ruth Maceová, zjistila, že ancestrální austronéská společnost byla matrilokální, přičemž ke změnám sociálního uspořádání od matrilokálního na patrilokálnímu došlo nezávisle na Filipínách, na Velkých Sundách a v Mikronésii (Jordan et al. 2009).

Obdobné metody byly uplatněny při studiu kulturní historie mluvčích indoevropských jazyků. Dnes tak víme nejen to, kde a kdy indoevropské jazyky vznikly, ale také, že jejich mluvčí praktikovali monogamní sňatky, platili cenu za nevěstu a jejich sociální organizace byla patrilokální. Nejenže dokážeme rekonstruovat proto-indoevropský jazyk, ale jsme dokonce schopní rekonstruovat příběhy, které si jeho mluvčí vyprávěli.

Někteří folkloristé 19. století, například Wilhelm Grimm[122], byli přesvědčeni, že evropské pohádky mají základ v ústní tradici sahající hluboko do minulosti, zatímco jiní předpokládali, že většina rozšířených pohádek je produktem novověké literární tradice. Řešení tohoto sporu, který se táhl po celé 20. století až do současnosti, naráželo na dva problémy. Je těžké určit, jak staré jsou mezinárodně rozšířené pohádky, protože jejich nejstarší psané verze pocházejí z 16. století nebo z pozdější doby a protože pohádky se, zejména v posledních staletích, rozšířily díky překladům. Tyto problémy, absenci fosilního záznamu a horizontální přenos, řečeno evolučně-biologickými pojmy, lze elegantně řešit pomocí evolučně-biologických metod.

Studie vzniklá pod vedení Jamshida Tehraniho se zabývala tím, jak hluboko do minulosti sahá historie evropských pohádek (da Silva & Tehrani 2016), konkrétně příběhů, které jsou v Aarne-Thompson-Utherově katalogu uvedeny pod označením ATU 300–ATU 749. Tyto příběhy zahrnují některé nejznámějších pohádek, například Jeníček a Mařenka (ATU 327A) nebo Kráska a zvíře (ATU 425C). Pomocí dvou testů horizontálního přenosu byly ze souboru vyloučeny pohádky, jejichž distribuce na fylogenetickém stromu je výsledkem horizontálního

---

[122] *“The outermost lines [of common heritage in stories] . . . are coterminous with those of the great race which is commonly called Indo-Germanic, and the relationship draws itself in constantly narrowing circles round the settlements of the Germans . . . It is my belief that the German stories do not belong to the northern and southern parts of our fatherland alone but that they are the absolutely common property of the nearly related Dutch, English and Scandinavians” (Grimm, 1884, p. 576)*

přenosu mezi nepříbuznými jazyky, i ty, jejichž geografická distribuce je výsledkem horizontálního přenosu mezi sousedními populacemi. Takto byli ze souboru vyloučeni například Jeníček a Mařenka, ale ne Kráska a zvíře. Dataset popisující přítomnost či nepřítomnost jednotlivých příběhů byl poté optimizován na topologii fylogenetického stromu indoevropských jazyků. Čtrnáct ze 76 příběhů, včetně Krásky a zvířete, sahá až ke společnému předkovi západní větve indoevropských jazyků (tj. balto-slovanských, germánských, keltských a románských) a čtyři z nich sahají až ke společnému předkovi všech indoevropských jazyků. Příběh, jehož přítomnost u společného předka byla rekonstruována s pravděpodobností 87 % je Kovář a ďábel (ATU 330)[123]. V tomto příběhu kovář upíše svou duši ďáblu (v některých verzích příběhu smrti nebo džinovi), aby se stal mistrem ve svém řemesle. Díky takto získané dovednosti kovář přikove ďábla k lavici nebo ke stromu a vyváže se z dohody. Jednou z variant tohoto příběhu je pohádka *Dobře tak, že je smrt na světě* od Karla Jaromíra Erbena, jinou je *Dařbuján a Pandrhola* od Jana Drdy. Výsledek této studie nám dává nahlédnout do technologických znalostí proto-indoevropanů a mimo jiné dále podporuje kurganskou hypotézu, která předpokládá původ indoevropských jazyků v době bronzové.

Většina dosavadních kulturně-fylogenetických studií studovala průběh kulturní evoluce v rámci jednotlivých jazykových expanzí. Existují však výjimky, jako například studie náboženství, jejímiž autory jsou antropologové Hervey Peoplesová a Frank Marlowe z Univerzity v Cambridgi (Peoples et al. 2016). Peoplesová a Marlowe zkoumali, jak hluboko do minulosti sahají fundamentální prvky náboženství, jako například víra v posmrtný život nebo šamanismus. Využili k tomu globální fylogenezi lovecko-sběračských etnik, založenou na kombinaci genetických a lingvistických dat a etnografická data ze *Standardního kross-kulturního vzorku*. Z fylogenetické analýzy vyplývá, že nejstarším prvkem náboženství, sahajícím až ke společnému předkovi lovecko-sběračských etnik, je animismus, v souladu s předpoklady Edwarda Burnetta Tylora o fundamentální roli tohoto fenoménu (viz kapitolu Lindy Hroníkové). Nepotvrdily se však Tylorovi a Spencerovi spekulace o uctívání předků coby základu veškerého náboženství. Studie odhalila, že jednotlivé prvky náboženství spolu korelují (změna jednoho znaku na fylogenetickém stromu ovlivňuje pravděpodobnost změny druhého). Jeden ze studovaných znaků, víra v boha, nicméně nekoreluje s žádným jiným prvkem náboženství, ani s animismem. To podporuje předchozí studie, které ukazují, že víra

---

[123] Další příběhy, sahající k nejstarším předkům zahrnují právě Krásku a zvíře (ATU 425C), Tři pírka a Kocoura v botách (ATU 402), Bílého hada a Včelí královnu (ATU 673) nebo Kmotřičku Smrt (ATU 332).

v boha, ovlivňujícího pozemské dění a trestajícího morální přestupky, souvisí spíše se způsobem života než s ostatními prvky náboženství.

Síla a příslib kulturní fylogenetiky tkví v rigorózním přístupu a v neotřelé kombinaci dat a metod z oborů, které spolu v minulosti nedostatečně komunikovaly, přestože si často kladly podobné typy otázek a přitom čelily podobným typům metodologických problémům. Znalost příbuzenských vztahů a metod, sloužících ke studiu biologické evoluce, umožňuje využít bohatství dat, nashromážděných generacemi antropologů a etnografů počínaje Spencerem a Tylorem, k novým účelům. Mimo jiné k demonstraci toho, že biologické a evoluční teorie kultury přinášejí cenné poznatky, ke kterým bychom se nedobrali, kdybychom nevykročili za hranice jednotlivých oborů.

Literatura

Atkinson QD, Gray RD (2005) Curious parallels and curious connections - Phylogenetic thinking in biology and historical linguistics. *Syst Biol, 54*(4), 513-526. doi:10.1080/10635150590950317

Barbrook AC, Howe CJ, Blake N, Robinson P (1998) The phylogeny of The Canterbury Tales. *Nature, 394*(6696), 839-839. doi:10.1038/29667

Berezkin Y (2005) Cosmic Hunt: Variants of Siberian-North American Myth. *Folklore: Electronic Journal of Folklore*(31), 79-100.

Blust R (2013) Terror from the Sky: Unconventional Linguistic Clues to the Negrito Past. *Human Biology, 85*(1-3), 401-416.

Borgerhoff Mulder M, Nunn CL, Towner MC (2006) Cultural macroevolution and the transmission of traits. *Evolutionary Anthropology, 15*(2), 52-64. doi:10.1002/evan.20088

Bouckaert R, Lemey P, Dunn M, Greenhill SJ, Alekseyenko AV, Drummond AJ, . . . Atkinson QD (2012) Mapping the origins and expansion of the Indo-European language family. *Science, 337*(6097), 957-960. doi:10.1126/science.1219669

Boyd R, Richerson PJ (2012) *V genech není všechno*. Praha: Academia.

Cannon WF (1961) The impact of uniformitarianism: two letters from John Herschel to Charles Lyell, 1836-1837. *Proceedings of the American Philosophical Society, 105*(3), 301-314.

Cochrane EE, Lipo CP (2010) Phylogenetic analyses of Lapita decoration do not support branching evolution or regional population structure during colonization of Remote Oceania. *Philosophical Transactions of the Royal Society B-Biological Sciences, 365*(1559), 3889-3902. doi:10.1098/rstb.2010.0091

Collard M, Shennan SJ, Tehrani JJ (2006) Branching, blending, and the evolution of cultural similarities and differences among human populations. *Evolution and Human Behavior, 27*(3), 169-184. doi:10.1016/j.evolhumbehav.2005.07.003

Creanza N, Ruhlen M, Pemberton TJ, Rosenberg NA, Feldman MW, Ramachandran S (2015) A comparison of worldwide phonemic and genetic variation in human populations. *Proceedings of the National Academy of Sciences of the United States of America, 112*(5), 1265-1272. doi:10.1073/pnas.1424033112

Currie TE, Mace R (2011) Mode and tempo in the evolution of socio-political organization: reconciling 'Darwinian' and 'Spencerian' evolutionary approaches in anthropology. *Philosophical Transactions of the Royal Society B-Biological Sciences, 366*(1567), 1108-1117. doi:10.1098/rstb.2010.0318

Čelakovský FL (1853) *Čtení o srovnávací mluvnici slovanské na universitě pražské. Spisů musejních čísli XLVII.* Praha: V komisi u Františka Řivnáče.

da Silva SG, Tehrani JJ (2016) Comparative phylogenetic analyses uncover the ancient roots of Indo-European folktales. *Royal Society Open Science, 3*(1), 11. doi:10.1098/rsos.150645

Darwin C (1859) *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life* (1st ed.). London: John Murray.

Darwin C (1871) *The Descent of Man, and Selection in Relation to Sex* (1st ed.). London: John Murray.

Darwin C (2006) *O původu člověka*. Praha: Academia.

Darwin Correspondence Project, "Letter no. 346," accessed on 30 November 2016, http://www.darwinproject.ac.uk/DCP-LETT-346

Darwin C (2007) *O vzniku druhů přírodním výběrem*. Praha, Academia.

Dennett DC (1996) *Darwin's dangerous idea: Evolution and the meanings of life*. New York: Simon & Schuster.

Diamond JM (1988) Express train to Polynesia. *Nature, 336*(6197), 307-308.

Diamond JM (2008) *Kolaps: Proč společnosti zanikají a přežívají*. Praha: Academia.

Dunn M, Terrill A, Reesink G, Foley RA, Levinson SC (2005) Structural phylogenetics and the reconstruction of ancient language history. *Science, 309*(5743), 2072-2075. doi:10.1126/science.1114615

Gray RD, Atkinson QD (2003) Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature, 426*(6965), 435-439. doi:10.1038/nature02029

Gray RD, Jordan FM (2000) Language trees support the express-train sequence of Austronesian expansion. *Nature, 405*(6790), 1052-1055. doi:10.1038/35016575

Gray RD, Drummond AJ, Greenhill SJ (2009) Language phylogenies reveal expansion pulses and pauses in Pacific settlement. *Science, 323*(5913), 479-483. doi:10.1126/science.1166858

Grimm W (1884) *Children's and Household Tales* (3rd ed.). London: George Bell.

Guglielmino CR, Viganotti C, Hewlett B, Cavalli-Sforza, LL (1995) Cultural variation in Africa: Role of mechanisms of transmission and adaptation. *Proceedings of the National Academy of Sciences of the United States of America, 92*(16), 7585-7589. doi:10.1073/pnas.92.16.7585

Haak W, Lazaridis I, Patterson N, Rohland N, Mallick S, Llamas B, . . . Reich D (2015) Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature, 522*(7555), 207-+. doi:10.1038/nature14317

Henrich J (2004) Demography and cultural evolution: How adaptive cultural processes can produce maladaptive losses - The Tasmanian case. *American Antiquity, 69*(2), 197-214. doi:10.2307/4128416

Chang W, Cathcart C, Hall D, Garrett A (2015) Ancestry-constrained phylogenetic analysis supports the Indo-European steppe hypothesis. *Language, 91*(1), 194-244.

Jäger G (2015) Support for linguistic macrofamilies from weighted sequence alignment. *Proceedings of the National Academy of Sciences, 112*(41), 12752-12757. doi:10.1073/pnas.1500331112

Jefferson T (1984) *Reprint of Notes on the State of Virginia*. Library of America, New York: Literary classics of the United States.

Jordan F M, Gray RD, Greenhill SJ, Mace R (2009) Matrilocal residence is ancestral in Austronesian societies. *Proceedings of the Royal Society B-Biological Sciences, 276*(1664), 1957-1964. doi:10.1098/rspb.2009.0088

Laland KN, Brown GR (2011) *Sense and nonsense: Evolutionary perspectives on human behaviour*: Oxford University Press.

Lipo CP, O'Brien MJ, Collard M, Shennan S (eds.) (2006) *Mapping our ancestors: Phylogenetic approaches in anthropology and prehistory*. New York: Aldine.

Lycett SJ (2015) Cultural evolutionary approaches to artifact variation over time and space: basis, progress, and prospects. *Journal of Archaeological Science, 56*, 21-31. doi:10.1016/j.jas.2015.01.004

Lycett SJ, Collard M, McGrew WC (2007) Phylogenetic analyses of behavior support existence of culture among wild chimpanzees. *Proceedings of the National Academy of Sciences*, *104*(45), 17588-17592. doi:10.1073/pnas.0707930104

Mace R, Pagel M (1994) The Comparative Method in Anthropology. *Current Anthropology, 35*(5), 549-564. doi:10.1086/204317

Mace R, Pagel M (1995) A latitudinal gradient in the density of human languages in North America. *Proceedings of the Royal Society B-Biological Sciences, 261*(1360), 117-121. doi:10.1098/rspb.1995.0125

Mace R, Holden CJ, Shennan S (eds.) (2005) *The evolution of cultural diversity: a phylogenetic approach*. London: University College London Press.

Mesoudi A (2011) *Cultural Evolution: How Darwinian Theory Can Explain Human Culture and Synthesize the Social Sciences*. Chicago: University of Chicago Press.

Moore JL, Manne L, Brooks T, Burgess ND, Davies R, Rahbek C, . . . Balmford A (2002) The distribution of cultural and biological diversity in Africa. *Proceedings of the Royal Society B-Biological Sciences, 269*(1501), 1645-1653. doi:10.1098/rspb.2002.2075

Nunn PD, Reid NJ (2016) Aboriginal Memories of Inundation of the Australian Coast Dating from More than 7000 Years Ago. *Australian Geographer, 47*(1), 11-47. doi:10.1080/00049182.2015.1077539

Oppenheimer S, Richards M (2001) Fast trains, slow boats, and the ancestry of the Polynesian islanders. *Science Progress, 84*(3), 157-181. doi:10.3184/003685001783238989

Pagel M (2009) Human language as a culturally transmitted replicator. *Nature Reviews Genetics, 10*(6), 405-415. doi:10.1038/nrg2560

Pagel M, Atkinson QD, Meade A (2007) Frequency of word-use predicts rates of lexical evolution throughout Indo-European history. *Nature, 449*(7163), 717-U717. doi:10.1038/nature06176

Peoples HC, Duda P, Marlowe FW (2016) Hunter-Gatherers and the Origins of Religion. *Human Nature-an Interdisciplinary Biosocial Perspective, 27*(3), 261-282. doi:10.1007/s12110-016-9260-0

Rivero DG, O'Brien MJ (2014) Phylogenetic Analysis Shows That Neolithic Slate Plaques from the Southwestern Iberian Peninsula Are Not Genealogical Recording Systems. *Plos One, 9*(2), 18. doi:10.1371/journal.pone.0088296

Rogers DS, Feldman MW, Ehrlich PR (2009) Inferring population histories using cultural data. *Proceedings of the Royal Society B-Biological Sciences, 276*(1674), 3835-3843. doi:10.1098/rspb.2009.1088

Schleicher A (1863) *Die Darwinsche Theorie und die Sprachwissenschaft – offenes Sendschreiben an Herrn Dr. Ernst Haeckel*. Weimar: H. Böhlau.

Skoglund P, Posth C, Sirak K, Spriggs M, Valentin F, Bedford S, . . . Reich D (2016) Genomic insights into the peopling of the Southwest Pacific. *Nature, 538*(7626), 510-513. doi:10.1038/nature19844

Soukup V (2004) *Přehled antropologických teorií kultury*. Praha: Portál.

Tehrani JJ (2013) The Phylogeny of Little Red Riding Hood. *Plos One, 8*(11), e78871. doi:10.1371/journal.pone.0078871

Tehrani JJ, Riede F (2008) Towards an archaeology of pedagogy: learning, teaching and the generation of material culture traditions. *World Archaeology, 40*(3), 316-331. doi:10.1080/00438240802261267

Tehrani J, Collard M (2002) Investigating cultural evolution through biological phylogenetic analyses of Turkmen textiles. *Journal of Anthropological Archaeology, 21*(4), 443-463. doi:10.1016/s0278-4165(02)00002-8

Wheeler WC, Whiteley PM (2015) Historical linguistics as a sequence optimization problem: the evolution and biogeography of Uto-Aztecan languages. *Cladistics, 31*(2), 113-125. doi:10.1111/cla.12078

Willems M, Lord E, Laforest L, Labelle G, Lapointe F-J, Di Sciullo AM, Makarenkov V (2016) Using hybridization networks to retrace the evolution of Indo-European languages. *Bmc Evolutionary Biology, 16*(1), 180. doi:10.1186/s12862-016-0745-6

Summary of the results and future perspectives

The aim of this thesis was to infer a supertree of modern human populations based on published genetic and linguistic phylogenetic trees, and to use this supertree as a framework to study coevolution between genes and languages and evolution of culture (religion in particular), on a global scale.

The resulting supertree (Duda & Zrzavý, 2016) includes 186 selected populations that represent all world regions and major linguistic groups. It is the most comprehensive phylogenetic tree of human populations published to date. It is also the first use of the supertree method to reconstruct intraspecific phylogeny.

We devised a new method inspired by the "sensitivity analysis" (Wheeler, 1995) to assess the stability of the inferred supertree topology. We also identified individual populations whose phylogenetic position is particularly unstable ("wildcard taxa"), using the *Iter*PCR method (Pol & Escapa, 2009; Goloboff & Szumik, 2015).

The sensitivity analysis identified three large sections of topological instability, West Eurasia, Mainland and Island Southeast Asia and Oceania, and East Asia, reflecting the existing phylogenetic controversies (Friedlaender et al., 2008; Jinam et al., 2012; Busby et al., 2015). The populations with the least stable phylogenetic position were Andamanese and Malagasy, reflecting the complex population history of these populations (Regueiro et al., 2008; Chaubey & Endicott, 2013; Aghakhanian et al., 2015).

The question of coevolution of genes and languages is considered fundamental but rarely studied by formal phylogenetic methods (Pakendorf, 2014). In order to investigate gene-culture coevolution on a global scale and to test for monophyly of the proposed linguistic groupings (language families and macrofamilies), we measured the congruence between linguistic data and the topology of the supertree and constrained this topology with linguistic data. This allowed us to reconstruct the relationships between the groups of populations speaking related languages based on genetic data. Linguistic classification fits rather poorly on the supertree topology, supporting a view that direct coevolution between genes and languages is far from universal. Most of the controversial linguistic macrofamilies were not supported by the resulting topology of the supertree.

Recent studies of the evolution of religion have focused on characteristics of the large prosocial religions that have emerged since the advent of agriculture. Relatively little attention has been paid to the religion of hunter-gatherers whose religious beliefs and behaviors have been evolving during the vast majority of human history (Lee & DeVore, 1968). Despite established speculations about various beliefs and behaviors that may represent an original form of religion, specific traits of nascent religiosity and the sequence in which they emerged have remained unknown. In order to study the origins and evolution of religion, we inferred a supertree of 33 hunter-gatherer populations included in the *Standard Cross-Cultural Sample* and *Ethnographic Atlas*. We reconstructed ancestral states for seven character describing hunter-gatherer religious beliefs and behaviors, and tested for coevolution between these characters and the direction of cultural change (Peoples et al., 2016).

Our results suggest that the oldest trait of religion, present in the most recent common ancestor of present-day hunter-gatherers, was animism, in agreement with long-standing beliefs about the fundamental role of this trait (Tylor, 1871). Belief in an afterlife emerged later, followed by shamanism and ancestor worship. Despite established speculation by Spencer (1870) and Tylor (1871) that ancestor worship represented the beginning of religion, our results suggest that worship of dead kin was not among the oldest traits of religion. Ancestor spirits or high gods who are active in human affairs were absent in early humans, indicating a deep history for the egalitarian nature of hunter-gatherer societies.

The results suggest that all studied characters evolved in concerted fashion as an integrated system of religious beliefs and practices but neither high gods nor active high gods exhibit correlated evolution with any other characters. This is in line with a variety of evidence from recent studies (Peoples and Marlowe 2012; Norenzayan, 2013; Botero et al., 2014) suggesting that if a society acquires belief in an omniscient and potentially morally punishing creator deity, it does so regardless of other aspects of its religion but more as a reflection of its social and political structure.

The thesis also includes two book chapters from books published by Academia, The Publishing House of the Academy of Sciences of the Czech Republic. They concern evolutionary history of Central African pygmies (Duda, 2015) and phylogenetic approaches to evolution of culture (Duda, *in press*).

The methods and results developed and presented in this thesis should help to improve the scope and quality of this research. Since its publication, the supertree has been used as a framework to investigate the association between men's social status and reproductive success in nonindustrial societies (von Rueden & Jaeggi, 2016). More studies investigating evolution of a broad range of behavioral, cultural and ecological traits will follow. They will require a larger, more comprehensive supertree, which is currently in preparation. Based on the updated dataset, a time-calibrated supertree of populations included in the *Standard Cross-Cultural Sample* is being prepared. This supertree will be used to reevaluate the existing hypotheses about ecological and cultural determinants of mating system and parental investment in human cultures in collaboration with Riana Minocher and Adrian Jaeggi at the Department of Anthropology of the Emory University in Atlanta. A supertree that will include over 1,000 human populations, including some extinct ones, is also in preparation. It will be used to study geographical and ecological patterns of cultural diversification in collaboration with David Storch at the Center for Theoretical Study in Prague.

References

Aghakhanian, F., Yunus, Y., Naidu, R., Jinam, T., Manica, A., Peng, H. B., & Phipps, M. E. (2015). Unravelling the Genetic History of Negritos and Indigenous populations of Southeast Asia. *Genome Biology and Evolution, 7*(5), 1206-1215. doi:10.1093/gbe/evv065

Botero, C. A., Gardner, B., Kirby, K. R., Bulbulia, J., Gavin, M. C., & Gray, R. D. (2014). The ecology of religious beliefs. *Proceedings of the National Academy of Sciences of the United States of America, 111*(47), 16784-16789. doi:10.1073/pnas.1408701111

Busby, G. B. J., Hellenthal, G., Montinaro, F., Tofanelli, S., Bulayeva, K., Rudan, I., . . . Capelli, C. (2015). The Role of Recent Admixture in Forming the Contemporary West Eurasian Genomic Landscape. *Current Biology, 25*(19), 2518-2526. doi:10.1016/j.cub.2015.08.007

Chaubey, G., & Endicott, P. (2013). The Andaman Islanders in a regional genetic context: Reexamining the evidence for an early peopling of the archipelago from South Asia. *Human Biology, 85*(1-3), 153-171.

Duda, P. (2015). Pygmejové pohledem evoluční biologie. In L. Hroníková & M. Schierová (Eds.), *Pygmejové: nejmenší lidé pohledem antropologie & Šebestova sbírka v Hrdličkově muzeu člověka PřF UK* (pp. 18-48). Praha: Academia.

Duda, P. (*in press*). Kulturní fylogenetika - Využití fylogenetických metod ke studiu evoluční historie jazyka a kultury. In L. Ovčáčková (Ed.), *Biologické a evoluční teorie kultury*. Praha: Academia.

Duda, P., & Zrzavý, J. (2016). Human population history revealed by a supertree approach. *Scientific Reports, 6,* 10. doi:10.1038/srep29890

Friedlaender, J. S., Friedlaender, F. R., Reed, F. A., Kidd, K. K., Kidd, J. R., Chambers, G. K., . . . Weber, J. L. (2008). The genetic structure of Pacific islanders. *Plos Genetics, 4*(1). doi:10.1371/journal.pgen.0040019

Goloboff, P. A., & Szumik, C. A. (2015). Identifying unstable taxa: Efficient implementation of triplet-based measures of stability, and comparison with Phyutility and RogueNaRok. *Molecular Phylogenetics and Evolution, 88,* 93-104.

Jinam, T. A., Hong, L. C., Phipps, M. E., Stoneking, M., Ameen, M., Edo, J., . . . Consortium, H. P.-A. S. (2012). Evolutionary history of continental Southeast Asians:“Early train” hypothesis based on genetic analysis of mitochondrial and autosomal DNA data. *Molecular Biology and Evolution, 29*(11), 3513-3527. doi:10.1093/molbev/mss169

Lee, R. B., & DeVore, I. (1968). Problems in the study of hunters and gatherers. In R. B. Lee & I. DeVore (Eds.) *Man the Hunter* (pp. 1–12). Chicago: Aldine.

Norenzayan, A. (2013). *Big Gods: How Religion Transformed Cooperation and Conflict*. Princeton: Princeton University Press.

Pakendorf, B. (2014). Coevolution of languages and genes. *Current opinion in genetics & development, 29,* 39-44. doi:10.1016/j.gde.2014.07.006

Peoples, H. C., & Marlowe, F. W. (2012). Subsistence and the evolution of religion. *Human Nature, 23*(3), 253-269. doi:10.1007/s12110-012-9148-6

Peoples, H. C., Duda, P., & Marlowe, F. W. (2016). Hunter-gatherers and the origins of religion. *Human Nature-an Interdisciplinary Biosocial Perspective, 27*(3), 261-282. doi:10.1007/s12110-016-9260-0

Pol, D., & Escapa, I. H. (2009). Unstable taxa in cladistic analysis: identification and the assessment of relevant characters. *Cladistics, 25*(5), 515-527. doi:10.1111/j.1096-0031.2009.00258.x

Regueiro, M., Mirabal, S., Lacau, H., Caeiro, J. L., Garcia-Bertrand, R. L., & Herrera, R. J. (2008). Austronesian genetic signature in East African Madagascar and Polynesia. *Journal of Human Genetics, 53*(2), 106-120. doi:10.1007/s10038-007-0224-4

Spencer, H. (1870). On ancestor worship and other peculiar beliefs. *Fortnightly Review, 13*, 535-550.

Tylor, E. B. (1871). *Primitive Culture: Researches into the Development of Mythology, Philosophy, Religion, Art, and Custom* (Vol. 2). London: Murray.

von Rueden, C. R., & Jaeggi, A. V. (2016). Men's status and reproductive success in 33 nonindustrial societies: Effects of subsistence, marriage system, and reproductive strategy. *Proceedings of the National Academy of Sciences of the United States of America, 113*(39), 10824-10829. doi:10.1073/pnas.1606800113

Wheeler, W. C. (1995). Sequence alignment, parameter sensitivity, and the phylogenetic analysis of molecular data. *Syst Biol, 44*(3), 321-331. doi:10.2307/2413595

Appendix I

Supplementary information for Chapter I

Human population history revealed by a supertree approach

Pavel Duda & Jan Zrzavý

Department of Zoology, Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic

Correspondence and request for material should be addressed to P. D. (email: dudapa01@.gmail.com)

SUPPLEMENTARY INFORMATION

Includes Supplementary Tables S1-S8, Supplementary Figures S1-S24 and Supplementary Methods.

SUPPLEMENTARY TABLES

Supplementary Table S1. Type populations used to replace populations groups, subpopulations, and populations of unspecified ethnic origin in the source trees.

Taxon name(s) (source trees)	References	Taxon name (supertree)	ISO 639-3	Comment
!Kung; Bushmen	1-7	!Kung	vaj, knw	
African	8	Yoruba	yri	Individuals from Nigeria
Africans; South African; South African Blacks	9-11	Xhosa	xho	Individuals from Johanesburg, South Africa
		Zulu	zul	
Algonquian	12	Ojibwa	ojc, ojg, ojb, ojs	
		Cree	cre	
Andamanese; Andaman Islanders	1,13	Andamanese	oon, anq, gac	
Ashkenazi Jews; Polish Jews and Russian Jews	14-19	Ashkenazi Jew	ydd, yih	
Australian aboriginals	various studies	AUSTRALIAN		
Central Australian Aboriginals	20	Warlpiri	wbp	
CEU; European(s); European (Northwestern); European-American; U.S.; American White; US Caucasian; N.American Caucasoids; N.A. Caucasoids; Caiucasoids; Canada; Canadians	various genome-wide SNP- and STR-based studies	English	eng	The CEU sample[21] shows ancestry most similar to British in England and Scotland[22]
Ethiopian	23	Amhara Oromo	amh gax	
Great Andamanese	24,25	Andamanese	oon, anq, gac	
Guarani	various HLA-based studies	Guarani	grn	
Han Chinese from Beijing, Shanghai, Shandong, and Wuxi	various studies	Northern Han Chinese	cmn	
Han Chinese from Hongkong, Fujian, Chaosan, Huizhou, Hainan, Guangdong, Guangxi, Yunnan, Guizhou, Huangzhou (Han Cantonese), and Xiaamen		Southern Han Chinese	cmn, yue	
Han Chinese from Singapore	various studies	Singapore Chinese	cmn	
HGDP Papuan; PNG highland; PNG Highlanders; New Guinea Highlands	various studies	Goroka	for, gim	
Chinese from Taiwan	various studies	Min Nan Taiwanese	nan	
Chinese; Taiwan(ese); Taiwan Han; China Taiwan	2,3,26-33	Min Nan Taiwanese	nan	

Jarawa	24	Andamanese	oon, anq, gac	
Lapp(s); Laps; Eur. Saam(i); Saami (Lapps); Swedish Lapp; Norwegian Lapp; Finnish Lapp; Skolt Saami; North Saami; Ume Saami; Sweden South Sami; Sweden North Sami; Saami Russia	1,11,20,23,27,34-45	Saami	sia, smn, sjd, sms, sjt	
Madagascar	46,47	Malagasy	plt	
Micronesia(n)	48; various genome-wide SNP- and STR-based studies	Kosraean	kos	
NAN Melanesian; HGDP Melanesian; Melanesian; Southeast Bougainville	23; various genome-wide SNP- and STR-based studies	Naasioi	nas	
Nigeria	4,49	Yoruba	yor	
Non-Ashkenazi Jews from Israel; Syria; Iraq; Iran; Turkey; Georgia and Azerbaijan; Hebrew	16,17; various HLA-based studies	Mizrahi Jew		
North Dravidian	23	Brahui	brh	
North European	50,51	German	deu	
Onge	25,52,53	Andamanese	oon, anq, gac	
Polynesia(n)	48; various genome-wide SNP- and STR-based studies	Samoan	smo	
Romania; Romanian(s); Rumanian; Romanian List; Vlach	17,30,54-70	Romanian	ron, rup	
South Dravidian	25,71	Tamil	tam	
South Indian	17,23,25,36,72	Malayalam	mal	
Taiwan; Taiwan Aborigine(s); Taiwanese; Indigenous Taiwanese	36,49,73-76	Atayal	tay	
Uralic Siberian	23	Nenets	yrk	

References

1 Sellen, D. W. & Mace, R. Fertility and Mode of Subsistence: A Phylogenetic Analysis. *Current Anthropology* **38**, 878-889, doi:10.1086/204677 (1997).

2 Comas, D. *et al.* Trading genes along the silk road: mtDNA sequences and the origin of central Asian populations. *American Journal of Human Genetics* **63**, 1824-1838, doi:10.1086/302133 (1998).

3 Chu, C. C. *et al.* Diversity of HLA among Taiwan's indigenous tribes and the Ivatans in the Philippines. *Tissue Antigens* **58**, 9-18, doi:10.1034/j.1399-0039.2001.580102.x (2001).

4 Nasidze, I. *et al.* Alu insertion polymorphisms and the genetic structure of human populations from the Caucasus. *European Journal of Human Genetics* **9**, 267-272, doi:10.1038/sj.ejhg.5200615 (2001).

5 Xing, J. C. *et al.* Fine-scaled human genetic structure revealed by SNP microarrays. *Genome Research* **19**, 815-825, doi:10.1101/gr.085589.108 (2009).

6 Alkorta-Aranburu, G. *et al.* The Genetic Architecture of Adaptations to High Altitude in Ethiopia. *Plos Genetics* **8**, doi:10.1371/journal.pgen.1003110 (2012).

7 Wangkumhang, P. *et al.* Insight into the Peopling of Mainland Southeast Asia from Thai Population Genetic Structure. *PloS one* **8**, e79522 (2013).

8 Yuasa, I. *et al.* A hypervariable STR polymorphism in the CFI gene: Southern origin of East Asian-specific group H alleles. *Legal Medicine* **15**, 239-243, doi:10.1016/j.legalmed.2013.04.001 (2013).

9 Bannai, M. *et al.* Analysis of HLA genes and haplotypes in Ainu (from Hokkaido, northern Japan) supports the premise that they descent from Upper Paleolithic populations of East Asia. *Tissue Antigens* **55**, 128-139, doi:10.1034/j.1399-0039.2000.550204.x (2000).

10 Arnaiz-Villena, A. *et al.* HLA alleles and haplotypes in the Turkish population: relatedness to Kurds, Armenians and other Mediterraneans. *Tissue Antigens* **57**, 308-317, doi:10.1034/j.1399-0039.2001.057004308.x (2001).

11 Martinez-Laso, J. *et al.* HLA in Jaidukama: an Amerindian secluded Colombian population with new haplotypes and Asian and Pacific-shared alleles. *Molecular Biology Reports* **38**, 3689-3701, doi:10.1007/s11033-010-0483-6 (2011).

12 Rubicz, R. *et al.* Genetic Structure of Native Circumpolar Populations Based on Autosomal, Mitochondrial, and Y Chromosome DNA Markers. *American Journal of Physical Anthropology* **143**, 62-74, doi:10.1002/ajpa.21290 (2010).

13 Walker, R. S., Hill, K. R., Flinn, M. V. & Ellsworth, R. M. Evolutionary History of Hunter-Gatherer Marriage Practices. *Plos One* **6**, doi:10.1371/journal.pone.0019066 (2011).

14 Kopelman, N. M. *et al.* Genomic microsatellites identify shared Jewish ancestry intermediate between Middle Eastern and European populations. *Bmc Genetics* **10**, doi:10.1186/1471-2156-10-80 (2009).

15 Li, H., Cho, K., Kidd, J. R. & Kidd, K. K. Genetic Landscape of Eurasia and "Admixture" in Uyghurs. *American Journal of Human Genetics* **85**, 934-937, doi:10.1016/j.ajhg.2009.10.024 (2009).

16 Atzmon, G. *et al.* Abraham's Children in the Genome Era: Major Jewish Diaspora Populations Comprise Distinct Genetic Clusters with Shared Middle Eastern Ancestry. *American Journal of Human Genetics* **86**, 850-859, doi:10.1016/j.ajhg.2010.04.015 (2010).

17 Behar, D. M. *et al.* The genome-wide structure of the Jewish people. *Nature* **466**, 238-U112, doi:10.1038/nature09103 (2010).

18 Haber, M. *et al.* Genome-Wide Diversity in the Levant Reveals Recent Structuring by Culture. *Plos Genetics* **9**, 8, doi:10.1371/journal.pgen.1003316 (2013).

19 Pemberton, T. J., DeGiorgio, M. & Rosenberg, N. A. Population Structure in a Comprehensive Genomic Data Set on Human Microsatellite Variation. *G3-Genes Genomes Genetics* **3**, 891-907, doi:10.1534/g3.113.005728 (2013).

20 Nei, M. & Roychoudhury, A. K. Evolutionary relationships of human populations on a global scale. *Molecular Biology and Evolution* **10**, 927-943 (1993).

21 Gibbs, R. A. *et al.* The International HapMap Project. *Nature* **426**, 789-796, doi:10.1038/nature02168 (2003).

22 Shriner, D., Tekola-Ayele, F., Adeyemo, A. & Rotimi, C. N. Genome-wide genotype and sequence-based reconstruction of the 140,000 year history of modern human ancestry. *Scientific Reports* **4**, doi:10.1038/srep06055 (2014).

23 Cavalli-Sforza, L. L., Menozzi, P. & Piazza, A. *The history and geography of human genes*. (Princeton university press, 1994).

24 Abbi, A. Is Great Andamanese genealogically and typologically distinct from Onge and Jarawa? *Language Sciences* **31**, 791-812, doi:10.1016/j.langsci.2008.02.002 (2009).

25 Pugach, I., Delfin, F., Gunnarsdottir, E., Kayser, M. & Stoneking, M. Genome-wide data substantiate Holocene gene flow from India to Australia. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 1803-1808, doi:10.1073/pnas.1211927110 (2013).

26 Horai, S. *et al.* mtDNA polymorphism in East Asian populations, with special reference to the peopling of Japan. *American Journal of Human Genetics* **59**, 579-590 (1996).

27 Wells, R. S. *et al.* The Eurasian Heartland: A continental perspective on Y-chromosome diversity. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 10244-10249, doi:10.1073/pnas.171305098 (2001).

28 Tajima, A. *et al.* Mitochondrial DNA polymorphisms in nine aboriginal groups of Taiwan: implications for the population history of aboriginal Taiwanese. *Human Genetics* **113**, 24-33, doi:10.1007/s00439-003-0945-1 (2003).

29 Tajima, A. *et al.* Genetic origins of the Ainu inferred from combined DNA analyses of maternal and paternal lineages. *Journal of Human Genetics* **49**, 187-193, doi:10.1007/s10038-004-0131-x (2004).

30 Nowak, J. *et al.* Allele and extended haplotype polymorphism of HLA-A, -C, -B, -DRB1 and -DQB1 loci in Polish population and genetic affinities to other populations. *Tissue Antigens* **71**, 193-205, doi:10.1111/j.1399-0039.2007.00991.x (2008).

31 Alfonso-Sanchez, M. A., Perez-Miranda, A. M. & Herrera, R. J. Autosomal microsatellite variability of the Arrernte people of Australia. *American Journal of Human Biology* **20**, 91-99, doi:10.1002/ajhb.20685 (2008).

32 Zhu, B. F. *et al.* Distributions of HLA-A and -B alleles and haplotypes in the Yi ethnic minority of Yunnan, China: relationship to other populations. *Journal of Zhejiang University-Science B* **11**, 127-135, doi:10.1631/jzus.B0900232 (2010).

33 Yuliwulandari, R. *et al.* Polymorphisms of HLA genes in Western Javanese (Indonesia): close affinities to Southeast Asian populations. *Tissue Antigens* **73**, 46-53, doi:10.1111/j.1399-0039.2008.01178.x (2009).

34 Holden, C. & Mace, R. Phylogenetic analysis of the evolution of lactose digestion in adults. *Human Biology* **69**, 605-628 (1997).

35 Ingman, M., Kaessmann, H., Paabo, S. & Gyllensten, U. Mitochondrial genome variation and the origin of modern humans. *Nature* **408**, 708-713 (2000).

36 Ingman, M. & Gyllensten, U. Mitochondrial genome variation and evolutionary history of Australian and New Guinean aborigines. *Genome Research* **13**, 1600-1606, doi:10.1101/gr.686603 (2003).

37 Nasidze, I., Sarkisian, T., Kerimov, A. & Stoneking, M. Testing hypotheses of language replacement in the Caucasus: evidence from the Y-chromosome. *Human Genetics* **112**, 255-261, doi:10.1007/s00439-002-0874-4 (2003).

38 Agrawal, S., Srivastava, S. K., Borkar, M. & Chaudhuri, T. K. Genetic affinities of north and northeastern populations of India: inference from HLA-based study. *Tissue Antigens* **72**, 120-130, doi:10.1111/j.1399-0039.2008.01083.x (2008).

39 Moscoso, J. *et al.* HLA genes of Aleutian Islanders living between Alaska (USA) and Kamchatka (Russia) suggest a possible southern Siberia origin. *Molecular Immunology* **45**, 1018-1026, doi:10.1016/j.molimm.2007.07.024 (2008).

40 Biro, A. Z., Zalan, A., Volgyi, A. & Pamjav, H. A Y-Chromosomal Comparison of the Madjars (Kazakhstan) and the Magyars (Hungary). *American Journal of Physical Anthropology* **139**, 305-310, doi:10.1002/ajpa.20984 (2009).

41 Arnaiz-Villena, A. *et al.* The Origin of Amerindians and the Peopling of the Americas According to HLA Genes: Admixture with Asian and Pacific People. *Current Genomics* **11**, 103-114, doi:10.2174/138920210790886862 (2010).

42 Krause, J. *et al.* The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. *Nature* **464**, 894-897, doi:10.1038/nature08976 (2010).

43 Huyghe, J. R. *et al.* A genome-wide analysis of population structure in the Finnish Saami with implications for genetic association studies. *European Journal of Human Genetics* **19**, 347-352, doi:10.1038/ejhg.2010.179 (2011).

44 Suslova, T. A. *et al.* HLA gene and haplotype frequencies in Russians, Bashkirs and Tatars, living in the Chelyabinsk Region (Russian South Urals). *International Journal of Immunogenetics* **39**, 394-408, doi:10.1111/j.1744-313X.2012.01117.x (2012).

45 Honkola, T. *et al.* Cultural and climatic changes shape the evolutionary history of the Uralic languages. *Journal of Evolutionary Biology* **26**, 1244-1253, doi:10.1111/jeb.12107 (2013).

46 Regueiro, M. *et al.* Austronesian genetic signature in East African Madagascar and Polynesia. *Journal of Human Genetics* **53**, 106-120, doi:10.1007/s10038-007-0224-4 (2008).

47 Mirabal, S., Cadenas, A. M., Garcia-Bertrand, R. & Herrera, R. J. Ascertaining the Role of Taiwan as a Source for the Austronesian Expansion. *American Journal of Physical Anthropology* **150**, 551-564, doi:10.1002/ajpa.22226 (2013).

48 Omoto, K. & Saitou, N. Genetic origins of the Japanese: A partial support for the dual structure hypothesis. *American Journal of Physical Anthropology* **102**, 437-446 (1997).

49 Edinur, H. A. *et al.* HLA polymorphism in six Malay subethnic groups in Malaysia. *Human Immunology* **70**, 518-526, doi:10.1016/j.humimm.2009.04.003 (2009).

50 Mesa, N. R. *et al.* Autosomal, mtDNA, and Y-chromosome diversity in Amerinds: Pre- and post-Columbian patterns of gene flow in South America. *American Journal of Human Genetics* **67**, 1277-1286 (2000).

51 Ayub, Q. *et al.* Reconstruction of human evolutionary tree using polymorphic autosomal microsatellites. *American Journal of Physical Anthropology* **122**, 259-268, doi:10.1002/ajpa.10234 (2003).

52 Reesink, G., Singer, R. & Dunn, M. Explaining the Linguistic Diversity of Sahul Using Population Models. *Plos Biology* **7**, doi:10.1371/journal.pbio.1000241 (2009).

53 Chaubey, G. & Endicott, P. The Andaman Islanders in a Regional Genetic Context: Reexamining the Evidence for an Early Peopling of the Archipelago from South Asia. *Human Biology* **85**, 153-171 (2013).

54 Munkhbat, B. *et al.* Molecular analysis of HLA polymorphism in Khoton-Mongolians. *Tissue Antigens* **50**, 124-134, doi:10.1111/j.1399-0039.1997.tb02851.x (1997).

55 Sanchez-Velasco, P., Karadsheh, N. S., Garcia-Martin, A., de Alegria, C. R. & Leyva-Cobian, F. Molecular analysis of HLA allelic frequencies and haplotypes in Jordanians and comparison with other related populations. *Human Immunology* **62**, 901-909, doi:10.1016/s0198-8859(01)00289-0 (2001).

56 Sanchez-Velasco, P. & Leyva-Cobian, F. The HLA class I and class II allele frequencies studied at the DNA level in the Svanetian population (Upper Caucasus) and their relationships to Western European populations. *Tissue Antigens* **58**, 223-233, doi:10.1034/j.1399-0039.2001.580402.x (2001).

57 Ivanova, M. *et al.* HLA polymorphism in Bulgarians defined by high-resolution typing methods in comparison with other populations. *Tissue Antigens* **60**, 496-504, doi:10.1034/j.1399-0039.2002.600605.x (2002).

58 Arnaiz-Villena, A. *et al.* HLA genes in the chuvashian population from European Russia: Admixture of central European and Mediterranean populations. *Human Biology* **75**, 375-392, doi:10.1353/hub.2003.0040 (2003).

59 Gray, R. D. & Atkinson, Q. D. Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature* **426**, 435-439, doi:10.1038/nature02029 (2003).

60 Pagel, M. & Meade, A. in *The evolution of cultural diversity: a phylogenetic approach* (eds Ruth Mace, Clare J Holden, & Stephen Shennan) Ch. 13, 235-256 (UCL Press, 2005).

61 Fortunato, L., Holden, C. & Mace, R. From bridewealth to dowry? *Human Nature* **17**, 355-376, doi:10.1007/s12110-006-1000-4 (2006).

62 Pagel, M., Atkinson, Q. D. & Meade, A. Frequency of word-use predicts rates of lexical evolution throughout Indo-European history. *Nature* **449**, 717-U717, doi:10.1038/nature06176 (2007).

63 Serva, M. & Petroni, F. Indo-European languages tree by Levenshtein distance. *EPL (Europhysics Letters)* **81**, doi:10.1209/0295-5075/81/68005 (2008).

64 Geisler, H. & List, J.-M. in *Arbeitstagung der Indogermanischen Gesellschaft 2009: Die Ausbreitung des Indogermanischen Thesen aus Sprachwissenschaft, Archäologie und Genetik* (Würzburg, Germany, 2009).

65 Sulcebe, G. *et al.* HLA allele and haplotype frequencies in the Albanian population and their relationship with the other European populations. *International Journal of Immunogenetics* **36**, 337-343, doi:10.1111/j.1744-313X.2009.00868.x (2009).

66 Greenhill, S. J., Atkinson, Q. D., Meade, A. & Gray, R. D. The shape and tempo of language evolution. *Proceedings of the Royal Society B-Biological Sciences* **277**, 2443-2450, doi:10.1098/rspb.2010.0051 (2010).

67 Rasmussen, M. *et al.* An Aboriginal Australian Genome Reveals Separate Human Dispersals into Asia. *Science* **334**, 94-98, doi:10.1126/science.1211177 (2011).

68 Bouckaert, R. *et al.* Mapping the Origins and Expansion of the Indo-European Language Family. *Science* **337**, 957-960, doi:10.1126/science.1219669 (2012).

69 Delmestri, A. & Cristianini, N. Linguistic Phylogenetic Inference by PAM-like Matrices. *Journal of Quantitative Linguistics* **19**, 95-120, doi:10.1080/09296174.2012.659001 (2012).

70 Mendizabal, I. *et al.* Reconstructing the Population History of European Romani from Genome-wide Data. *Current Biology* **22**, 2342-2349, doi:10.1016/j.cub.2012.10.039 (2012).

71 Melton, T. *et al.* Polynesian genetic affinities with Southeast Asian populations as identified by mtDNA analysis. *American Journal of Human Genetics* **57**, 403-414 (1995).

72 Chhaya, S., Desai, S. & Saranath, D. HLA polymorphisms in Sindhi community in Mumbai, India. *International Journal of Immunogenetics* **37**, 373-377, doi:10.1111/j.1744-313X.2010.00936.x (2010).

73 Redd, A. J. & Stoneking, M. Peopling of Sahul: mtDNA variation in Aboriginal Australian and Papua New Guinean populations. *American Journal of Human Genetics* **65**, 808-828, doi:10.1086/302533 (1999).

74 Shaw, C. K., Chen, L. L., Lee, A. & Lee, T. D. Distribution of HLA gene and haplotype frequencies in Taiwan: a comparative study among Min-nan, Hakka, Aborigines and Mainland Chinese. *Tissue Antigens* **53**, 51-64, doi:10.1034/j.1399-0039.1999.530106.x (1999).

75 Nasidze, I. & Stoneking, M. Mitochondrial DNA variation and language replacements in the Caucasus. *Proceedings of the Royal Society B-Biological Sciences* **268**, 1197-1206 (2001).

76 Fu, Q. M. *et al.* DNA analysis of an early modern human from Tianyuan Cave, China. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 2223-2227, doi:10.1073/pnas.1221359110 (2013).

Supplementary Table S2. The representative dataset. 186 populations included in the study, ISO 639-3 code, principal source study, country of origin and geographic coordinates (based on principal source study), presence of the population in the genomic source trees, and linguistic affiliation accepted from *Ethnologue*, Ruhlen, and Greenberg & Ruhlen.

Taxon name (supertree)	**ISO 639-3**	**Principal source(s)**	**Location**	**Latitude (Degrees North)**	**Longitude (Degrees East)**	**Genomic source trees**	**Linguistic classification[1]**	**Linguistic classification[2,3]**
Khwe	xuu	Schlebusch et al. (2012)	Angola	-17.363921	22.950439	Y	Khoisan: Southern Africa	Khoisan
Ju'|hoan	ktz	Rosenberg et al. (2002, 2005); Ramachandran et al. (2005); Schlebusch et al. (2012)	Namibia	-19.597399	20.494995	Y	Khoisan: Southern Africa	Khoisan
!Kung	vaj, knw	Schlebusch et al. (2012)	Angola	-14.628943	17.666016	Y	Khoisan: Southern Africa	Khoisan
‡Khomani	ngh	Schlebusch et al. (2012)	South Africa	-26.974138	20.794373	Y	Khoisan: Southern Africa	Khoisan
Nama	naq	Schlebusch et al. (2012)	Namibia	-22.558559	17.072754	Y	Khoisan: Southern Africa	Khoisan
Mbuti Pygmy	efe	Rosenberg et al. (2002, 2005); Ramachandran et al. (2005)	Congo	1	29	Y	Nilo-Saharan	Nilo-Saharan
Aka Pygmy	axk	Rosenberg et al. (2002, 2005); Ramachandran et al. (2005)	Central African Republic	4	17	Y	Niger-Congo: Bantoid	Niger-Kordofanian
Fulani	fuv	Tishkoff et al. (2009)	Nigeria	11	11	Y	Niger-Congo: Atlantic	Niger-Kordofanian
Masana	mcn	Tishkoff et al. (2009)	Cameroon	10.3	15.3	Y	Afro-Asiatic: Chadic	Afro-Asiatic
Kotoko	kot	Tishkoff et al. (2009)	Cameroon	11.8	14.8	Y	Afro-Asiatic: Chadic	Afro-Asiatic
Podokwo	pbi	Tishkoff et al. (2009)	Cameroon	11	12.1	Y	Afro-Asiatic: Chadic	Afro-Asiatic

Ndebele	nde	Rosenberg et al. (2002, 2005); Ramachandran et al. (2005)	South Africa	-25.5692643	24.25	Y	Niger-Congo: Bantoid	Niger-Kordofanian
Swati	ssw	Rosenberg et al. (2002, 2005); Ramachandran et al. (2005)	South Africa	-25.5692643	24.25	Y	Niger-Congo: Bantoid	Niger-Kordofanian
Xhosa	xho	Tishkoff et al. (2009)	South Africa	-32	28	Y	Niger-Congo: Bantoid	Niger-Kordofanian
Zulu	zul	Rosenberg et al. (2002, 2005); Ramachandran et al. (2005)	South Africa	-25.5692643	24.25	Y	Niger-Congo: Bantoid	Niger-Kordofanian
Mandinka	mnk	Rosenberg et al. (2002, 2005); Ramachandran et al. (2005)	Senegal	12	-12	Y	Niger-Congo: Mande	Niger-Kordofanian
Yoruba	yor	HapMap	Nigeria	7.4	3.9	Y	Niger-Congo: Defoid	Niger-Kordofanian
Hausa	hau	Tishkoff et al. (2009)	Nigeria	10	7	Y	Afro-Asiatic: Chadic	Afro-Asiatic
Kikuyu	kik	Tishkoff et al. (2009)	Kenya	-1	37	Y	Niger-Congo: Bantoid	Niger-Kordofanian
Dinka	dip	Tishkoff et al. (2009)	Sudan	8	30	Y	Nilo-Saharan	Nilo-Saharan
Shilluk	shk	Tishkoff et al. (2009)	Sudan	10	32	Y	Nilo-Saharan	Nilo-Saharan
Anuak	anu	Pagani et al. (2012)	Ethiopia	8	34	Y	Nilo-Saharan	Nilo-Saharan
Hadza	hts	Tishkoff et al. (2009)	Tanzania	-3.8	35.3	Y	Khoisan: Hatsa	Khoisan
Sandawe	sad	Tishkoff et al. (2009)	Tanzania	-5.5	35.5	Y	Khoisan: Sandawe	Khoisan
Maasai	mas	Tishkoff et al. (2009)	Tanzania	-4	37	Y	Nilo-Saharan	Nilo-Saharan
Iraqw	irk	Tishkoff et al. (2009)	Tanzania	-4	35.5	Y	Afro-Asiatic: Cushtic	Afro-Asiatic
Aari	aiw	Pagani et al. (2012)	Ethiopia	6	37	Y	Afro-Asiatic: Omotic	Afro-Asiatic
Dogon	dds	Tishkoff et al. (2009)	Mali	14	-3	Y	Niger-Congo:	Niger-Kordofanian

							Dogon	an
Somali	som	Pagani et al. (2012)	Ethiopia	9	42	Y	Afro-Asiatic: Cushtic	Afro-Asiatic
Beta Israel	amh	Tishkoff et al. (2009)	Ethiopia	12	38	Y	Afro-Asiatic: Cushtic	Afro-Asiatic
Aweer	bob	Tishkoff et al. (2009)	Kenya	3.5	37	Y	Afro-Asiatic: Cushtic	Afro-Asiatic
Amhara	amh	Behar et al. (2010)	Ethiopia	10	39	Y	Afro-Asiatic: Semitic	Afro-Asiatic
Oromo	gax	Tishkoff et al. (2009)	Kenya	3	37.5	Y	Afro-Asiatic: Cushtic	Afro-Asiatic
Rendille	rel	Tishkoff et al. (2009)	Kenya	2.3	37.5	Y	Afro-Asiatic: Cushtic	Afro-Asiatic
Beja Hadandawa	bej	Tishkoff et al. (2009)	Sudan	21	36	Y	Afro-Asiatic: Cushtic	Afro-Asiatic
Berber	shi	Izaabel et al. (1998)	Morocco	30.93	-7.22	Y	Afro-Asiatic: Berber	Afro-Asiatic
Mozabite	mzb	Rosenberg et al. (2002, 2005); Ramachandran et al. (2005)	Algeria (Mzab)	32	3	Y	Afro-Asiatic: Berber	Afro-Asiatic
Algerian	ara (arq)	Arnaiz-Villena et al. (1995)	Algeria	36.75	3.2	Y	Afro-Asiatic: Semitic	Afro-Asiatic
Moroccan	ara (ary)	Gomez-Casado et al. (2000)	Morocco	33.2	-8.5	Y	Afro-Asiatic: Semitic	Afro-Asiatic
Egyptian	ara (arz)	Behar et al. (2010)	Egypt	29.31	30.84	Y	Afro-Asiatic: Semitic	Afro-Asiatic
Qatari	ara (arb)	Hunter-Zinck et al. (2010)	Qatar	25.4	51.2	Y	Afro-Asiatic: Semitic	Afro-Asiatic
Yemeni	ara (arb)	Behar et al. (2010)	Yemen	15.37	44.19	Y	Afro-Asiatic: Semitic	Afro-Asiatic
Rwala Bedouin	ayl	Rosenberg et al. (2002, 2005); Ramachandran et al. (2005)	Israel (Negev)	31	35	Y	Afro-Asiatic: Semitic	Afro-Asiatic
Saudi Arabian	ara (arb)	Behar et al. (2010)	Saudi Arabia	24.72	46.70	Y	Afro-Asiatic: Semitic	Afro-Asiatic

Palestinian	ara (ajp)	Rosenberg et al. (2002, 2005); Ramachandran et al. (2005)	Israel (Central)	32	35	Y	Afro-Asiatic: Semitic	Afro-Asiatic
Druze	apc	Rosenberg et al. (2002, 2005); Ramachandran et al. (2005)	Israel (Carmel)	32	35	Y	Afro-Asiatic: Semitic	Afro-Asiatic
Armenian	hye	Behar et al. (2010)	Armenia	40.18	44.51	Y	Indo-European	Indo-Hittite
Ashkenazi Jew	ydd, yih	Kopelman et al. (2009)	Israel (Ashkelon)	31.666667	34.566667	Y	Indo-European	Indo-Hittite
Mizrahi Jew	heb	Kopelman et al. (2009)	Israel (Ashkelon)	31.666667	34.566667	Y	Afro-Asiatic: Semitic	Afro-Asiatic
Moroccan Jew	heb, aju	Kopelman et al. (2009)	Israel (Ashkelon)	31.666667	34.566667	Y	Afro-Asiatic: Semitic	Afro-Asiatic
Syrian	ara (apc)	Behar et al. (2010)	Syria	33.51	36.28	Y	Afro-Asiatic: Semitic	Afro-Asiatic
Lebanese	ara (apc)	Clayton et al. (1997)	Lebanon	33.89	35.50	Y	Afro-Asiatic: Semitic	Afro-Asiatic
Kurd	ckb	Xing et al. (2010)	Iraq	36.74	43.89	Y	Indo-European	Indo-Hittite
Turk	tur	Behar et al. (2010)	Turkey	36.95	32.84	Y	Altaic: Turkic	Macro-Altaic
Greek	ell	Mendizabal et al. (2012)	Greece	40.6422	22.9456	Y	Indo-European	Indo-Hittite
Romanian	ron, rup	Behar et al. (2010)	Romania	44.43	26.10	Y	Indo-European	Indo-Hittite
Tuscan	ita	Rosenberg et al. (2002, 2005); Ramachandran et al. (2005)	Italy	43	11	Y	Indo-European	Indo-Hittite
Italian	ita	Rosenberg et al. (2002, 2005); Ramachandran et al. (2005)	Italy (Bergamo)	46	10	Y	Indo-European	Indo-Hittite
Sardinian	src	Rosenberg et al. (2002, 2005); Ramachandran et al. (2005)	Italy	40	9	Y	Indo-European	Indo-Hittite

French	fra	Rosenberg et al. (2002, 2005); Ramachandran et al. (2005)	France	46	2	Y	Indo-European	Indo-Hittite
Basque	eus	Rosenberg et al. (2002, 2005); Ramachandran et al. (2005)	France	43	0	Y	Language isolate	Dene-Caucasian
Spanish	spa	Martinez-Laso et al. 1995	Spain	40.42	-3.70	Y	Indo-European	Indo-Hittite
German	deu	Imanishi et al. (1992)	Germany	51.34	12.37	N	Indo-European	Indo-Hittite
English	eng	???	UK	51.5	-0.13	Y	Indo-European	Indo-Hittite
Polish	pol	Nowak et al. (2008)	Poland	51.11	17.03	Y	Indo-European	Indo-Hittite
Hungarian	hun	Mendizabal et al., 2012	Hungary	47.45	19.04	Y	Uralic	Uralic
Orcadian	sco	Rosenberg et al. (2002, 2005); Ramachandran et al. (2005)	Orkney Islands	59	-3	Y	Indo-European	Indo-Hittite
Russian	rus	Rosenberg et al. (2002, 2005); Ramachandran et al. (2005)	Russia	61	40	Y	Indo-European	Indo-Hittite
Finnish	fin	Clayton et al. (1997)	Finland	60.22	24.91	N	Uralic	Uralic
Saami	sia, smn, sjd, sms, sjt	Evseeva et al. (2002)	Russia (Kola peninsula)	68.00	35.02	Y	Uralic	Uralic
Georgian	kat	Behar et al. (2010)	Georgia	42.49	41.83	Y	Kartvelian	Kartvelian
Adygei	ady	Rosenberg et al. (2002, 2005); Ramachandran et al. (2005)	Russia (Caucasus)	44	39	Y	North Caucasian	Dene-Caucasian: Caucasian
Iranian	pes	Behar et al. (2010)	Iran	32.65	51.66	Y	Indo-European	Indo-Hittite
Makrani	bcc	Rosenberg et al. (2002, 2005); Ramachand	Pakistan	26	64	Y	Indo-European	Indo-Hittite

		ran et al. (2005)						
Brahui	brh	Rosenberg et al. (2002, 2005); Ramachandran et al. (2005)	Pakistan	30.49871492	66.5	Y	Dravidian	Dravidian
Balochi	bgp	Rosenberg et al. (2002, 2005); Ramachandran et al. (2005)	Pakistan	30.49871492	66.5	Y	Indo-European	Indo-Hittite
Sindhi	snd	Rosenberg et al. (2002, 2005); Ramachandran et al. (2005)	Pakistan	25.49063551	69	Y	Indo-European	Indo-Hittite
Pashtun	pst	Rosenberg et al. (2002, 2005); Ramachandran et al. (2005)	Pakistan	33.48700562	70.5	Y	Indo-European	Indo-Hittite
Kalash	kls	Rosenberg et al. (2002, 2005); Ramachandran et al. (2005)	Pakistan	35.99366014	71.5	Y	Indo-European	Indo-Hittite
Madiga	tel	Reich et al. (2009)	India (Andhra Pradesh)	17.58	79.35	Y	Dravidian	Dravidian
Tamil	tam	Rosenberg et al. (2006)	India (Tamil Nadu)	11.10167224	77.94827586	Y	Dravidian	Dravidian
Malayalam	mal	Rosenberg et al. (2006)	India (Kerala)	10	76.25	Y	Dravidian	Dravidian
Burusho	bsk	Rosenberg et al. (2002, 2005); Ramachandran et al. (2005)	Pakistan	36.49838568	74	Y	Language isolate	Dene-Caucasian
Hazara	haz	Rosenberg et al. (2002, 2005); Ramachandran et al. (2005)	Pakistan	33.49855601	70	Y	Indo-European	Indo-Hittite
Uyghur	uig	Rosenberg et al. (2002, 2005);	China	44	81	Y	Altaic: Turkic	Macro-Altaic

		Ramachandran et al. (2005)						
Kensiu	kns	The HUGO Pan-Asian SNP Consortium	Malaysia	5.7	100.9	Y	Austro-Asiatic: Mon-Khmer	Austric: Austroasiatic
Jehai	jhi	The HUGO Pan-Asian SNP Consortium	Malaysia	5.4	101.1	Y	Austro-Asiatic: Mon-Khmer	Austric: Austroasiatic
Amis	ami	Friedlaender et al. (2008)	Taiwan	23.3	121	Y	Austronesian	Austric: Austronesian
Paiwan	pwn	Gray et al. (2009)	Taiwan	22.6	120.8	N	Austronesian	Austric: Austronesian
Taroko	trv	Friedlaender et al. (2008)	Taiwan	23.3	121	Y	Austronesian	Austric: Austronesian
Atayal	tay	The HUGO Pan-Asian SNP Consortium	Taiwan	24.6	121.4	N	Austronesian	Austric: Austronesian
Tagalog	tgl	The HUGO Pan-Asian SNP Consortium	Philiippines	14.6	121.0	Y	Austronesian: Malayo-Polynesian	Austric: Austronesian
Cebuano	vis	The HUGO Pan-Asian SNP Consortium	Philippines	6.9	122.1	Y	Austronesian: Malayo-Polynesian	Austric: Austronesian
Toraja	sda, rob	The HUGO Pan-Asian SNP Consortium	Indonesia	-4.7	119.7	Y	Austronesian: Malayo-Polynesian	Austric: Austronesian
Bugis	bug	Hatin et al. (2011)	Malaysia	2.04	103.31	N	Austronesian: Malayo-Polynesian	Austric: Austronesian
Dayak Ngaju	nij	Gray et al. (2009)	Indonesia (Northwest Kalimantan)	-1.9	114.5	N	Austronesian: Malayo-Polynesian	Austric: Austronesian
Malagasy	plt	Gray et al. (2009)	Madagascar	-18.9	47.5	N	Austronesian: Malayo-Polynesian	Austric: Austronesian
Melayu	zsm	The HUGO Pan-Asian SNP Consortium	Indonesia	-3.0	104.7	Y	Austronesian: Malayo-Polynesian	Austric: Austronesian
Iban		Xing et al. (2009)	Malaysia (Sarawak)	3	113	N	Austronesian: Malayo-	Austric: Austronesian

							Polynesian	
Singapore Malay	zlm	The HUGO Pan-Asian SNP Consortium	Singapore	1.4	103.8	Y	Austronesian: Malayo-Polynesian	Austric: Austronesian
Malay	zlm	The HUGO Pan-Asian SNP Consortium	Malaysia	5.3	102.0	Y	Austronesian: Malayo-Polynesian	Austric: Austronesian
Javanese	jav	The HUGO Pan-Asian SNP Consortium	Indonesia	-6.2	106.7	Y	Austronesian: Malayo-Polynesian	Austric: Austronesian
Bidayuh	sne	The HUGO Pan-Asian SNP Consortium	Malaysia	1.4	110.2	Y	Austronesian: Malayo-Polynesian	Austric: Austronesian
Mamanwa	mmn	The HUGO Pan-Asian SNP Consortium	Philiippines	9.7	125.6	Y	Austronesian: Malayo-Polynesian	Austric: Austronesian
Agta	agt	The HUGO Pan-Asian SNP Consortium	Philiippines	13.7	123.3	Y	Austronesian: Malayo-Polynesian	Austric: Austronesian
Aeta	sbl	The HUGO Pan-Asian SNP Consortium	Philiippines	14.9	120.2	Y	Austronesian: Malayo-Polynesian	Austric: Austronesian
Andamanese	oon, anq, gac	Rosenberg et al. (2006)	India (Andaman Islands)	10.30	92.30	Y	Andamanese	Indo-Pacific
Naasioi	nas	Rosenberg et al. (2002, 2005); Friedlaender et al. (2008)	Bougainville	-6.483	155.833	Y	South Bougainville	Indo-Pacific: East Papuan
Goroka	for, gim	Rosenberg et al. (2002, 2005); Friedlaender et al. (2008)	New Guinea	-6.083	145.4	Y	Trans-New Guinea	Indo-Pacific
Ngarinyin	ung	Reesink et al. (2009)	Australia	-16.4	126.4	N	Australian	Australian
Warlpiri	wbp	Reesink et al. (2009)	Australia	-23.7	133.9	N	Australian: Pama-Nyungan	Australian
AUSTRALIAN		Rasmussen et al. (2011)	Australia	?	?	Y	Australian	Australian
Manggarai	mqy	The HUGO Pan-Asian SNP Consortium	Indonesia	-8.6	120.1	Y	Austronesian: Malayo-Polynesian	Austric: Austronesian

Kambera	xbr	The HUGO Pan-Asian SNP Consortium	Indonesia	-9.8	120.0	Y	Austronesian: Malayo-Polynesian	Austric: Austronesian
Alorese	aol	The HUGO Pan-Asian SNP Consortium	Indonesia	-8.3	124.7	Y	Austronesian: Malayo-Polynesian	Austric: Austronesian
Lembata	lmf, lmj	The HUGO Pan-Asian SNP Consortium	Indonesia	-8.3	124.7	Y	Austronesian: Malayo-Polynesian	Austric: Austronesian
Kosraean	kos	Friedlaender et al. (2008)	Micronesia	5.31	163	Y	Austronesian: Malayo-Polynesian	Austric: Austronesian
Samoan	smo	Friedlaender et al. (2008)	Polynesia	-13.35	-172.2	Y	Austronesian: Malayo-Polynesian	Austric: Austronesian
Tongan	ton	Gray et al. (2009)	Polynesia	-21.18	-175.20	Y	Austronesian: Malayo-Polynesian	Austric: Austronesian
Hawaiian	haw	Gray et al. (2009)	Polynesia	21.30	-157.85	N	Austronesian: Malayo-Polynesian	Austric: Austronesian
Maori	mri	Friedlaender et al. (2008)	Polynesia	-41	174	Y	Austronesian: Malayo-Polynesian	Austric: Austronesian
Rumai	rbb	The HUGO Pan-Asian SNP Consortium	Thailand	19.9	99.2	Y	Austro-Asiatic: Mon-Khmer	Austric: Austroasiatic
Lawa	lwl, lcp	The HUGO Pan-Asian SNP Consortium	Thailand	18.4	98.1	Y	Austro-Asiatic: Mon-Khmer	Austric: Austroasiatic
Yakut	sah	Rosenberg et al. (2002, 2005); Ramachandran et al. (2005)	Siberia	62.98287845	129.5	Y	Altaic: Turkic	Macro-Altaic
Tuvan	tyv	Martinez-Laso et al. (2001)	Russia (Republic of Tuva)	51.71	94.45	Y	Altaic: Turkic	Macro-Altaic
Nenets	yrk	Wang et al. (2008)	Siberia	66.08	76.5	Y	Uralic	Uralic
Khalkha Mongol	khk, mvn	Rosenberg et al. (2002, 2005); Ramachandran et al.	Mongolia	48	107	Y	Altaic: Mongolic	Macro-Altaic

		(2005); Reich et al. (2012)						
Daur	tdd	Rosenberg et al. (2002, 2005); Ramachandran et al. (2005)	China	48.49753416	124	Y	Altaic: Mongolic	Macro-Altaic
Hezhen	gld	Rosenberg et al. (2002, 2005); Ramachandran et al. (2005)	China	47.4976192	133.5	Y	Altaic: Tungusic	Macro-Altaic
Oroqen	orh	Rosenberg et al. (2002, 2005); Ramachandran et al. (2005)	China	50.43389257	126.5	Y	Altaic: Tungusic	Macro-Altaic
Xibo	sjo	Rosenberg et al. (2002, 2005); Ramachandran et al. (2005)	China	43.49792973	81.5	Y	Altaic: Tungusic	Macro-Altaic
Tu	mjg	Rosenberg et al. (2002, 2005); Ramachandran et al. (2005)	China	36	101	Y	Altaic: Mongolic	Macro-Altaic
Yi	yif	Rosenberg et al. (2002, 2005); Ramachandran et al. (2005)	China	28	103	Y	Sino-Tibetan: Tibeto-Burman	Sino-Tibetan
Naxi	nbf	Rosenberg et al. (2002, 2005); Ramachandran et al. (2005)	China	26	100	Y	Sino-Tibetan: Tibeto-Burman	Sino-Tibetan
Tibetan	bod	Kang et al. (2010)	China	29.65	91.17	Y	Sino-Tibetan: Tibeto-Burman	Sino-Tibetan
Korean	kor	The HUGO Pan-Asian SNP Consortium	Korea	36.9	127.5	Y	Koreanic	Macro-Altaic
Japanese	jpn	HapMap	Japan (Honshu)	35.7	139.8	Y	Japonic	Macro-Altaic

Ainu	ain	Bannai et al. (200)	Japan (Hokkaido)	42.35	142.39	Y	Language isolate	Eurasiatic
Northern Han Chinese	cmn	Rosenberg et al. (2002, 2005); Ramachandran et al. (2005)	China	32.26566812	114	Y	Sino-Tibetan: Chinese	Sino-Tibetan
Min nan Taiwanese'	nan	The HUGO Pan-Asian SNP Consortium	Taiwan	25.0	121.5	Y	Sino-Tibetan: Chinese	Sino-Tibetan
Southern Han Chinese	cmn, yue	The HUGO Pan-Asian SNP Consortium	China	23.3	113.5	Y	Sino-Tibetan: Chinese	Sino-Tibetan
Singapore Chinese	cmn	The HUGO Pan-Asian SNP Consortium	Singapore	1.4	103.8	Y	Sino-Tibetan: Chinese	Sino-Tibetan
Tujia	tji	Rosenberg et al. (2002, 2005); Ramachandran et al. (2005)	China	29	109	Y	Sino-Tibetan: Tibeto-Burman	Sino-Tibetan
She	shx	Rosenberg et al. (2002, 2005); Ramachandran et al. (2005)	China	27	119	Y	Hmong-Mien	Austric: Miao-Yao
Miao	hmy	Rosenberg et al. (2002, 2005); Ramachandran et al. (2005)	China	28	109	Y	Hmong-Mien	Austric: Miao-Yao
Dai	tdd	Rosenberg et al. (2002, 2005); Ramachandran et al. (2005)	China	21	100	Y	Tai-Kadai	Austric: Daic
Lahu	lhu	Rosenberg et al. (2002, 2005); Ramachandran et al. (2005)	China	22	100	Y	Sino-Tibetan: Tibeto-Burman	Sino-Tibetan
Cambodian	khm	Rosenberg et al. (2002, 2005); Ramachandran et al.	Cambodia	12	105	Y	Austro-Asiatic: Mon-Khmer	Austric

		(2005)						
Evenki	evn	Grahovac et al. (1998); Reich et al. (2012)	Siberia (Evenkiysky District)	64.1	95.4	Y	Altaic: Tungusic	Macro-Altaic
Yukaghir	ykg	Reich et al. (2012)	Siberia (Srednekolymsky District)	68	150	Y	Yukaghir	Uralic
Ket	ket	Grahovac et al. (1998); Reich et al. (2012)	Siberia (Evenkiysky District)	63.8	87.4	Y	Yeniseian	Dene-Caucasian: Caucasian
Tlingit	tli	Imanishi et al. (1992)	Canada (British Colombia)	51	-123	N	Eyak-Athabascan	Na-Dene
Nivkh	niv	Grahovac et al. (1998)	Siberia (Sakhalin Island)	52.7	142.8	N	Language isolate	Eurasiatic
Dakelh	crx, bcr	Monsalve et al. (1998)	Canada (British Colombia)	53	-123	N	Eyak-Athabascan: Athabascan	Na-Dene
Koryak	kpy	Grahovac et al. (1998); Reich et al. (2012)	Siberia	59.4	163	Y	Chukotko-Kamchatkan	Chukchi-Kamchatkan
Chukchi	ckt	Grahovac et al. (1998); Reich et al. (2012)	Siberia (Providensky District)	65.1	-173.5	Y	Chukotko-Kamchatkan	Chukchi-Kamchatkan
Siberian Yupik	ess, ynk, ysr	Grahovac et al. (1998)	Siberia (Chukotsky District)	66	-175.2	Y	Eskimo-Aleut	Eskimo-Aleut
Greenland Inuit	kal	Reich et al. (2012)	Greenland	65.3	-52	Y	Eskimo-Aleut	Eskimo-Aleut
Haida	hai (hdn, hax)	Ward et al. (1993)	Canada	53.25	-132	N	Haida	Dene-Caucasian: Na-Dene
Dene	chp	Wang et al. (2008)	Canada	59.55	-107.3	Y	Eyak-Athabascan: Athabascan	Dene-Caucasian: Na-Dene
Navajo	nav	Budowle et al. (2002)	USA (South Dakota)	35.66	-109.07	N	Eyak-Athabascan: Athabascan	Dene-Caucasian: Na-Dene:
Cree	cre	Wang et al. (2008)	Canada	50.33	-102.5	Y	Algic	Amerind: Almosan

Ojibwa	ojc, ojg, ojb, ojs	Wang et al. (2008)	Canada	46.5	-81	Y	Algic	Amerind: Almosan
Mapuche	huh, arn	Wang et al. (2008)	Chile	-41	-73	Y	Mapudungu	Amerind: Andean
Pima	pia	Rosenberg et al. (2002, 2005); Ramachandran et al. (2005)	Mexico	29	-108	Y	Uto-Aztecan	Amerind: Central
Yucatec Maya	yua	Rosenberg et al. (2002, 2005); Ramachandran et al. (2005)	Mexico	19	-91	Y	Mayan	Amerind: Penutian
Lakota	lkt	Leffell et al. (2004)	USA (South Dakota)	35.66	-109.07	N	Siouan-Catawban	Amerind: Keresiouan
Aymara	aym (ayr, ayc)	Wang et al. (2008)	Chile	-22	-70	Y	Aymaran	Amerind: Andean
Quechua	que	Wang et al. (2008)	Peru	-14	-74	Y	Quechuan	Amerind: Andean
Tehuelche	teh	Reich et al. (2012)	Chile	-45	-74	Y	Chon	Amerind: Andean
Yahgan	yag	Reich et al. (2012)	Chile	-55	-68	Y	Language isolate	Amerind: Andean
Cherokee	chr	Malhi et al. (2001)	USA (Oklahoma)	35.81	-94.63	N	Iroquoian	Amerind: Keresiouan
Mixtec	mix	Wang et al. (2008)	Mexico	17	-97	Y	Otomanguean	Amerind: Central
Zapotec	several dialects	Wang et al. (2008)	Mexico	16	-97	Y	Otomanguean	Amerind: Central
Kaingang	xok, zkp	Wang et al. (2008)	Brazil	-24	-52.5	Y	Jean	Amerind: Macro-Ge
Arhuaco	arc	Wang et al. (2008)	Colombia	11	-73.8	Y	Chibchan	Amerind: Chibchan
Kogi	kog	Wang et al. (2008)	Colombia	11	-74	Y	Chibchan	Amerind: Chibchan
Guarani	grn	Wang et al. (2008)	Brazil	-23	-54	Y	Tupian	Amerind: Equatorial
Wichi	mzh, mpt, wlv	Cerna et al. (1993)	Argentina	-31	-59	Y	Matacoan	Amerind: Macro-Panoan
Toba	tob	Cerna et al. (1993)	Argentina	-28	-59	Y	Guaykuruan	Amerind: Macro-Panoan
Piapoco	pio	Rosenberg et al. (2002, 2005);	Colombia	3	-68	Y	Maipurean	Amerind: Equatorial

		Ramachandran et al. (2005)						
Curripaco	kpc	Rosenberg et al. (2002, 2005); Ramachandran et al. (2005)	Colombia	3	-68	Y	Maipurean	Amerind: Equatorial
Karitiana	ktn	Rosenberg et al. (2002, 2005); Ramachandran et al. (2005)	Brazil	-10	-63	Y	Tupian	Amerind: Equatorial
Surui	sru	Rosenberg et al. (2002, 2005); Ramachandran et al. (2005)	Brazil	-11	-62	Y	Tupian	Amerind: Equatorial
Ache	guq	Wang et al. (2008)	Paraguay	-24	-56	Y	Tupian	Amerind: Equatorial
Waiwai	waw	Hutz et al. (2002)	Brazil	0.4	-57.55	N	Cariban	Amerind: Macro-Carib

References

1 Lewis, M., Simons, G. & Fennig, C. *Ethnologue: Languages of the World, Seventeenth Edition*. (SIL international Dallas, TX, 2013).

2 Ruhlen, M. *A guide to the world's languages: classification*. Vol. 1 (Stanford University Press, 1991).

3 Greenberg, J. H. & Ruhlen, M. *An Amerind Etymological Dictionary*. (Stanford University, Department of Anthropological Sciences, Stanford, 2007).

Supplementary Table S3. Resolution (number of nodes), tree length, consistency index (CI), and retention index (RI) for 16 semistrict consensus supertrees based on the sensitivity analysis. (**a**) Representative dataset. (**b**) Representative dataset excluding four wildcard taxa. (**c**) HGDP dataset.

a

	Number of nodes (max 185)	**Tree length**	**CI**	**RI**
1.A	164	11764	0.538	0.838
1.B	164	11820	0.536	0.837
1.C	120	15954	.0397	0.714
1.D	111	18620	0.340	0.635
2.A	149	12520	0.506	0.816
2.B	148	12539	0.505	0.815
2.C	146	12978	0.488	0.802
2.D	146	12872	0.492	0.805
3.A	159	12200	0.519	0.825
3.B	148	13399	0.472	0.790
3.C	148	13399	0.472	0.790
3.D	162	12191	0.519	0.826
4.A	148	12863	0.492	0.806
4.B	148	12863	0.492	0.806
4.C	160	12863	0.507	0.817
4.D	155	12483	0.503	0.814

b

	Number of nodes (max 181)	**Tree length**	**CI**	**RI**
1.A	160	11716	0.537	0.838
1.B	160	11726	0.537	0.838
1.C	152	11921	0.528	0.832
1.D	155	11803	0.533	0.836
2.A	165	11730	0.537	0.838
2.B	164	11749	0.536	0.837
2.C	163	11982	0.525	0.830
2.D	162	11957	0.526	0.831
3.A	165	11940	0.527	0.832
3.B	163	11966	0.526	0.831
3.C	163	11966	0.526	0.831
3.D	164	12006	0.524	0.830
4.A	161	12165	0.517	0.825
4.B	161	12165	0.517	0.825
4.C	161	12156	0.517	0.825
4.D	162	12105	0.520	0.827

c

	Number of nodes (max 54)	**Tree length**	**CI**	**RI**
1.A	50	5953	0.605	0.857
1.B	54	5912	0.609	0.860
1.C	54	5912	0.609	0.860
1.D	54	5887	0.612	0.861
2.A	53	5956	0.605	0.857
2.B	53	5956	0.605	0.857
2.C	53	5956	0.605	0.857
2.D	52	5953	0.605	0.857
3.A	53	5956	0.605	0.857
3.B	53	5956	0.605	0.857
3.C	53	5956	0.605	0.857
3.D	52	5953	0.605	0.857
4.A	53	6035	0.597	0.852
4.B	53	6035	0.597	0.852
4.C	53	6035	0.597	0.852
4.D	53	6052	0.595	0.851

Supplementary Table S4. Similarity of the resulting supertrees based on sensitivity analyses measured by SPR distances. Number of SPR moves and percentage similarity between the two supertrees (in parentheses). (**a**) Representative dataset. (**b**) Representative dataset excluding four wildcard taxa. (**c**) HGDP dataset.

a

	1.A	**1.B**	**1.C**	**1.D**	**2.A**	**2.B**	**2.C**	**2.D**	**3.A**	**3.B**	**3.C**	**3.D**	**4.A**	**4.B**	**4.C**	**4.D**
1.A	-	9 (0.9519)	8 (0.9572)	10 (0.9465)	28 (0.8503)	28 (0.8503)	37 (0.8021)	37 (0.8021)	38 (0.7968)	38 (0.7968)	38 (0.7968)	49 (0.7380)	38 (0.7968)	40 (0.7861)	42 (0.7754)	40 (0.7861)
1.B	9 (0.9519)	-	6 (0.9679)	8 (0.9572)	26 (0.8610)	26 (0.8610)	34 (0.8182)	32 (0.8289)	36 (0.8075)	33 (0.8235)	34 (0.8182)	45 (0.7594)	39 (0.7914)	39 (0.7914)	43 (0.7701)	41 (0.7807)
1.C	8 (0.9572)	6 (0.9679)	-	3 (0.9840)	14 (0.9251)	15 (0.9198)	20 (0.8930)	24 (0.8717)	24 (0.8717)	24 (0.8717)	23 (0.8770)	29 (0.8449)	26 (0.8610)	27 (0.8556)	28 (0.8503)	28 (0.8503)
1.D	10 (0.9465)	8 (0.9572)	3 (0.9840)	-	14 (0.9251)	15 (0.9198)	18 (0.9037)	17 (0.9091)	22 (0.8824)	20 (0.8930)	20 (0.8930)	23 (0.8770)	24 (0.8717)	24 (0.8717)	27 (0.8556)	23 (0.8770)
2.A	28 (0.8503)	26 (0.8610)	14 (0.9251)	14 (0.9251)	-	1 (0.9947)	13 (0.9305)	16 (0.9144)	18 (0.9037)	18 (0.9037)	18 (0.9037)	25 (0.8663)	35 (0.8128)	36 (0.8075)	37 (0.8021)	40 (0.7861)
2.B	28 (0.8503)	26 (0.8610)	15 (0.9198)	15 (0.9198)	1 (0.9947)	-	13 (0.9305)	16 (0.9144)	18 (0.9037)	17 (0.9091)	18 (0.9037)	25 (0.8663)	35 (0.8128)	35 (0.8128)	37 (0.8021)	40 (0.7861)
2.C	37 (0.8021)	34 (0.8182)	20 (0.8930)	18 (0.9037)	13 (0.9305)	13 (0.9305)	-	9 (0.9519)	21 (0.8877)	20 (0.8930)	19 (0.8984)	34 (0.8182)	36 (0.8075)	37 (0.8021)	39 (0.7914)	43 (0.7701)
2.D	37 (0.8021)	32 (0.8289)	24 (0.8717)	17 (0.9091)	16 (0.9144)	16 (0.9144)	9 (0.9519)	-	20 (0.8930)	20 (0.8930)	20 (0.8930)	30 (0.8396)	38 (0.7968)	38 (0.7968)	38 (0.7968)	39 (0.7914)
3.A	38 (0.7968)	36 (0.8075)	24 (0.8717)	22 (0.8824)	18 (0.9037)	18 (0.9037)	21 (0.8877)	20 (0.8930)	-	2 (0.9893)	2 (0.9893)	13 (0.9305)	40 (0.7861)	41 (0.7807)	43 (0.7701)	41 (0.7807)
3.B	38 (0.7968)	33 (0.8235)	24 (0.8717)	20 (0.8930)	18 (0.9037)	17 (0.9091)	20 (0.8930)	20 (0.8930)	2 (0.9893)	-	1 (0.9947)	12 (0.9358)	38 (0.7968)	38 (0.7968)	41 (0.7807)	39 (0.7914)
3.C	38 (0.7968)	34 (0.8182)	23 (0.8770)	20 (0.8930)	18 (0.9037)	18 (0.9037)	19 (0.8984)	20 (0.8930)	2 (0.9893)	1 (0.9947)	-	12 (0.9358)	38 (0.7968)	39 (0.7914)	40 (0.7861)	39 (0.7914)
3.D	49 (0.7380)	45 (0.7594)	29 (0.8449)	23 (0.8770)	25 (0.8663)	25 (0.8663)	34 (0.8182)	30 (0.8396)	13 (0.9305)	12 (0.9358)	12 (0.9358)	-	45 (0.7594)	46 (0.7540)	49 (0.7380)	42 (0.7754)
4.A	38 (0.7968)	39 (0.7914)	26 (0.8610)	24 (0.8717)	35 (0.8128)	35 (0.8128)	36 (0.8075)	38 (0.7968)	40 (0.7861)	38 (0.7968)	38 (0.7968)	45 (0.7594)	-	1 (0.9947)	2 (0.9893)	6 (0.9679)
4.B	40 (0.7861)	39 (0.7914)	27 (0.8556)	24 (0.8717)	36 (0.8075)	35 (0.8128)	37 (0.8021)	38 (0.7968)	41 (0.7807)	38 (0.7968)	39 (0.7914)	46 (0.7540)	1 (0.9947)	-	2 (0.9893)	6 (0.9679)
4.C	42 (0.7754)	43 (0.7701)	28 (0.8503)	27 (0.8556)	37 (0.8021)	37 (0.8021)	39 (0.7914)	38 (0.7968)	43 (0.7701)	41 (0.7807)	40 (0.7861)	49 (0.7380)	2 (0.9893)	2 (0.9893)	-	7 (0.9626)
4.D	40 (0.7861)	41 (0.7807)	28 (0.8503)	23 (0.8770)	40 (0.7861)	40 (0.7861)	43 (0.7701)	39 (0.7914)	41 (0.7807)	39 (0.7914)	39 (0.7914)	42 (0.7754)	6 (0.9679)	6 (0.9679)	7 (0.9626)	-

b

	1.A	1.B	1.C	1.D	2.A	2.B	2.C	2.D	3.A	3.B	3.C	3.D	4.A	4.B	4.C	4.D
1.A	-	8 (0.9563)	11 (0.9399)	14 (0.9235)	27 (0.8525)	28 (0.8470)	44 (0.7596)	46 (0.7486)	35 (0.8087)	38 (0.7923)	38 (0.7923)	45 (0.7541)	40 (0.7814)	41 (0.7760)	41 (0.7760)	41 (0.7760)
1.B	8 (0.9563)	-	12 (0.9344)	15 (0.9180)	25 (0.8634)	25 (0.8634)	41 (0.7760)	39 (0.7869)	35 (0.8087)	34 (0.8142)	35 (0.8087)	43 (0.7650)	41 (0.7760)	41 (0.7760)	42 (0.7705)	42 (0.7705)
1.C	11 (0.9399)	12 (0.9344)	-	3 (0.9836)	30 (0.8361)	31 (0.8306)	36 (0.8033)	38 (0.7923)	30 (0.8361)	30 (0.8361)	29 (0.8415)	36 (0.8033)	36 (0.8033)	37 (0.7978)	36 (0.8033)	36 (0.8033)
1.D	14 (0.9235)	15 (0.9180)	3 (0.9836)	-	30 (0.8361)	31 (0.8306)	38 (0.7923)	34 (0.8142)	33 (0.8197)	31 (0.8306)	31 (0.8306)	34 (0.8142)	39 (0.7869)	40 (0.7814)	40 (0.7814)	36 (0.8033)
2.A	27 (0.8525)	25 (0.8634)	30 (0.8361)	30 (0.8361)	-	1 (0.9945)	25 (0.8634)	24 (0.8689)	17 (0.9071)	17 (0.9071)	17 (0.9071)	22 (0.8798)	40 (0.7814)	41 (0.7760)	41 (0.7760)	43 (0.7650)
2.B	28 (0.8470)	25 (0.8634)	31 (0.8306)	31 (0.8306)	1 (0.9945)	-	25 (0.8634)	24 (0.8689)	17 (0.9071)	16 (0.9126)	17 (0.9071)	22 (0.8798)	40 (0.7814)	40 (0.7814)	41 (0.7760)	43 (0.7650)
2.C	44 (0.7596)	41 (0.7760)	36 (0.8033)	38 (0.7923)	25 (0.8634)	25 (0.8634)	-	5 (0.9727)	22 (0.8798)	21 (0.8852)	20 (0.8907)	33 (0.8197)	43 (0.7650)	44 (0.7596)	43 (0.7650)	44 (0.7596)
2.D	46 (0.7486)	39 (0.7869)	38 (0.7923)	34 (0.8142)	24 (0.8689)	24 (0.8689)	5 (0.9727)	-	25 (0.8634)	24 (0.8689)	24 (0.8689)	33 (0.8197)	49 (0.7322)	50 (0.7268)	50 (0.7268)	47 (0.7432)
3.A	35 (0.8087)	35 (0.8087)	30 (0.8361)	33 (0.8197)	17 (0.9071)	17 (0.9071)	22 (0.8798)	25 (0.8634)	-	2 (0.9891)	2 (0.9891)	11 (0.9399)	40 (0.7814)	41 (0.7760)	41 (0.7760)	40 (0.7814)
3.B	38 (0.7923)	34 (0.8142)	30 (0.8361)	31 (0.8306)	17 (0.9071)	16 (0.9126)	21 (0.8852)	24 (0.8689)	2 (0.9891)	-	1 (0.9945)	10 (0.9454)	40 (0.7814)	40 (0.7814)	41 (0.7760)	40 (0.7814)
3.C	38 (0.7923)	35 (0.8087)	29 (0.8415)	31 (0.8306)	17 (0.9071)	17 (0.9071)	20 (0.8907)	24 (0.8689)	2 (0.9891)	1 (0.9945)	-	10 (0.9454)	40 (0.7814)	41 (0.7760)	40 (0.7814)	40 (0.7814)
3.D	45 (0.7541)	43 (0.7650)	36 (0.8033)	34 (0.8142)	22 (0.8798)	22 (0.8798)	33 (0.8197)	33 (0.8197)	11 (0.9399)	10 (0.9454)	10 (0.9454)	-	44 (0.7596)	45 (0.7541)	45 (0.7541)	42 (0.7705)
4.A	40 (0.7814)	41 (0.7760)	36 (0.8033)	39 (0.7869)	40 (0.7814)	40 (0.7814)	43 (0.7650)	49 (0.7322)	40 (0.7814)	40 (0.7814)	40 (0.7814)	44 (0.7596)	-	1 (0.9945)	1 (0.9945)	4 (0.9781)
4.B	41 (0.7760)	41 (0.7760)	37 (0.7978)	40 (0.7814)	41 (0.7760)	40 (0.7814)	44 (0.7596)	50 (0.7268)	41 (0.7760)	40 (0.7814)	41 (0.7760)	45 (0.7541)	1 (0.9945)	-	1 (0.9945)	4 (0.9781)
4.C	41 (0.7760)	42 (0.7705)	36 (0.8033)	40 (0.7814)	41 (0.7760)	41 (0.7760)	43 (0.7650)	50 (0.7268)	41 (0.7760)	41 (0.7760)	40 (0.7814)	45 (0.7541)	1 (0.9945)	1 (0.9945)	-	4 (0.9781)
4.D	41 (0.7760)	42 (0.7705)	36 (0.8033)	36 (0.8033)	43 (0.7650)	43 (0.7650)	44 (0.7596)	47 (0.7432)	40 (0.7814)	40 (0.7814)	40 (0.7814)	42 (0.7705)	4 (0.9781)	4 (0.9781)	4 (0.9781)	-

C

	1.A	1.B	1.C	1.D	2.A	2.B	2.C	2.D	3.A	3.B	3.C	3.D	4.A	4.B	4.C	4.D
1.A	-	13 (0.7636)	13 (0.7636)	13 (0.7636)	14 (0.7455)	15 (0.7273)	15 (0.7273)	15 (0.7273)	14 (0.7455)	15 (0.7273)	15 (0.7273)	15 (0.7273)	15 (0.7273)	16 (0.7091)	16 (0.7091)	17 (0.6909)
1.B	13 (0.7636)	-	1 (0.9818)	1 (0.9818)	7 (0.8727)	7 (0.8727)	8 (0.8545)	8 (0.8545)	7 (0.8727)	7 (0.8727)	8 (0.8545)	8 (0.8545)	9 (0.8364)	9 (0.8364)	10 (0.8182)	11 (0.8000)
1.C	13 (0.7636)	1 (0.9818)	-	1 (0.9818)	7 (0.8727)	8 (0.8545)	7 (0.8727)	8 (0.8545)	7 (0.8727)	8 (0.8545)	7 (0.8727)	8 (0.8545)	9 (0.8364)	10 (0.8182)	9 (0.8364)	11 (0.8000)
1.D	13 (0.7636)	1 (0.9818)	1 (0.9818)	-	7 (0.8727)	8 (0.8545)	8 (0.8545)	7 (0.8727)	7 (0.8727)	8 (0.8545)	8 (0.8545)	7 (0.8727)	9 (0.8364)	10 (0.8182)	10 (0.8182)	10 (0.8182)
2.A	14 (0.7455)	7 (0.8727)	7 (0.8727)	7 (0.8727)	-	1 (0.9818)	1 (0.9818)	1 (0.9818)	0 (1.0000)	1 (0.9818)	1 (0.9818)	1 (0.9818)	5 (0.9091)	6 (0.8909)	6 (0.8909)	7 (0.8727)
2.B	15 (0.7273)	7 (0.8727)	8 (0.8545)	8 (0.8545)	1 (0.9818)	-	1 (0.9818)	1 (0.9818)	0 (1.0000)	0 (1.0000)	1 (0.9818)	1 (0.9818)	5 (0.9091)	5 (0.9091)	6 (0.8909)	7 (0.8727)
2.C	15 (0.7273)	8 (0.8545)	7 (0.8727)	8 (0.8545)	1 (0.9818)	1 (0.9818)	-	1 (0.9818)	0 (1.0000)	1 (0.9818)	0 (1.0000)	1 (0.9818)	5 (0.9091)	6 (0.8909)	5 (0.9091)	7 (0.8727)
2.D	15 (0.7273)	8 (0.8545)	8 (0.8545)	7 (0.8727)	1 (0.9818)	1 (0.9818)	1 (0.9818)	-	0 (1.0000)	1 (0.9818)	1 (0.9818)	0 (1.0000)	5 (0.9091)	6 (0.8909)	6 (0.8909)	6 (0.8909)
3.A	14 (0.7455)	7 (0.8727)	7 (0.8727)	7 (0.8727)	0 (1.0000)	0 (1.0000)	0 (1.0000)	0 (1.0000)	-	1 (0.9818)	1 (0.9818)	1 (0.9818)	5 (0.9091)	6 (0.8909)	6 (0.8909)	7 (0.8727)
3.B	15 (0.7273)	7 (0.8727)	8 (0.8545)	8 (0.8545)	1 (0.9818)	0 (1.0000)	1 (0.9818)	1 (0.9818)	1 (0.9818)	-	1 (0.9818)	1 (0.9818)	5 (0.9091)	5 (0.9091)	6 (0.8909)	7 (0.8727)
3.C	15 (0.7273)	8 (0.8545)	7 (0.8727)	8 (0.8545)	1 (0.9818)	1 (0.9818)	0 (1.0000)	1 (0.9818)	1 (0.9818)	1 (0.9818)	-	1 (0.9818)	5 (0.9091)	6 (0.8909)	5 (0.9091)	7 (0.8727)
3.D	15 (0.7273)	8 (0.8545)	8 (0.8545)	7 (0.8727)	1 (0.9818)	1 (0.9818)	1 (0.9818)	0 (1.0000)	1 (0.9818)	1 (0.9818)	1 (0.9818)	-	5 (0.9091)	6 (0.8909)	6 (0.8909)	6 (0.8909)
4.A	15 (0.7273)	9 (0.8364)	9 (0.8364)	9 (0.8364)	5 (0.9091)	5 (0.9091)	5 (0.9091)	5 (0.9091)	5 (0.9091)	5 (0.9091)	5 (0.9091)	5 (0.9091)	-	1 (0.9818)	1 (0.9818)	2 (0.9636)
4.B	16 (0.7091)	9 (0.8364)	10 (0.8182)	10 (0.8182)	6 (0.8909)	5 (0.9091)	6 (0.8909)	6 (0.8909)	6 (0.8909)	5 (0.9091)	6 (0.8909)	6 (0.8909)	1 (0.9818)	-	1 (0.9818)	2 (0.9636)
4.C	16 (0.7091)	10 (0.8182)	9 (0.8364)	10 (0.8182)	6 (0.8909)	6 (0.8909)	5 (0.9091)	6 (0.8909)	6 (0.8909)	6 (0.8909)	5 (0.9091)	6 (0.8909)	1 (0.9818)	1 (0.9818)	-	2 (0.9636)
4.D	17 (0.6909)	11 (0.8000)	11 (0.8000)	10 (0.8182)	7 (0.8727)	7 (0.8727)	7 (0.8727)	6 (0.8909)	7 (0.8727)	7 (0.8727)	7 (0.8727)	6 (0.8909)	2 (0.9636)	2 (0.9636)	2 (0.9636)	-

Supplementary Table S5. (**a**) Wildcard taxa identified in each parameter set of the sensitivity analysis of the representative dataset, with the number of additional nodes of the consensus gained by their exclusion (in parenthesis). Wildcard taxa decreasing the resolution of the consensus by 5 nodes or more are in bold. (**b**) Sum of wildcard taxa identified in each parameter set and the sum of additional nodes of the consensus gained by their exclusion. (**c**) Wildcard taxa identified in all parameters of the sensitivity analysis, the sum of parameter sets in which these taxa acted as wildcard taxa and the sum of additional nodes of the consensus gained by their exclusion.

a

	A	**B**	**C**	**D**
1	Amhara (1) Aari (1) Alorese (1) Dogon (1) Egyptian (1) Georgian (2) Guarani, Toba, Wichi (1) Mozabite (1)	Aari (1) Amhara (1) Andamanese (2) Burusho (1) Dogon (1) Egyptian (1) Guarani, Toba, Wichi (1) Mozabite (1) Yahgan, Tehuelche (3)	Ache (1) **Andamanese (21)** Armenian (1) Egyptian (1) Georgian (2) **Malagasy (7)** Singapore Malay (1) Waiwai (1) Yahgan, Tehuelche (3)	Ache (1) **Andamanese (5)** Armenian (1) **Dayak Ngaju (10)** Egyptian (1) **Malagasy (20)** Mapuche, Yahgan, Tehuelche (4) Singapore Malay (1) Waiwai (1)
2	Amhara (1) Andamanese (3) Dogon (2) Guarani, Toba, Wichi (1) Cherokee (1) **Malagasy (23)**	Amhara (1) Andamanese (2) Dogon (2) Guarani, Toba, Wichi (1) Cherokee (1) **Malagasy (23)**	Ache (1) Cherokee (1) Malagasy (3) Nenets (1) **Qatari (21)** Waiwai (2)	Ache (1) Amhara (1) Andamanese (1) Dogon (2) Nenets (1) **Qatari (21)** Waiwai (2)
3	**Andamanese (7)** Cherokee (1) Guarani, Toba, Wichi (1) **Malagasy (5)** Nenets (1) Orcadian (2)	**Andamanese (14)** Cherokee (1) Guarani, Toba, Wichi (1) Malagasy (3) Nenets (1) Orcadian (2)	**Andamanese (14)** Cherokee (1) Guarani, Toba, Wichi (1) Malagasy (3) Nenets (1) Orcadian (2)	Guarani, Toba, Wichi (1) Andamanese (1) **Malagasy (5)** Nenets (1)
4	**Andamanese (7)** Cherokee (1) **Malagasy (9)** Tibetan (2) Toba, Wichi (1) Yahgan (1)	**Andamanese (7)** Cherokee (1) **Malagasy (9)** Tibetan (2) Toba, Wichi (1) Yahgan (1)	Amhara (1) **Andamanese (7)** Bidayuh (1) Dogon (2) Cherokee (1) Mozabite (1) Tibetan (2) Toba, Wichi (1) Yahgan (1)	Andamanese (1) Cherokee (1) Dayak Ngaju (1) **Malagasy (9)** Tibetan (2) Toba, Wichi (1) Yahgan (1)

b

	A	B	C	D	Sum
1	8/9	9/12	9/38	9/44	35/103
2	6/31	6/30	6/29	7/29	25/119
3	6/17	6/22	6/22	4/9	22/70
4	6/21	6/21	9/17	7/16	28/75
Sum	26/78	27/85	30/106	27/98	

c

	Sum of parameter sets	Sum of nodes
Aari	2	4
Ache	4	4
Alorese	1	1
Amhara	6	6
Andamanese	14	92
Armenian	2	2
Burusho	1	1
Dayak Ngaju	2	11
Dogon	6	10
Egyptian	4	4
Georgian	2	4
Guarani, Toba, Wichi	8	8
Cherokee	10	10
Malagasy	12	119
Mapuche, Yahgan, Tehuelche	1	4
Mozabite	3	3
Nenets	6	6
Orcadian	3	6
Qatari	2	42
Singapore_Malay	2	2
Tibetan	4	8
Toba, Wichi	4	1
Waiwai	4	7
Yahgan, Tehuelche	2	6

Supplementary Table S6. List of characters included in the linguistic classification datasets. Characters that are phylogenetically informative for 186 taxa in the representative dataset are in black. Characters that are phylogenetically informative for 186 taxa in the representative dataset are in gray. Characters not used to constrain the topology of the language-constrained supertree (Fig. 3) are in red**.** (**a**) *Ethnologue* dataset. (**b**) Greenberg–Ruhlen dataset.

a) *Ethnologue* dataset

[1] Khoisan
[2] Khoisan: Hatsa
[3] Khoisan: Sandawe
[4] Khoisan: Southern Africa
[5] Niger-Congo
[6] Niger-Congo: Mande
[7] Niger-Congo: Atlantic-Congo
[8] Niger-Congo: Atlantic-Congo: Atlantic
[9] Niger-Congo: Atlantic-Congo: Volta-Congo: Dogon
[10] Niger-Congo: Atlantic-Congo: Volta-Congo: Kwa
[11] Niger-Congo: Atlantic-Congo: Volta-Congo: North
[12] Niger-Congo: Atlantic-Congo: Volta-Congo: Benue-Congo
[13] Niger-Congo: Atlantic-Congo: Volta-Congo: Benue-Congo: Bantoid
[14] Niger-Congo: Atlantic-Congo: Volta-Congo: Benue-Congo: Nupoid
[15] Niger-Congo: Atlantic-Congo: Volta-Congo: Benue-Congo: Defoid
[16] Niger-Congo: Atlantic-Congo: Volta-Congo: Benue-Congo: Igboid
[17] Nilo-Saharan
[18] Afro-Asiatic
[19] Afro-Asiatic: Berber
[20] Afro-Asiatic: Chadic
[21] Afro-Asiatic: Cushtic
[22] Afro-Asiatic: Omotic
[23] Afro-Asiatic: Semitic
[24] Dravidian
[25] Kartvelian
[26] North Caucasian
[27] Indo-European
[28] Uralic
[29] Yukaghir
[30] Chukotko-Kamchatkan
[31] Eskimo-Aleut
[32] Altaic
[33] Altaic: Turkic
[34] Altaic: Mongolic

[35] Altaic: Tungusic
[36] Japonic
[37] Sino-Tibetan
[38] Sino-Tibetan: Chinese
[39] Sino-Tibetan: Tibeto-Burman
[40] Yeniseian
[41] Haida
[42] Eyak-Athabascan
[43] Eyak-Athabascan: Athabascan
[44] Hmong-Mien
[45] Tai-Kadai
[46] Austro-Asiatic
[47] Austro-Asiatic: Mon-Khmer
[48] Austro-Asiatic: Munda
[49] Austronesian
[50] Austronesian: Malayo-Polynesian
[51] Austronesian: Malayo-Polynesian: Oceanic
[52] Andamanese
[53] Yele-West New Britain
[54] East New Britain
[55] North Bougainville
[56] South Bougainville
[57] Central Solomons
[58] Torricelli
[59] Sepik
[60] Trans-New Guinea
[61] Australian
[62] Australian: Daly
[63] Australian: Gunwingguan
[64] Australian: Pama-Nyungan
[65] Algic
[66] Salish
[67] Yokutsan
[68] Muskogean
[69] Wakashan
[70] Siouan-Catawban
[71] Totonacan
[72] Pomoan
[73] Iroquoian
[74] Mayan
[75] Otomanguean
[76] Mixe-Zoquean
[77] Uto-Aztecan
[78] Chibchan

[79] Yanomaman
[80] Tarascan
[81] Paezan
[82] Barbacoan
[83] Chocoan
[84] Aymaran
[85] Quechuan
[86] Mapudungu
[87] Kaweskaran
[88] Chon
[89] Maipurean
[90] Guajiboan
[91] Jivaroan
[92] Chapacuran
[93] Salivan
[94] Mascoyan
[95] Arauan
[96] Zamucoan
[97] Chipaya-Uru
[98] Cariban
[99] Tupian
[100] Jean
[101] Tucanoan
[102] Puinavean
[103] Matacoan
[104] Guaykuruan
[105] Mosetenan

b) Greenberg-Ruhlen dataset

[1] Khoisan
[2] Niger-Kordofanian
[3] Nilo-Saharan
[4] Afro-Asiatic
[5] Dravidian
[6] Caucasian
[6] Kartvelian
[7] Indo-Hittite
[8] Uralic
[9] Chukchi-Kamchatkan
[10] Eskimo-Aleut
[11] Macro-Altaic

[12] Sino-Tibetan
[13] Na-Dene
[14] Austric
[15] Austric: Austroasiatic
[16] Austric: Miao-Yao
[17] Austric: Daic
[18] Austric: Austronesian
[19] Indo-Pacific
[20] Indo-Pacific: Kusunda
[21] Indo-Pacific: Andaman Islands
[22] Indo-Pacific: East Papuan
[23] Indo-Pacific: Toricelli
[24] Indo-Pacific: Sepik-Ramu
[25] Indo-Pacific: Trans-New Guinea
[26] Australian
[27] Australian: Gunwinyguan
[28] Australian: Wororan
[29] Australian: Pama-Nyungan
[30] Amerind
[31] Amerind: Almosan
[32] Amerind: Keresiouan
[33] Amerind: Penutian
[34] Amerind: Hokan
[35] Amerind: Central
[36] Amerind: Chibchan
[37] Amerind: Paezan
[38] Amerind: Andean
[39] Amerind: Equatorial
[40] Amerind: Macro-Tucanoan
[41] Amerind: Macro-Ge
[42] Amerind: Macro-Panoan
[43] Amerind: Macro-Carib

Supplementary Table S7. Consistency index (CI), retention index (RI), and normalized CI and RI values for linguistic groupings resulting from the optimization of the linguistic classification datasets on the topology of the purely genetic and combined supertree (based on parameter set 1.A of the sensitivity analysis). (**a**) *Ethnologue* dataset. (**b**) Greenberg–Ruhlen dataset. (**c**) Additional linguistic macrofamilies accepted from Ruhlen[1].

a

		c.i.		**r.i.**		**normalized c.i.**		**normalized r.i.**	
Linguistic grouping	**N taxa**	**Genetic tree**	**Combined tree**	**Genetic tree**	**Combined tree**	**Genetic tree**	**Combined tree**	**Genetic tree**	**Combined tree**
Khoisan	7	0.33333333	0.33333333	0.66666667	0.66666667	2.33333331	2.33333331	4.66666669	4.66666669
Khoisan: Southern Africa	5	1	1	1	1	5	5	5	5
Niger-Congo	10	0.16666667	0.16666667	0.44444444	0.44444444	1.6666667	1.6666667	4.4444444	4.4444444
Niger-Congo: Atlantic-Congo	9	0.16666667	0.16666667	0.375	0.375	1.50000003	1.50000003	3.375	3.375
Niger-Congo: Atlantic-Congo: Volta-Congo: Benue-Congo	6	0.33333333	0.33333333	0.6	0.6	1.99999998	1.99999998	3.6	3.6
Niger-Congo: Atlantic-Congo: Volta-Congo: Benue-Congo: Bantoid	6	0.33333333	0.33333333	0.6	0.6	1.99999998	1.99999998	3.6	3.6
Nilo-Saharan	5	0.25	0.25	0.25	0.25	1.25	1.25	1.25	1.25
Afro-Asiatic	28	0.11111111	0.16666667	0.7037037	0.81481481	3.11111108	4.66666676	19.7037036	22.81481468
Afro-Asiatic: Berber	2	0.5	0.5	0	0	1	1	0	0
Afro-Asiatic: Chadic	4	0.5	0.5	0.66666667	0.66666667	2	2	2.66666668	2.66666668
Afro-Asiatic: Cushtic	7	0.16666667	0.25	0.16666667	0.5	1.16666669	1.75	1.16666669	3.5
Afro-Asiatic: Semitic	14	0.16666667	0.2	0.61538462	0.69230769	2.33333338	2.8	8.61538468	9.69230766
Dravidian	4	0.5	0.5	0.66666667	0.66666667	2	2	2.66666668	2.66666668
Indo-European	22	0.11111111	0.125	0.61904762	0.66666667	2.44444442	2.75	13.61904764	14.66666674
Uralic	4	0.33333333	0.33333333	0.33333333	0.33333333	1.33333332	1.33333332	1.33333332	1.33333332
Chukotko-Kamchatkan	2	0.5	0.5	0	0	1	1	0	0
Eskimo-Aleut	2	0.5	0.5	0	0	1	1	0	0
Altaic	11	0.16666667	0.16666667	0.5	0.5	1.83333337	1.83333337	5.5	5.5
Altaic: Turkic	4	0.25	0.25	0	0	1	1	0	0
Altaic: Mongolic	3	0.33333333	0.33333333	0	0	0.99999999	0.99999999	0	0
Altaic: Tungusic	4	0.33333333	0.33333333	0.33333333	0.33333333	1.33333332	1.33333332	1.33333332	1.33333332
Sino-Tibetan	9	0.25	0.33333333	0.625	0.75	2.25	2.99999997	5.625	6.75
Sino-Tibetan: Chinese	4	1	1	1	1	4	4	4	4
Sino-Tibetan: Tibeto-Burman	5	0.33333333	0.33333333	0.5	0.5	1.66666665	1.66666665	2.5	2.5
Eyak-Athabascan	4	0.25	0.25	0	0	1	1	0	0
Eyak-Athabascan: Athabascan	3	0.33333333	0.33333333	0	0	0.99999999	0.99999999	0	0
Hmong-Mien	2	0.5	0.5	0	0	1	1	0	0
Austro-Asiatic	5	0.33333333	0.33333333	0.5	0.5	1.66666665	1.66666665	2.5	2.5

Austro-Asiatic: Mon-Khmer	5	0.33333333	0.33333333	0.5	0.5	1.66666665	1.66666665	2.5	2.5
Austronesian	28	0.2	0.25	0.85185185	0.88888889	5.6	7	23.8518518	24.88888892
Austronesian: Malayo-Polynesian	24	0.11111111	0.25	0.65217391	0.86956522	2.66666664	6	15.65217384	20.86956528
Austronesian: Malayo-Polynesian: Oceanic	5	0.5	0.33333333	0.75	0.5	2.5	1.66666665	3.75	2.5
Australian	3	0.5	0.33333333	0.5	0	1.5	0.99999999	1.5	0
Algic	2	1	1	1	1	2	2	2	2
Chibchan	2	1	1	1	1	2	2	2	2
Maipurean	2	1	1	1	1	2	2	2	2
Tupian	4	0.5	0.5	0.66666667	0.66666667	2	2	2.66666668	2.66666668

b

		c.i.		r.i.		rescaled c.i.		rescaled r.i.	
Linguistic grouping	**N taxa**	**Genetic tree**	**Combined tree**	**Genetic tree**	**Combined tree**	**Genetic tree**	**Combined tree**	**Genetic tree**	**Combined tree**
Khoisan	7	0.33333333	0.33333333	0.66666667	0.66666667	2.33333331	2.33333331	4.66666669	4.66666669
Niger-Kordofanian	10	0.16666667	0.16666667	0.44444444	0.44444444	1.6666667	1.6666667	4.4444444	4.4444444
Nilo-Saharan	5	0.25	0.25	0.25	0.25	1.25	1.25	1.25	1.25
Afro-Asiatic	28	0.11111111	0.16666667	0.7037037	0.81481481	3.11111108	4.66666676	19.7037036	22.81481468
Dravidian	4	0.5	0.5	0.66666667	0.66666667	2	2	2.66666668	2.66666668
Indo-Hittite	22	0.1	0.125	0.57142857	0.66666667	2.2	2.75	12.57142854	14.66666674
Uralic	5	0.25	0.25	0.25	0.25	1.25	1.25	1.25	1.25
Chukchi-Kamchatkan	2	0.5	0.5	0	0	1	1	0	0
Eskimo-Aleut	2	0.5	0.5	0	0	1	1	0	0
Macro-Altaic	13	0.14285714	0.14285714	0.5	0.5	1.85714282	1.85714282	6.5	6.5
Sino-Tibetan	9	0.25	0.25	0.625	0.625	2.25	2.25	5.625	5.625
Na-Dene	5	0.2	0.2	0	0	1	1	0	0
Austric	36	0.1	0.1	0.74285714	0.77142857	3.6	3.6	26.74285704	27.77142852
Austric: Austroasiatic	5	0.5	0.5	0.66666667	0.66666667	2.5	2.5	3.33333335	3.33333335
Austric: Miao-Yao	2	0.5	0.5	0	0	1	1	0	0
Austric: Austronesian	28	0.14285714	0.16666667	0.77777778	0.81481481	3.99999992	4.66666676	21.77777784	22.81481468
Indo-Pacific	3	0.33333333	0.33333333	0	0	0.99999999	0.99999999	0	0
Australian	3	0.5	0.33333333	0.5	0	1.5	0.99999999	1.5	0
Amerind	25	0.5	0.5	0.95833333	0.95833333	12.5	12.5	23.95833325	23.95833325
Amerind: Almosan	2	1	1	1	1	2	2	2	2
Amerind: Keresiouan	2	0.5	0.5	0	0	1	1	0	0
Amerind: Central	3	0.33333333	0.33333333	0	0	0.99999999	0.99999999	0	0
Amerind: Chibchan	2	1	1	1	1	2	2	2	2
Amerind: Andean	4	0.25	0.25	0.25	0.25	1	1	1	1
Amerind: Equatorial	6	0.33333333	0.33333333	0.6	0.6	1.99999998	1.99999998	3.6	3.6
Amerind: Macro-Panoan	2	1	1	1	1	2	2	2	2

c

		c.i.		**r.i.**		**rescaled c.i.**		**rescaled r.i.**	
Linguistic grouping	**N taxa**	**Genetic tree**	**Combined tree**	**Genetic tree**	**Combined tree**	**Genetic tree**	**Combined tree**	**Genetic tree**	**Combined tree**
Eurasiatic	46	0.05882353	0.0625	0.64444444	0.66666667	2.70588238	2.875	29.64444424	30.66666682
Nostratic (incl. Afro-Asiatic and Dravidian)	76	0.0625	0.07692308	0.8	0.84	4.75	5.84615408	60.8	63.84
Na-Dene (incl. Yeniseian)	6	0.16666667	0.16666667	0	0	1.00000002	1.00000002	0	0
Dene-Caucasian	36	0.07142857	0.07142857	0.27777778	0.27777778	2.57142852	2.57142852	10.00000008	10.00000008

Supplementary Table S8. Supertree dataset references.

1 Abbi, A. Is Great Andamanese genealogically and typologically distinct from Onge and Jarawa? *Language Sciences* **31**, 791-812, doi:10.1016/j.langsci.2008.02.002 (2009).

2 Abdulla, M. A. *et al.* Mapping human genetic diversity in Asia. *Science* **326**, 1541-1545, doi:10.1126/science.1177074 (2009).

3 Agrawal, S., Srivastava, S. K., Borkar, M. & Chaudhuri, T. K. Genetic affinities of north and northeastern populations of India: inference from HLA-based study. *Tissue Antigens* **72**, 120-130, doi:10.1111/j.1399-0039.2008.01083.x (2008).

4 Alfonso-Sanchez, M. A., Perez-Miranda, A. M. & Herrera, R. J. Autosomal microsatellite variability of the Arrernte people of Australia. *American Journal of Human Biology* **20**, 91-99, doi:10.1002/ajhb.20685 (2008).

5 Alkorta-Aranburu, G. *et al.* The genetic architecture of adaptations to high altitude in Ethiopia. *PLOS Genetics* **8**, doi:10.1371/journal.pgen.1003110 (2012).

6 Arnaiz-Villena, A. *et al.* HLA genes in Macedonians and the sub-Saharan origin of the Greeks. *Tissue Antigens* **57**, 118-127, doi:10.1034/j.1399-0039.2001.057002118.x (2001).

7 Arnaiz-Villena, A. *et al.* HLA genes in Uros from Titikaka Lake, Peru: origin and relationship with other Amerindians and worldwide populations. *International Journal of Immunogenetics* **36**, 159-167, doi:10.1111/j.1744-313X.2009.00841.x (2009).

8 Arnaiz-Villena, A. *et al.* The origin of Cretan populations as determined by characterization of HLA alleles. *Tissue Antigens* **53**, 213-226, doi:10.1034/j.1399-0039.1999.530301.x (1999).

9 Arnaiz-Villena, A. *et al.* HLA alleles and haplotypes in the Turkish population: relatedness to Kurds, Armenians and other Mediterraneans. *Tissue Antigens* **57**, 308-317, doi:10.1034/j.1399-0039.2001.057004308.x (2001).

10 Arnaiz-Villena, A. *et al.* Relatedness among Basques, Portuguese, Spaniards, and Algerians studied by HLA allelic frequencies and haplotypes. *Immunogenetics* **47**, 37-43, doi:10.1007/s002510050324 (1997).

11 Arnaiz-Villena, A. *et al.* HLA genes in the chuvashian population from European Russia: admixture of central European and Mediterranean populations. *Human Biology* **75**, 375-392, doi:10.1353/hub.2003.0040 (2003).

12 Arnaiz-Villena, A. *et al.* HLA genes in Mayos population from northeast Mexico. *Current Genomics* **8**, 466-475 (2007).

13 Arnaiz-Villena, A. *et al.* The origin of Amerindians and the peopling of the Americas according to HLA genes: admixture with Asian and Pacific people. *Current Genomics* **11**, 103-114, doi:10.2174/138920210790886862 (2010).

14 Arnaiz-Villena, A. *et al.* Origin of Aymaras from Bolivia and their relationship with other Amerindians according to HLA genes. *Tissue Antigens* **65**, 379-390, doi:10.1111/j.1399-0039.2005.00356.x (2005).

15 Arnaiz-Villena, A. *et al.* HLA genes in Mexican Mazatecans, the peopling of the Americas and the uniqueness of Amerindians. *Tissue Antigens* **56**, 405-416,

doi:10.1034/j.1399-0039.2000.560503.x (2000).

16 Atzmon, G. *et al.* Abraham's children in the genome era: major Jewish diaspora populations comprise distinct genetic clusters with shared Middle Eastern Ancestry. *American Journal of Human Genetics* **86**, 850-859, doi:10.1016/j.ajhg.2010.04.015 (2010).

17 Ayub, Q. *et al.* Reconstruction of human evolutionary tree using polymorphic autosomal microsatellites. *American Journal of Physical Anthropology* **122**, 259-268, doi:10.1002/ajpa.10234 (2003).

18 Balanovsky, O. *et al.* Parallel evolution of genes and languages in the Caucasus region. *Molecular Biology and Evolution* **28**, 2905-2920, doi:10.1093/molbev/msr126 (2011).

19 Bannai, M. *et al.* Analysis of HLA genes and haplotypes in Ainu (from Hokkaido, northern Japan) supports the premise that they descent from Upper Paleolithic populations of East Asia. *Tissue Antigens* **55**, 128-139, doi:10.1034/j.1399-0039.2000.550204.x (2000).

20 Barrantes, R. *et al.* Microevolution in lower Central America: genetic characterization of the Chibcha-speaking groups of Costa Rica and Panama, and a consensus taxonomy based on genetic and linguistic affinity. *American Journal of Human Genetics* **46**, 63-84 (1990).

21 Behar, D. M. *et al.* The genome-wide structure of the Jewish people. *Nature* **466**, 238-U112, doi:10.1038/nature09103 (2010).

22 Bharadwaj, U., Khan, F., Srivastava, S., Goel, H. & Agrawal, S. Phylogenetic applications of HLA class II loci. *International Journal of Human Genetics* **7**, 123 (2007).

23 Bieber, H., Bieber, S. W., Rodewald, A. & Barrantes, R. Microevolution and genetic affinities among six Amerindian tribes of lower Central America: comparative genetic study of serum proteins. *Human Biology* **68**, 929-953 (1996).

24 Biro, A. Z., Zalan, A., Volgyi, A. & Pamjav, H. A Y-Chromosomal Comparison of the Madjars (Kazakhstan) and the Magyars (Hungary). *American Journal of Physical Anthropology* **139**, 305-310, doi:10.1002/ajpa.20984 (2009).

25 Blagitko, N. *et al.* Polymorphism of the HLA-DRB1 locus in Colombian, Ecuadorian, and Chilean Amerinds. *Human Immunology* **54**, 74-81, doi:10.1016/s0198-8859(97)00005-0 (1997).

26 Bouckaert, R. *et al.* Mapping the origins and expansion of the Indo-European language family. *Science* **337**, 957-960, doi:10.1126/science.1219669 (2012).

27 Bruges-Armas, J. *et al.* HLA in the Azores Archipelago: possible presence of Mongoloid genes. *Tissue Antigens* **54**, 349-359, doi:10.1034/j.1399-0039.1999.540404.x (1999).

28 Bryc, K. *et al.* Genome-wide patterns of population structure and admixture in West Africans and African Americans. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 786-791, doi:10.1073/pnas.0909559107 (2010).

29 Callegari-Jacques, S. M. *et al.* Autosome STRs in native South America—testing models of association with geography and language. *American Journal of Physical Anthropology* **145**, 371-381, doi:10.1002/ajpa.21505 (2011).

30 Cavalli-Sforza, L. L., Menozzi, P. & Piazza, A. *The History and Geography of Human Genes* (Princeton university press, 1994).

31 Comas, D. *et al.* Trading genes along the silk road: mtDNA sequences and the origin of central Asian populations. *American Journal of Human Genetics* **63**, 1824-1838, doi:10.1086/302133 (1998).

32 Corella, A., Bert, F., Perez-Perez, A., Gene, M. & Turbon, D. Mitochondrial DNA diversity of the Amerindian populations living in the Andean Piedmont of Bolivia: Chimane, Moseten, Aymara and Quechua. *Annals of Human Biology* **34**, 34-55, doi:10.1080/03014460601075819 (2007).

33 Crawford, M. H., Rubicz, R. C. & Zlojutro, M. Origins of Aleuts and the genetic structure of populations of the archipelago: molecular and archaeological perspectives. *Human Biology* **82**, 695-717 (2010).

34 Currie, T. E., Greenhill, S. J., Gray, R. D., Hasegawa, T. & Mace, R. Rise and fall of political complexity in island South-East Asia and the Pacific. *Nature* **467**, 801-804, doi:10.1038/nature09461 (2010).

35 Currie, T. E., Meade, A., Guillon, M. & Mace, R. Cultural phylogeography of the Bantu Languages of sub-Saharan Africa. *Proceedings of the Royal Society B-Biological Sciences* **280**, 8, doi:10.1098/rspb.2013.0695 (2013).

36 Deka, R. *et al.* Genetic variation at twentythree microsatellite loci in sixteen human populations. *Journal of Genetics* **78**, 99-121, doi:10.1007/bf02924561 (1999).

37 Delmestri, A. & Cristianini, N. Linguistic phylogenetic inference by PAM-like matrices. *Journal of Quantitative Linguistics* **19**, 95-120, doi:10.1080/09296174.2012.659001 (2012).

38 Dunn, M. Contact and phylogeny in Island Melanesia. *Lingua* **119**, 1664-1678, doi:10.1016/j.lingua.2007.10.026 (2009).

39 Dunn, M., Greenhill, S. J., Levinson, S. C. & Gray, R. D. Evolved structure of language shows lineage-specific trends in word-order universals. *Nature* **473**, 79-82, doi:10.1038/nature09923 (2011).

40 Dunn, M., Terrill, A., Reesink, G., Foley, R. A. & Levinson, S. C. Structural phylogenetics and the reconstruction of ancient language history. *Science* **309**, 2072-2075, doi:10.1126/science.1114615 (2005).

41 Edinur, H. A. *et al.* HLA polymorphism in six Malay subethnic groups in Malaysia. *Human Immunology* **70**, 518-526, doi:10.1016/j.humimm.2009.04.003 (2009).

42 Eshleman, J. A. *et al.* Mitochondrial DNA and prehistoric settlements: native migrations on the western edge of North America. *Human Biology* **76**, 55-75, doi:10.1353/hub.2004.0019 (2004).

43 Fadhlaoui-Zid, K. *et al.* Genome-wide and paternal diversity reveal a recent origin of human populations in North Africa. *PLOS ONE* **8**, doi:10.1371/journal.pone.0080293 (2013).

44 Farjadian, S. & Ghaderi, A. HLA class II similarities in Iranian Kurds and Azeris. *International Journal of Immunogenetics* **34**, 457-463, doi:10.1111/j.1744-313X.2007.00723.x (2007).

45 Farjadian, S., Moqadam, F. A. & Ghaderi, A. HLA class II gene polymorphism in Parsees and Zoroastrians of Iran. *International Journal of Immunogenetics* **33**, 185-

191, doi:10.1111/j.1744-313X.2006.00594.x (2006).

46 Farjadian, S. *et al.* Molecular analysis of HLA allele frequencies and haplotypes in Baloch of Iran compared with related populations of Pakistan. *Tissue Antigens* **64**, 581-587, doi:10.1111/j.1399-0039.2004.00302.x (2004).

47 Farjadian, S., Ota, M., Inoko, H. & Ghaderi, A. The genetic relationship among Iranian ethnic groups: an anthropological view based on HLA class II gene polymorphism. *Molecular Biology Reports* **36**, 1943-1950, doi:10.1007/s11033-008-9403-4 (2009).

48 Fedorova, S. A. *et al.* Autosomal and uniparental portraits of the native populations of Sakha (Yakutia): implications for the peopling of Northeast Eurasia. *BMC Evolutionary Biology* **13**, doi:10.1186/1471-2148-13-127 (2013).

49 Firasat, S. *et al.* Y-chromosomal evidence for a limited Greek contribution to the Pathan population of Pakistan. *European Journal of Human Genetics* **15**, 121-126, doi:10.1038/sj.ejhg.5201726 (2007).

50 Fortunato, L., Holden, C. & Mace, R. From bridewealth to dowry? *Human Nature* **17**, 355-376, doi:10.1007/s12110-006-1000-4 (2006).

51 Friedlaender, J. S. *et al.* The genetic structure of Pacific islanders. *PLOS Genetics* **4**, doi:10.1371/journal.pgen.0040019 (2008).

52 Fu, Q. M. *et al.* DNA analysis of an early modern human from Tianyuan Cave, China. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 2223-2227, doi:10.1073/pnas.1221359110 (2013).

53 Garcia-Ortiz, J. E. *et al.* High-resolution molecular characterization of the HLA class I and class II in the Tarahumara Amerindian population. *Tissue Antigens* **68**, 135-146, doi:10.1111/j.1399-0039.2006.00636.x (2006).

54 Gaya-Vidal, M. *et al.* Autosomal and X chromosome Alu insertions in Bolivian Aymaras and Quechuas: two languages and one genetic pool. *American Journal of Human Biology* **22**, 154-162, doi:10.1002/ajhb.20967 (2010).

55 Geisler, H. & List, J.-M. in *Arbeitstagung der Indogermanischen Gesellschaft 2009: Die Ausbreitung des Indogermanischen Thesen aus Sprachwissenschaft, Archäologie und Genetik* (Würzburg, Germany, 2009).

56 Gomez-Casado, E. *et al.* HLA genes in Arabic-speaking Moroccans: close relatedness to Berbers and Iberians. *Tissue Antigens* **55**, 239-249, doi:10.1034/j.1399-0039.2000.550307.x (2000).

57 Gomez-Casado, E. *et al.* Origin of Mayans according to HLA genes and the uniqueness of Amerindians. *Tissue Antigens* **61**, 425-436, doi:10.1034/j.1399-0039.2003.00040.x (2003).

58 Gray, R. D. & Atkinson, Q. D. Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature* **426**, 435-439, doi:10.1038/nature02029 (2003).

59 Gray, R. D., Drummond, A. J. & Greenhill, S. J. Language phylogenies reveal expansion pulses and pauses in Pacific settlement. *Science* **323**, 479-483, doi:10.1126/science.1166858 (2009).

60 Gray, R. D. & Jordan, F. M. Language trees support the express-train sequence of Austronesian expansion. *Nature* **405**, 1052-1055, doi:10.1038/35016575 (2000).

61 Greenhill, S. J. & Gray, R. D. in *The Evolution of Cultural Diversity: A Phylogenetic Approach* (eds Ruth Mace, Clare J Holden, & Stephen Shennan) Ch. 3, 31-52 (UCL Press, 2005).

62 Hajjej, A. *et al.* The contribution of HLA class I and II alleles and haplotypes to the investigation of the evolutionary history of Tunisians. *Tissue Antigens* **68**, 153-162, doi:10.1111/j.1399-0039.2005.00622.x (2006).

63 Hatin, W. I. *et al.* Population genetic structure of peninsular Malaysia Malay sub-ethnic groups. *PLOS ONE* **6**, doi:10.1371/journal.pone.0018312 (2011).

64 Henn, B. M. *et al.* Genomic ancestry of North Africans supports back-to-Africa migrations. *PLOS Genetics* **8**, doi:10.1371/journal.pgen.1002397 (2012).

65 Henn, B. M. *et al.* Hunter-gatherer genomic diversity suggests a southern African origin for modern humans. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 5154-5162, doi:10.1073/pnas.1017511108 (2011).

66 Hodoglugil, U. & Mahley, R. W. Turkish population structure and genetic ancestry reveal relatedness among Eurasian populations. *Annals of Human Genetics* **76**, 128-141, doi:10.1111/j.1469-1809.2011.00701.x (2012).

67 Holden, C. & Mace, R. Phylogenetic analysis of the evolution of lactose digestion in adults. *Human Biology* **69**, 605-628 (1997).

68 Holden, C. J. Bantu language trees reflect the spread of farming across sub-Saharan Africa: a maximum-parsimony analysis. *Proceedings of the Royal Society B-Biological Sciences* **269**, 793-799, doi:10.1098/rspb.2002.1955 (2002).

69 Holden, C. J. & Gray, R. D. in *Phylogenetic methods and the prehistory of languages* (eds Peter Forster & Colin Renfrew) 19-31 (McDonald Institute for Archaeological Research, 2006).

70 Holden, C. J., Meade, A. & Pagel, M. in *The Evolution of Cultural Diversity: A Phylogenetic Approach* (eds Ruth Mace, Clare J Holden, & Stephen Shennan) Ch. 4, 53-65 (UCL Press, 2005).

71 Honkola, T. *et al.* Cultural and climatic changes shape the evolutionary history of the Uralic languages. *Journal of Evolutionary Biology* **26**, 1244-1253, doi:10.1111/jeb.12107 (2013).

72 Horai, S. *et al.* mtDNA polymorphism in East Asian populations, with special reference to the peopling of Japan. *American Journal of Human Genetics* **59**, 579-590 (1996).

73 Hunley, K. *et al.* Genetic and linguistic coevolution in Northern Island Melanesia. *PLOS Genetics* **4**, doi:10.1371/journal.pgen.1000239 (2008).

74 Hunley, K. & Long, J. C. Gene flow across linguistic boundaries in Native North American populations. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 1312-1317, doi:10.1073/pnas.0409301102 (2005).

75 Hunley, K. L., Cabana, G. S., Merriwether, D. A. & Long, J. C. A formal test of linguistic and genetic coevolution in Native Central and South America. *American Journal of Physical Anthropology* **132**, 622-631, doi:10.1002/ajpa.20542 (2007).

76 Huyghe, J. R. *et al.* A genome-wide analysis of population structure in the Finnish Saami with implications for genetic association studies. *European Journal of Human Genetics* **19**, 347-352, doi:10.1038/ejhg.2010.179 (2011).

77 Chaubey, G. & Endicott, P. The Andaman Islanders in a regional genetic context: Reexamining the evidence for an early peopling of the archipelago from South Asia. *Human Biology* **85**, 153-171 (2013).

78 Chaubey, G. *et al.* Population genetic structure in Indian Austroasiatic speakers: the role of landscape barriers and sex-specific admixture. *Molecular Biology and Evolution* **28**, 1013-1024, doi:10.1093/molbev/msq288 (2011).

79 Chen, Y. S. *et al.* mtDNA variation in the South African Kung and Khwe—and their genetic relationships to other African populations. *American Journal of Human Genetics* **66**, 1362-1383, doi:10.1086/302848 (2000).

80 Chhaya, S., Desai, S. & Saranath, D. HLA polymorphisms in Sindhi community in Mumbai, India. *International Journal of Immunogenetics* **37**, 373-377, doi:10.1111/j.1744-313X.2010.00936.x (2010).

81 Chu, C. C. *et al.* Diversity of HLA among Taiwan's indigenous tribes and the Ivatans in the Philippines. *Tissue Antigens* **58**, 9-18, doi:10.1034/j.1399-0039.2001.580102.x (2001).

82 Chu, J. Y. *et al.* Genetic relationship of populations in China. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 11763-11768, doi:10.1073/pnas.95.20.11763 (1998).

83 Ingman, M. & Gyllensten, U. Mitochondrial genome variation and evolutionary history of Australian and New Guinean aborigines. *Genome Research* **13**, 1600-1606, doi:10.1101/gr.686603 (2003).

84 Ingman, M., Kaessmann, H., Paabo, S. & Gyllensten, U. Mitochondrial genome variation and the origin of modern humans. *Nature* **408**, 708-713 (2000).

85 Ivanova, M. *et al.* HLA polymorphism in Bulgarians defined by high-resolution typing methods in comparison with other populations. *Tissue Antigens* **60**, 496-504, doi:10.1034/j.1399-0039.2002.600605.x (2002).

86 Jarvis, J. P. *et al.* Patterns of ancestry, signatures of natural selection, and genetic association with stature in Western African pygmies. *PLOS Genetics* **8**, 299-313, doi:10.1371/journal.pgen.1002641 (2012).

87 Jin, F. *et al.* Population genetic studies on nine Aboriginal ethnic groups of Taiwan. I. Red cell enzyme systems. *Anthropological Science* **107**, 229-246 (1999).

88 Jinam, T. A. *et al.* Evolutionary history of continental Southeast Asians:“Early train” hypothesis based on genetic analysis of mitochondrial and autosomal DNA data. *Molecular Biology and Evolution* **29**, 3513-3527, doi:10.1093/molbev/mss169 (2012).

89 Jordan, F. M., Gray, R. D., Greenhill, S. J. & Mace, R. Matrilocal residence is ancestral in Austronesian societies. *Proceedings of the Royal Society B-Biological Sciences* **276**, 1957-1964, doi:10.1098/rspb.2009.0088 (2009).

90 Kang, L. L. *et al.* Genetic structures of the Tibetans and the Deng people in the Himalayas viewed from autosomal STRs. *Journal of Human Genetics* **55**, 270-277, doi:10.1038/jhg.2010.21 (2010).

91 Kharkov, V. N. *et al.* The origin of Yakuts: analysis of the Y-chromosome haplotypes. *Molecular Biology* **42**, 198-208, doi:10.1134/s0026893308020040 (2008).

92 Kitchen, A., Ehret, C., Assefa, S. & Mulligan, C. J. Bayesian phylogenetic analysis of Semitic languages identifies an Early Bronze Age origin of Semitic in the Near East.

Proceedings of the Royal Society B: Biological Sciences **276**, 2703-2710, doi:10.1098/rspb.2009.0408 (2009).

93 Knight, A. *et al.* African Y chromosome and mtDNA divergence provides insight into the history of click languages. *Current Biology* **13**, 464-473, doi:10.1016/s0960-9822(03)00130-1 (2003).

94 Kohlrausch, F. B. *et al.* Geography influences microsatellite polymorphism diversity in Amerindians. *American Journal of Physical Anthropology* **126**, 463-470, doi:10.1002/ajpa.20042 (2005).

95 Kopelman, N. M. *et al.* Genomic microsatellites identify shared Jewish ancestry intermediate between Middle Eastern and European populations. *BMC Genetics* **10**, doi:10.1186/1471-2156-10-80 (2009).

96 Krause, J. *et al.* The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. *Nature* **464**, 894-897, doi:10.1038/nature08976 (2010).

97 Lachance, J. *et al.* Evolutionary history and adaptation from high-coverage whole-genome sequences of diverse African hunter-gatherers. *Cell* **150**, 457-469, doi:10.1016/j.cell.2012.07.009 (2012).

98 Levinson, S. C., Greenhill, S. J., Gray, R. D. & Dunn, M. Universal typological dependencies should be detectable in the history of language families. *Linguistic Typology* **15**, 509–534, doi:10.1515/LITY.2011.034 (2011).

99 Li, H., Cho, K., Kidd, J. R. & Kidd, K. K. Genetic landscape of Eurasia and “admixture” in Uyghurs. *American Journal of Human Genetics* **85**, 934-937, doi:10.1016/j.ajhg.2009.10.024 (2009).

100 Li, J. Z. *et al.* Worldwide human relationships inferred from genome-wide patterns of variation. *Science* **319**, 1100-1104, doi:10.1126/science.1153717 (2008).

101 Lin, M. *et al.* Heterogeneity of Taiwan's indigenous population: possible relation to prehistoric Mongoloid dispersals. *Tissue Antigens* **55**, 1-9, doi:10.1034/j.1399-0039.2000.550101.x (2000).

102 Listman, J. B. *et al.* Southeast Asian origins of five Hill Tribe populations and correlation of genetic to linguistic relationships inferred with genome-wide SNP data. *American Journal of Physical Anthropology* **144**, 300-308, doi:10.1002/ajpa.21408 (2011).

103 Magalhaes, T. R. *et al.* HGDP and HapMap snalysis by Ancestry Mapper reveals local and global population relationships. *PLOS ONE* **7**, doi:10.1371/journal.pone.0049438 (2012).

104 Martinez-Cruz, B. *et al.* In the heartland of Eurasia: the multilocus genetic landscape of Central Asian populations. *European Journal of Human Genetics* **19**, 216-223, doi:10.1038/ejhg.2010.153 (2011).

105 Martinez-Laso, J. *et al.* HLA in Jaidukama: an Amerindian secluded Colombian population with new haplotypes and Asian and Pacific-shared alleles. *Molecular Biology Reports* **38**, 3689-3701, doi:10.1007/s11033-010-0483-6 (2011).

106 Martinez-Laso, J. *et al.* HLA molecular markers in Tuvinians: a population with both Oriental and Caucasoid characteristics. *Annals of Human Genetics* **65**, 245-261, doi:10.1046/j.1469-1809.2001.6530245.x (2001).

107 Martinez-Laso, J. *et al.* Origin of Bolivian Quechua Amerindians: their relationship

with other American Indians and Asians according to HLA genes. *European Journal of Medical Genetics* **49**, 169-185, doi:10.1016/j.ejmg.2005.04.005 (2006).

108 McEvoy, B. P. *et al.* Whole-genome genetic diversity in a sample of Australians with deep Aboriginal ancestry. *American Journal of Human Genetics* **87**, 297-305, doi:10.1016/j.ajhg.2010.07.008 (2010).

109 Melton, T. *et al.* Polynesian genetic affinities with Southeast Asian populations as identified by mtDNA analysis. *American Journal of Human Genetics* **57**, 403-414 (1995).

110 Mendizabal, I. *et al.* Reconstructing the population history of European Romani from genome-wide data. *Current Biology* **22**, 2342-2349, doi:10.1016/j.cub.2012.10.039 (2012).

111 Mesa, N. R. *et al.* Autosomal, mtDNA, and Y-chromosome diversity in Amerinds: pre- and post-Columbian patterns of gene flow in South America. *American Journal of Human Genetics* **67**, 1277-1286 (2000).

112 Mirabal, S., Cadenas, A. M., Garcia-Bertrand, R. & Herrera, R. J. Ascertaining the role of Taiwan as a source for the Austronesian expansion. *American Journal of Physical Anthropology* **150**, 551-564, doi:10.1002/ajpa.22226 (2013).

113 Mohyuddin, A. *et al.* HLA polymorphism in six ethnic groups from Pakistan. *Tissue Antigens* **59**, 492-501, doi:10.1034/j.1399-0039.2002.590606.x (2002).

114 Morlote, D. M., Gayden, T., Arvind, P., Babu, A. & Herrera, R. J. The Soliga, an isolated tribe from Southern India: genetic diversity and phylogenetic affinities. *Journal of Human Genetics* **56**, 258-269, doi:10.1038/jhg.2010.173 (2011).

115 Moscoso, J. *et al.* HLA genes of Aleutian Islanders living between Alaska (USA) and Kamchatka (Russia) suggest a possible southern Siberia origin. *Molecular Immunology* **45**, 1018-1026, doi:10.1016/j.molimm.2007.07.024 (2008).

116 Moscoso, J. *et al.* HLA genes in Lamas Peruvian-Amazonian Amerindians. *Molecular Immunology* **43**, 1881-1889, doi:10.1016/j.molimm.2005.10.013 (2006).

117 Munkhbat, B. *et al.* Molecular analysis of HLA polymorphism in Khoton-Mongolians. *Tissue Antigens* **50**, 124-134, doi:10.1111/j.1399-0039.1997.tb02851.x (1997).

118 Muro, M. *et al.* HLA polymorphism in the Murcia population (Spain): in the cradle of the archaeologic Iberians. *Human Immunology* **62**, 910-921, doi:10.1016/s0198-8859(01)00290-7 (2001).

119 Nasidze, I. *et al.* Alu insertion polymorphisms and the genetic structure of human populations from the Caucasus. *European Journal of Human Genetics* **9**, 267-272, doi:10.1038/sj.ejhg.5200615 (2001).

120 Nasidze, I., Sarkisian, T., Kerimov, A. & Stoneking, M. Testing hypotheses of language replacement in the Caucasus: evidence from the Y-chromosome. *Human Genetics* **112**, 255-261, doi:10.1007/s00439-002-0874-4 (2003).

121 Nasidze, I. & Stoneking, M. Mitochondrial DNA variation and language replacements in the Caucasus. *Proceedings of the Royal Society B-Biological Sciences* **268**, 1197-1206 (2001).

122 Nei, M. & Roychoudhury, A. K. Evolutionary relationships of human populations on a global scale. *Molecular Biology and Evolution* **10**, 927-943 (1993).

123 Nowak, J. *et al.* Allele and extended haplotype polymorphism of HLA-A, -C, -B, -

DRB1 and -DQB1 loci in Polish population and genetic affinities to other populations. *Tissue Antigens* **71**, 193-205, doi:10.1111/j.1399-0039.2007.00991.x (2008).

124 Nurbakova, D., Rusakov, S. & Alexandrov, V. Quantifying uncertainty in phylogenetic studies of the Slavonic languages. *Procedia Computer Science* **18**, 2269-2277, doi:10.1016/j.procs.2013.05.398 (2013).

125 Omoto, K. & Saitou, N. Genetic origins of the Japanese: a partial support for the dual structure hypothesis. *American Journal of Physical Anthropology* **102**, 437-446 (1997).

126 Pagani, L. *et al.* Ethiopian genetic diversity reveals linguistic stratification and complex influences on the Ethiopian gene pool. *American Journal of Human Genetics* **91**, 83-96, doi:10.1016/j.ajhg.2012.05.015 (2012).

127 Pagel, M., Atkinson, Q. D. & Meade, A. Frequency of word-use predicts rates of lexical evolution throughout Indo-European history. *Nature* **449**, 717-U717, doi:10.1038/nature06176 (2007).

128 Pagel, M. & Meade, A. in *The Evolution of Cultural Diversity: A Phylogenetic Approach* (eds Ruth Mace, Clare J Holden, & Stephen Shennan) Ch. 13, 235-256 (UCL Press, 2005).

129 Parolin, M. L. & Carnese, F. R. HLA-DRB1 alleles in four Amerindian populations from Argentina and Paraguay. *Genetics and Molecular Biology* **32**, 212-219 (2009).

130 Patin, E. *et al.* Inferring the demographic history of African farmers and Pygmy hunter–gatherers using a multilocus resequencing data set. *PLOS Genetics* **5**, doi:10.1371/journal.pgen.1000448 (2009).

131 Pemberton, T. J., DeGiorgio, M. & Rosenberg, N. A. Population structure in a comprehensive genomic data set on human microsatellite variation. *G3-Genes Genomes Genetics* **3**, 891-907, doi:10.1534/g3.113.005728 (2013).

132 Petersen, D. C. *et al.* Complex patterns of genomic admixture within southern Africa. *PLOS Genetics* **9**, doi:10.1371/journal.pgen.1003309 (2013).

133 Pickrell, J. K. *et al.* The genetic prehistory of southern Africa. *Nature Communications* **3**, doi:10.1038/ncomms2140 (2012).

134 Pickrell, J. K. & Pritchard, J. K. Inference of population splits and mixtures from genome-wide allele frequency data. *PLOS Genetics* **8**, doi:10.1371/journal.pgen.1002967 (2012).

135 Pickrell, J. K. & Pritchard, J. K. *Extending TreeMix to microsatellite data.* Published by the authors (2012).

136 Pugach, I., Delfin, F., Gunnarsdottir, E., Kayser, M. & Stoneking, M. Genome-wide data substantiate Holocene gene flow from India to Australia. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 1803-1808, doi:10.1073/pnas.1211927110 (2013).

137 Rama, T. & Singh, A. K. in *International Conference RANLP 2009* 355–359 (Borovets, Bulgaria, 2009).

138 Rasmussen, M. *et al.* An Aboriginal Australian genome reveals separate human dispersals into Asia. *Science* **334**, 94-98, doi:10.1126/science.1211177 (2011).

139 Rasmussen, M. *et al.* Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* **463**, 757-762, doi:10.1038/nature08835 (2010).

140 Redd, A. J. & Stoneking, M. Peopling of Sahul: mtDNA variation in Aboriginal Australian and Papua New Guinean populations. *American Journal of Human Genetics* **65**, 808-828, doi:10.1086/302533 (1999).

141 Reesink, G., Singer, R. & Dunn, M. Explaining the linguistic diversity of Sahul using population models. *PLOS Biology* **7**, doi:10.1371/journal.pbio.1000241 (2009).

142 Regueiro, M. *et al.* Austronesian genetic signature in East African Madagascar and Polynesia. *Journal of Human Genetics* **53**, 106-120, doi:10.1007/s10038-007-0224-4 (2008).

143 Reich, D. *et al.* Reconstructing Native American population history. *Nature* **488**, 370-+, doi:10.1038/nature11258 (2012).

144 Rexova, K., Bastin, Y. & Frynta, D. Cladistic analysis of Bantu languages: a new tree based on combined lexical and grammatical data. *Naturwissenschaften* **93**, 189-194, doi:10.1007/s00114-006-0088-z (2006).

145 Rey, D. *et al.* HLA genetic profile of Mapuche (Araucanian) Amerindians from Chile. *Molecular Biology Reports* **40**, 4257-4267, doi:10.1007/s11033-013-2509-3 (2013).

146 Rolf, B. *et al.* Microsatellite profiles reveal an unexpected genetic relationship between Asian populations. *Human Genetics* **102**, 647-652, doi:10.1007/s004390050757 (1998).

147 Rosenberg, N. A. *et al.* Low levels of genetic divergence across geographically and linguistically diverse populations from India. *PLOS Genetics* **2**, 2052-2061, doi:10.1371/journal.pgen.0020215 (2006).

148 Rosenberg, N. A. *et al.* Genetic structure of human populations. *Science* **298**, 2381-2385, doi:10.1126/science.1078311 (2002).

149 Rubicz, R. *et al.* Genetic structure of native circumpolar populations based on autosomal, mitochondrial, and Y chromosome DNA markers. *American Journal of Physical Anthropology* **143**, 62-74, doi:10.1002/ajpa.21290 (2010).

150 Ruiz-Linares, A. *et al.* Microsatellites provide evidence for Y chromosome diversity among the founders of the New World. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 6312-6317, doi:10.1073/pnas.96.11.6312 (1999).

151 Salmela, E. *et al.* Swedish population substructure revealed by genome-wide single nucleotide polymorphism data. *PLOS ONE* **6**, doi:10.1371/journal.pone.0016747 (2011).

152 Sanchez-Velasco, P., de Diego, J. E., Paz-Miguel, J. E., Ocejo-Vinyals, G. & Leyva-Cobian, F. HLA-DR, DQ nucleotide sequence polymorphisms in the Pasiegos (Pas valleys, Northern Spain) and comparison of the allelic and haplotypic frequencies with those of other European populations. *Tissue Antigens* **53**, 65-73, doi:10.1034/j.1399-0039.1999.530107.x (1999).

153 Sanchez-Velasco, P. *et al.* HLA alleles in isolated populations from North Spain: origin of the Basques and the ancient Iberians. *Tissue Antigens* **61**, 384-392, doi:10.1034/j.1399-0039.2003.00041.x (2003).

154 Sanchez-Velasco, P., Karadsheh, N. S., Garcia-Martin, A., de Alegria, C. R. & Leyva-Cobian, F. Molecular analysis of HLA allelic frequencies and haplotypes in Jordanians and comparison with other related populations. *Human Immunology* **62**, 901-909,

doi:10.1016/s0198-8859(01)00289-0 (2001).

155 Sanchez-Velasco, P. & Leyva-Cobian, F. The HLA class I and class II allele frequencies studied at the DNA level in the Svanetian population (Upper Caucasus) and their relationships to Western European populations. *Tissue Antigens* **58**, 223-233, doi:10.1034/j.1399-0039.2001.580402.x (2001).

156 Sellen, D. W. & Mace, R. Fertility and mode of subsistence: a phylogenetic analysis. *Current Anthropology* **38**, 878-889, doi:10.1086/204677 (1997).

157 Serva, M. & Petroni, F. Indo-European languages tree by Levenshtein distance. *EPL (Europhysics Letters)* **81**, doi:10.1209/0295-5075/81/68005 (2008).

158 Shaw, C. K., Chen, L. L., Lee, A. & Lee, T. D. Distribution of HLA gene and haplotype frequencies in Taiwan: a comparative study among Min-nan, Hakka, Aborigines and Mainland Chinese. *Tissue Antigens* **53**, 51-64, doi:10.1034/j.1399-0039.1999.530106.x (1999).

159 Schlebusch, C. M., Lombard, M. & Soodyall, H. MtDNA control region variation affirms diversity and deep sub-structure in populations from southern Africa. *BMC Evolutionary Biology* **13**, doi:10.1186/1471-2148-13-56 (2013).

160 Schlebusch, C. M. *et al.* Genomic variation in seven Khoe-San groups reveals adaptation and complex African history. *Science* **338**, 374-379, doi:10.1126/science.1227721 (2012).

161 Sikora, M., Laayouni, H., Calafell, F., Comas, D. & Bertranpetit, J. A genomic analysis identifies a novel component in the genetic structure of sub-Saharan African populations. *European Journal of Human Genetics* **19**, 84-88, doi:10.1038/ejhg.2010.141 (2011).

162 Starikovskaya, E. B. *et al.* Mitochondrial DNA diversity in indigenous populations of the southern extent of Siberia, and the origins of Native American haplogroups. *Annals of Human Genetics* **69**, 67-89, doi:10.1046/j.1529-8817.2003.00127.x (2005).

163 Sulcebe, G. *et al.* HLA allele and haplotype frequencies in the Albanian population and their relationship with the other European populations. *International Journal of Immunogenetics* **36**, 337-343, doi:10.1111/j.1744-313X.2009.00868.x (2009).

164 Suslova, T. A. *et al.* HLA gene and haplotype frequencies in Russians, Bashkirs and Tatars, living in the Chelyabinsk Region (Russian South Urals). *International Journal of Immunogenetics* **39**, 394-408, doi:10.1111/j.1744-313X.2012.01117.x (2012).

165 Tajima, A. *et al.* Genetic origins of the Ainu inferred from combined DNA analyses of maternal and paternal lineages. *Journal of Human Genetics* **49**, 187-193, doi:10.1007/s10038-004-0131-x (2004).

166 Tajima, A. *et al.* Mitochondrial DNA polymorphisms in nine aboriginal groups of Taiwan: implications for the population history of aboriginal Taiwanese. *Human Genetics* **113**, 24-33, doi:10.1007/s00439-003-0945-1 (2003).

167 Thomas, R., Nair, S. B. & Banerjee, M. HLA-B and HLA-C alleles and haplotypes in the Dravidian tribal populations of southern India. *Tissue Antigens* **64**, 58-65, doi:10.1111/j.1399-0039.2004.00244.x (2004).

168 Tishkoff, S. A. *et al.* History of click-speaking populations of Africa inferred from mtDNA and Y chromosome genetic variation. *Molecular Biology and Evolution* **24**, 2180-2195, doi:10.1093/molbev/msm155 (2007).

169 Tishkoff, S. A. *et al.* The Genetic structure and history of Africans and African Americans. *Science* **324**, 1035-1044, doi:10.1126/science.1172257 (2009).

170 Torroni, A. *et al.* mtDNA and Y-chromosome polymorphisms in four Native American populations from southern Mexico. *American Journal of Human Genetics* **54**, 303-318 (1994).

171 Tsuneto, L. T. *et al.* HLA class II diversity in seven Amerindian populations. Clues about the origins of the Ache. *Tissue Antigens* **62**, 512-526, doi:10.1046/j.1399-0039.2003.00139.x (2003).

172 Uinuk-ool, T. S., Takezaki, N., Derbeneva, O. A., Volodko, N. V. & Sukernik, R. I. Variation of HLA class II genes in the Nganasan and Ket, two aboriginal Siberian populations. *European Journal of Immunogenetics* **31**, 43-51, doi:10.1111/j.1365-2370.2004.00443.x (2004).

173 Uinuk-ool, T. S., Takezaki, N., Sukernik, R. I., Nagl, S. & Klein, J. Origin and affinities of indigenous Siberian populations as revealed by HLA class II gene frequencies. *Human Genetics* **110**, 209-226, doi:10.1007/s00439-001-0668-0 (2002).

174 Usme-Romero, S., Alonso, M., Hernandez-Cuervo, H., Yunis, E. J. & Yunis, J. J. Genetic differences between Chibcha and Non-Chibcha speaking tribes based on mitochondrial DNA (mtDNA) haplogroups from 21 Amerindian tribes from Colombia. *Genetics and Molecular Biology* **36**, 149-157 (2013).

175 Vargas-Alarcon, G. *et al.* HLA genes in Mexican Teeneks: HLA genetic relationship with other worldwide populations. *Molecular Immunology* **43**, 790-799, doi:10.1016/j.molimm.2005.07.017 (2006).

176 Vargas-Alarcon, G. *et al.* Origin of Mexican Nahuas (Aztecs) according to HLA genes and their relationships with worldwide populations. *Molecular Immunology* **44**, 747-755, doi:10.1016/j.molimm.2006.04.014 (2007).

177 Veeramah, K. R. *et al.* An early divergence of KhoeSan ancestors from those of other modern humans is supported by an ABC-based analysis of autosomal resequencing data. *Molecular Biology and Evolution* **29**, 617-630, doi:10.1093/molbev/msr212 (2012).

178 Verdu, P. *et al.* Origins and genetic diversity of pygmy hunter-gatherers from Western Central Africa. *Current Biology* **19**, 312-318, doi:10.1016/j.cub.2008.12.049 (2009).

179 Walker, R. S., Hill, K. R., Flinn, M. V. & Ellsworth, R. M. Evolutionary history of hunter-gatherer marriage practices. *PLOS ONE* **6**, doi:10.1371/journal.pone.0019066 (2011).

180 Walker, R. S. & Ribeiro, L. A. Bayesian phylogeography of the Arawak expansion in lowland South America. *Proceedings of the Royal Society B-Biological Sciences* **278**, 2562-2567, doi:10.1098/rspb.2010.2579 (2011).

181 Walker, R. S., Wichmann, S., Mailund, T. & Atkisson, C. J. Cultural phylogenetics of the Tupi language family in lowland South America. *PLOS ONE* **7**, doi:10.1371/journal.pone.0035025 (2012).

182 Walsh, S. J., Mitchell, R. J., Watson, N. & Buckleton, J. S. A comprehensive analysis of microsatellite diversity in Aboriginal Australians. *Journal of Human Genetics* **52**, 712-728, doi:10.1007/s10038-007-0172-z (2007).

183 Wang, B. B. *et al.* On the origin of Tibetans and their genetic basis in adapting high-

altitude environments. *PLOS ONE* **6**, doi:10.1371/journal.pone.0017002 (2011).

184 Wang, S. *et al.* Genetic variation and population structure in Native Americans. *PLOS Genetics* **3**, 2049-2067, doi:10.1371/journal.pgen.0030185 (2007).

185 Wangkumhang, P. *et al.* Insight into the peopling of mainland southeast Asia from Thai population genetic structure. *PLOS ONE* **8**, e79522 (2013).

186 Watkins, W. S. *et al.* Patterns of ancestral human diversity: An analysis of Alu-insertion and restriction-site polymorphisms. *American Journal of Human Genetics* **68**, 738-752, doi:10.1086/318793 (2001).

187 Watkins, W. S. *et al.* Genetic variation among world populations: inferences from 100 Alu insertion polymorphisms. *Genome Research* **13**, 1607-1618, doi:10.1101/gr.894603 (2003).

188 Wells, R. S. *et al.* The Eurasian Heartland: A continental perspective on Y-chromosome diversity. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 10244-10249, doi:10.1073/pnas.171305098 (2001).

189 Wollstein, A. *et al.* Demographic history of Oceania inferred from genome-wide data. *Current Biology* **20**, 1983-1992, doi:10.1016/j.cub.2010.10.040 (2010).

190 Xing, J. *et al.* Genomic analysis of natural selection and phenotypic variation in high-altitude Mongolians. *PLOS Genetics* **9**, doi:10.1371/journal.pgen.1003634 (2013).

191 Xing, J. C. *et al.* Toward a more uniform sampling of human genetic diversity: A survey of worldwide populations by high-density genotyping. *Genomics* **96**, 199-210, doi:10.1016/j.ygeno.2010.07.004 (2010).

192 Xing, J. C. *et al.* Fine-scaled human genetic structure revealed by SNP microarrays. *Genome Research* **19**, 815-825, doi:10.1101/gr.085589.108 (2009).

193 Xu, S. H. *et al.* Genetic evidence supports linguistic affinity of Mlabri - a hunter-gatherer group in Thailand. *BMC Genetics* **11**, doi:10.1186/1471-2156-11-18 (2010).

194 Xu, S. H. *et al.* Genetic dating indicates that the Asian-Papuan admixture through Eastern Indonesia corresponds to the Austronesian expansion. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 4574-4579, doi:10.1073/pnas.1118892109 (2012).

195 Yuasa, I. *et al.* A hypervariable STR polymorphism in the CFI gene: southern origin of East Asian-specific group H alleles. *Legal Medicine* **15**, 239-243, doi:10.1016/j.legalmed.2013.04.001 (2013).

196 Yuasa, I. *et al.* Population genetic studies on nine aboriginal ethnic groups of Taiwan II. Serum protein systems. *Anthropological Science* **109**, 257-273 (2001).

197 Yuliwulandari, R. *et al.* Polymorphisms of HLA genes in Western Javanese (Indonesia): close affinities to Southeast Asian populations. *Tissue Antigens* **73**, 46-53, doi:10.1111/j.1399-0039.2008.01178.x (2009).

198 Yunis, J. J., Yunis, E. J. & Yunis, E. Genetic relationship of the Guambino, Paez, and Ingano Amerindians of southwest Colombia using major histocompatibility complex class II haplotypes and blood groups. *Human Immunology* **62**, 970-978, doi:10.1016/s0198-8859(01)00295-6 (2001).

199 Yunis, J. J., Yunis, E. J. & Yunis, E. MHC Class II haplotypes of Colombian Amerindian tribes. *Genetics and Molecular Biology* **36**, 158-166 (2013).

200 Zhu, B. F. *et al.* Distributions of HLA-A and -B alleles and haplotypes in the Yi ethnic

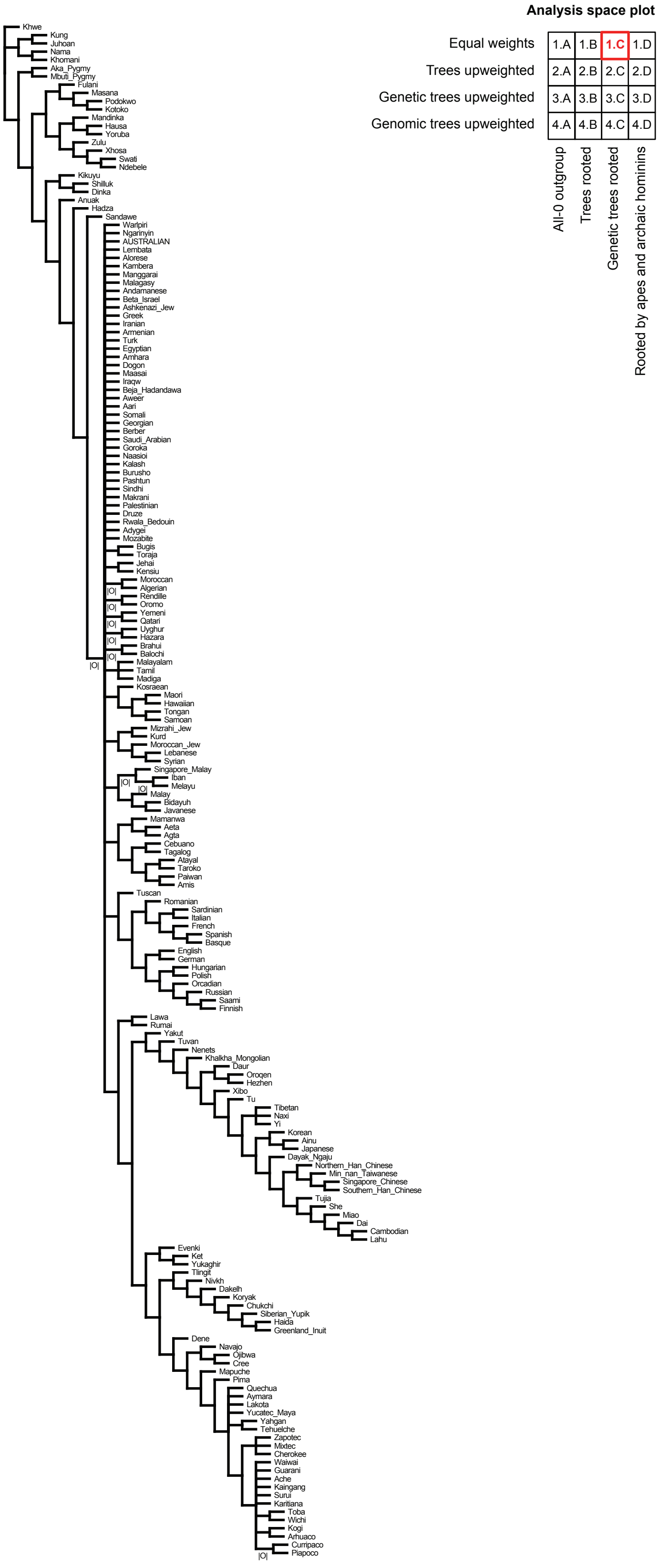
minority of Yunnan, China: relationship to other populations. *Journal of Zhejiang University-Science B* **11**, 127-135, doi:10.1631/jzus.B0900232 (2010).

SUPPLEMENTARY FIGURES

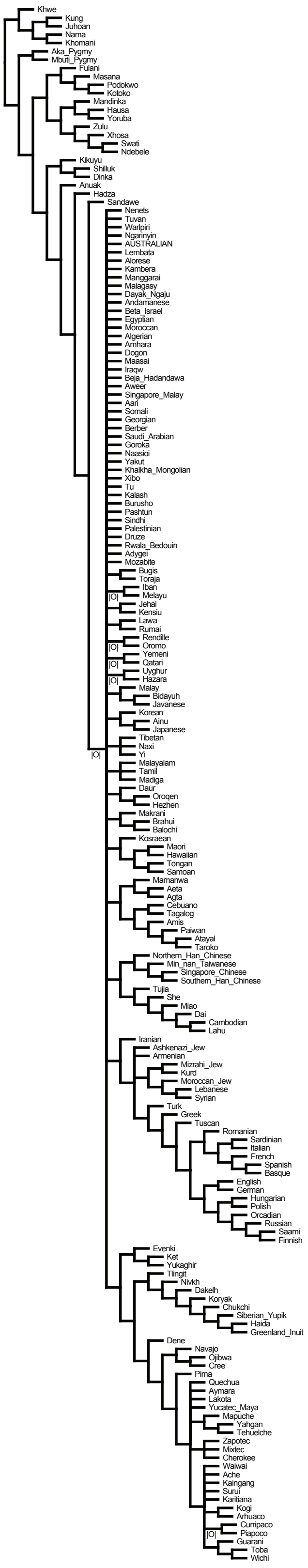
Supplementary Figure S1. Semistrict consensus supertree based on parameter set 1.A of the sensitivity analysis of the representative dataset. A symbol "|O|" indicates clades which would be collapsed into polytomies in a strict consensus supertree.

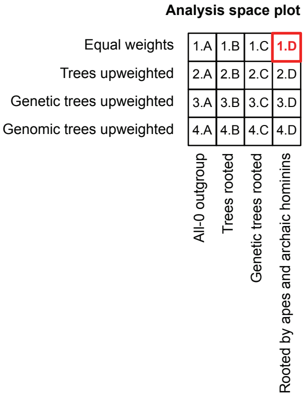
Supplementary Figure S2. Semistrict consensus supertree based on parameter set 1.B of the sensitivity analysis of the representative dataset. A symbol “|O|” indicates clades which would be collapsed into polytomies in a strict consensus supertree.

Supplementary Figure S3. Semistrict consensus supertree based on parameter set 1.C of the sensitivity analysis of the representative dataset. A symbol "|O|" indicates clades which would be collapsed into polytomies in a strict consensus supertree.

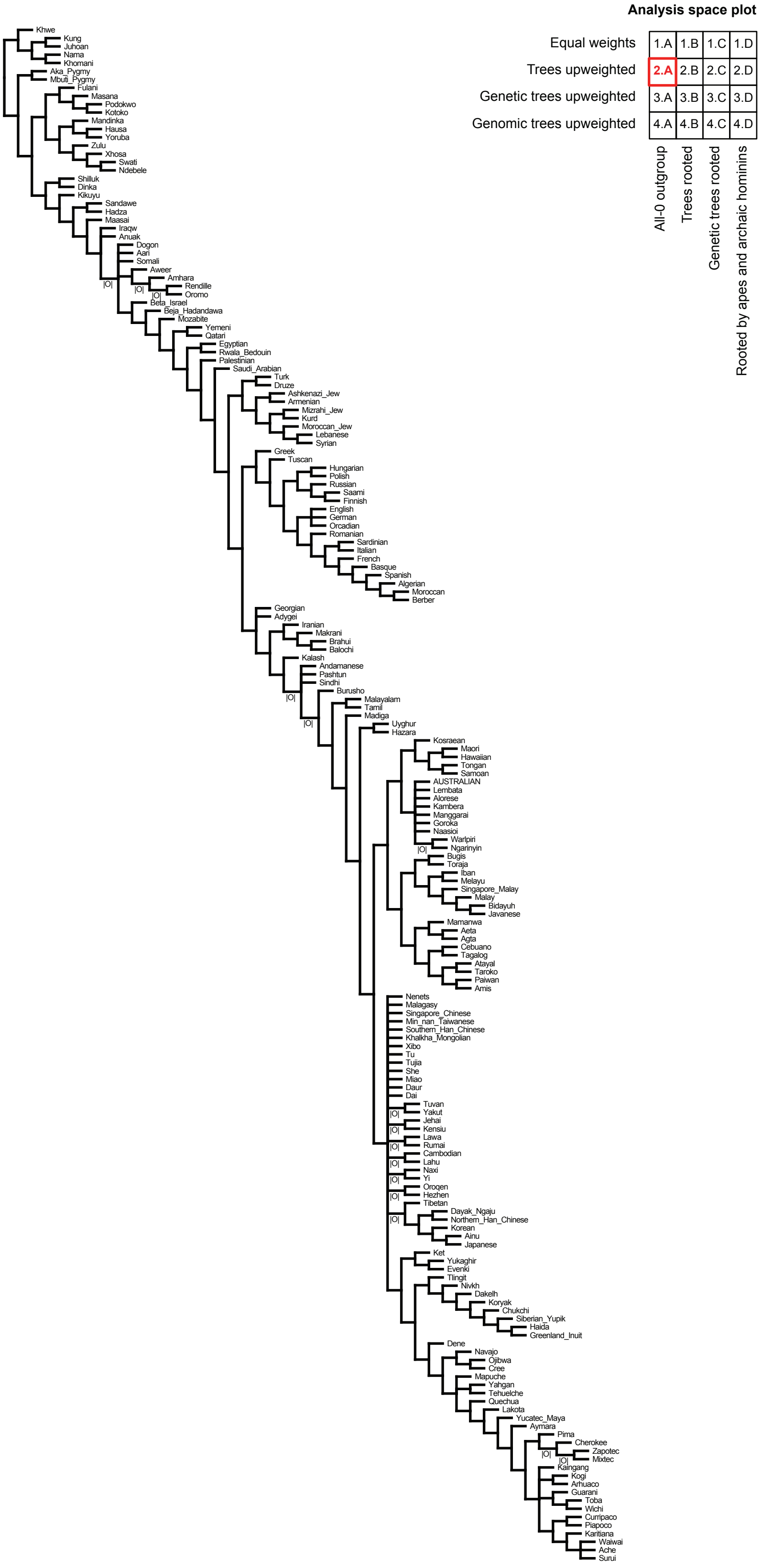


Supplementary Figure S4. Semistrict consensus supertree based on parameter set 1.D of the sensitivity analysis of the representative dataset. A symbol “|O|” indicates clades which would be collapsed into polytomies in a strict consensus supertree.

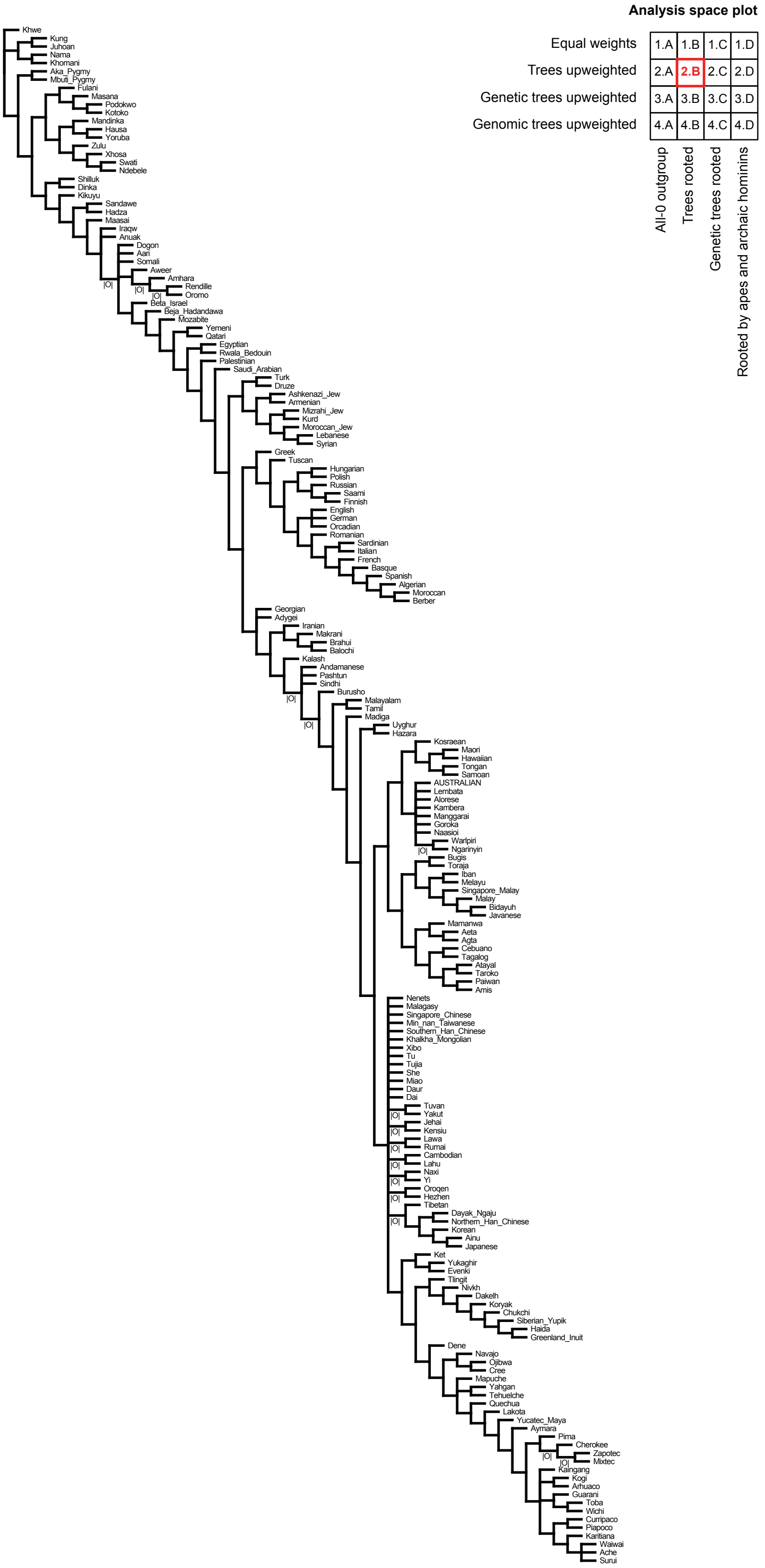




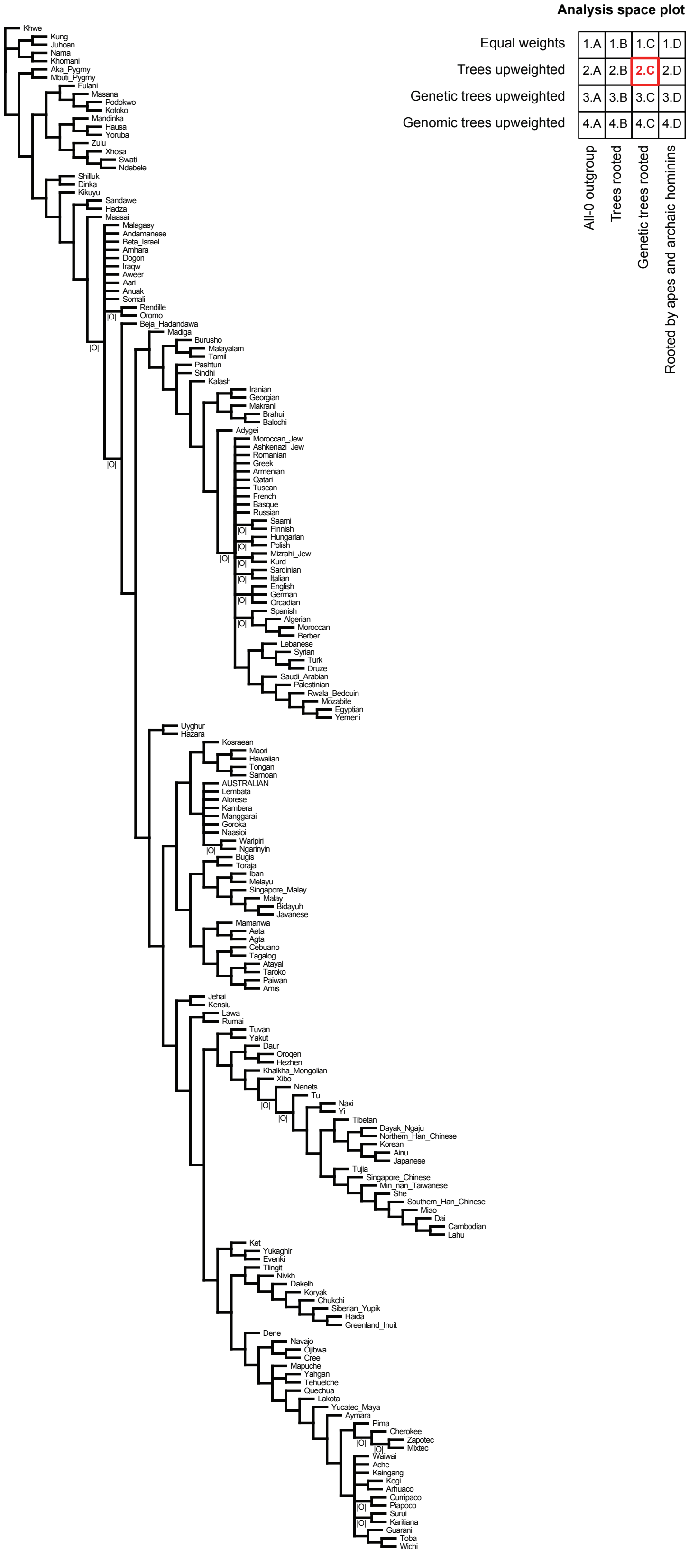
Supplementary Figure S5. Semistrict consensus supertree based on parameter set 2.A of the sensitivity analysis of the representative dataset. A symbol “|O|” indicates clades which would be collapsed into polytomies in a strict consensus supertree.



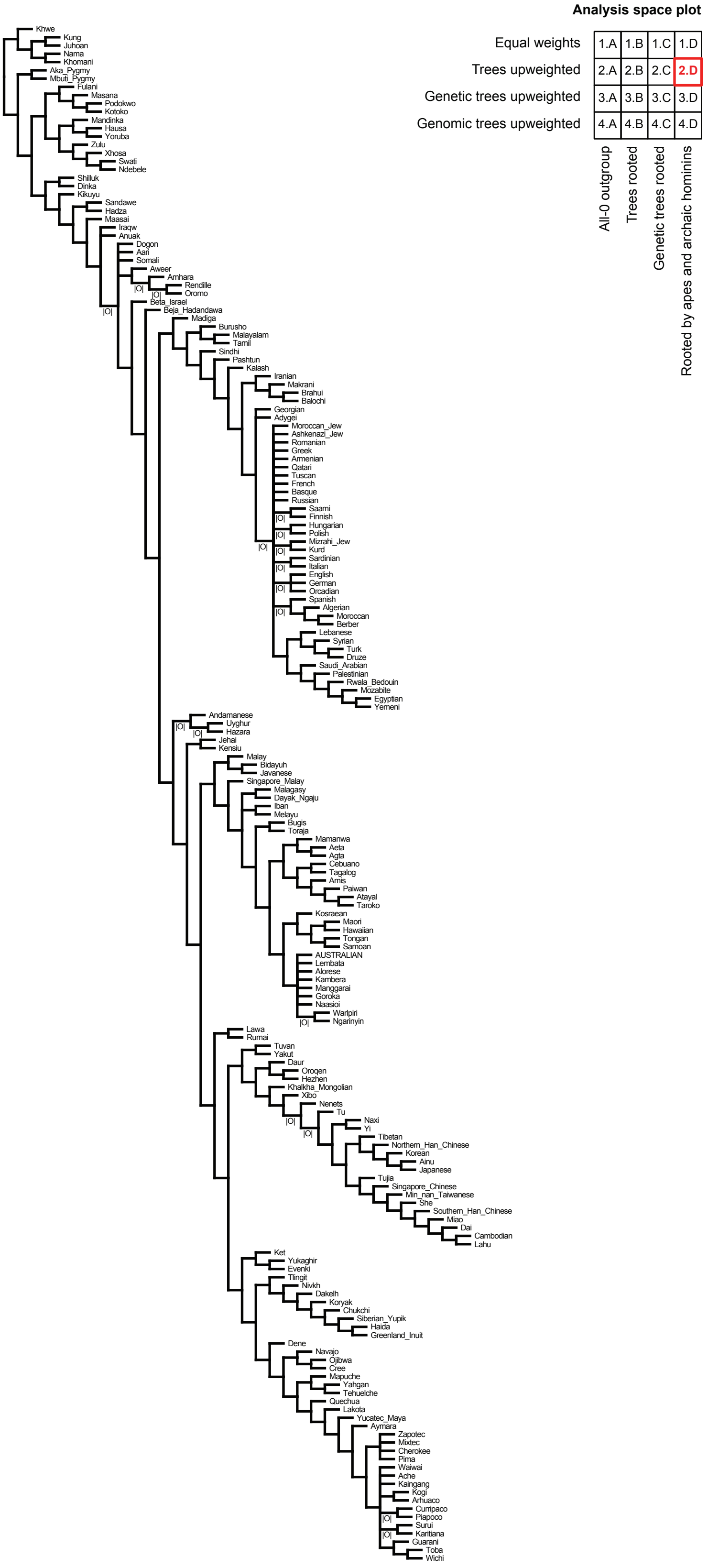
Supplementary Figure S6. Semistrict consensus supertree based on parameter set 2.B of the sensitivity analysis of the representative dataset. A symbol “|O|” indicates clades which would be collapsed into polytomies in a strict consensus supertree.



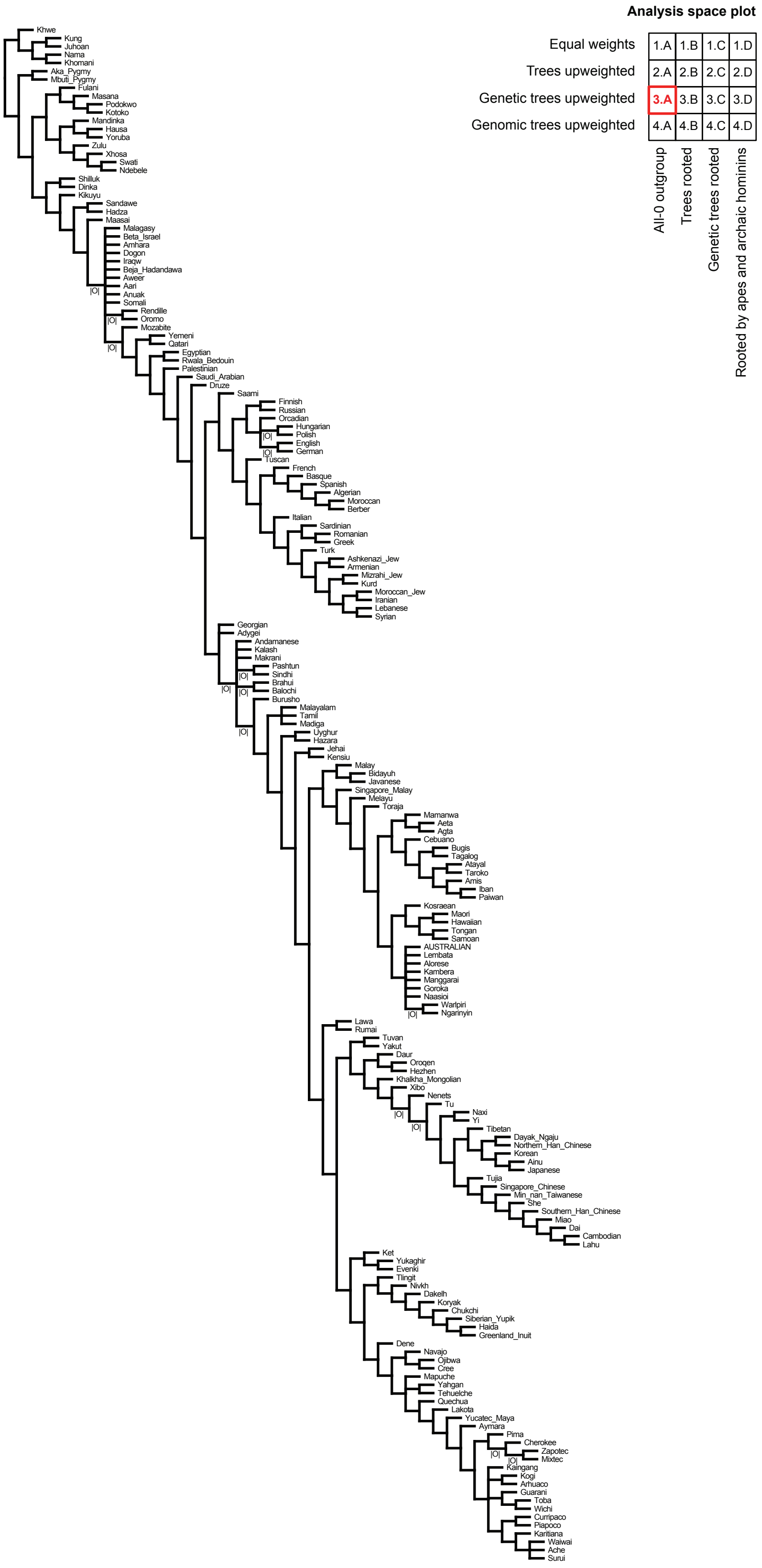
Supplementary Figure S7. Semistrict consensus supertree based on parameter set 2.C of the sensitivity analysis of the representative dataset. A symbol “|O|” indicates clades which would be collapsed into polytomies in a strict consensus supertree.



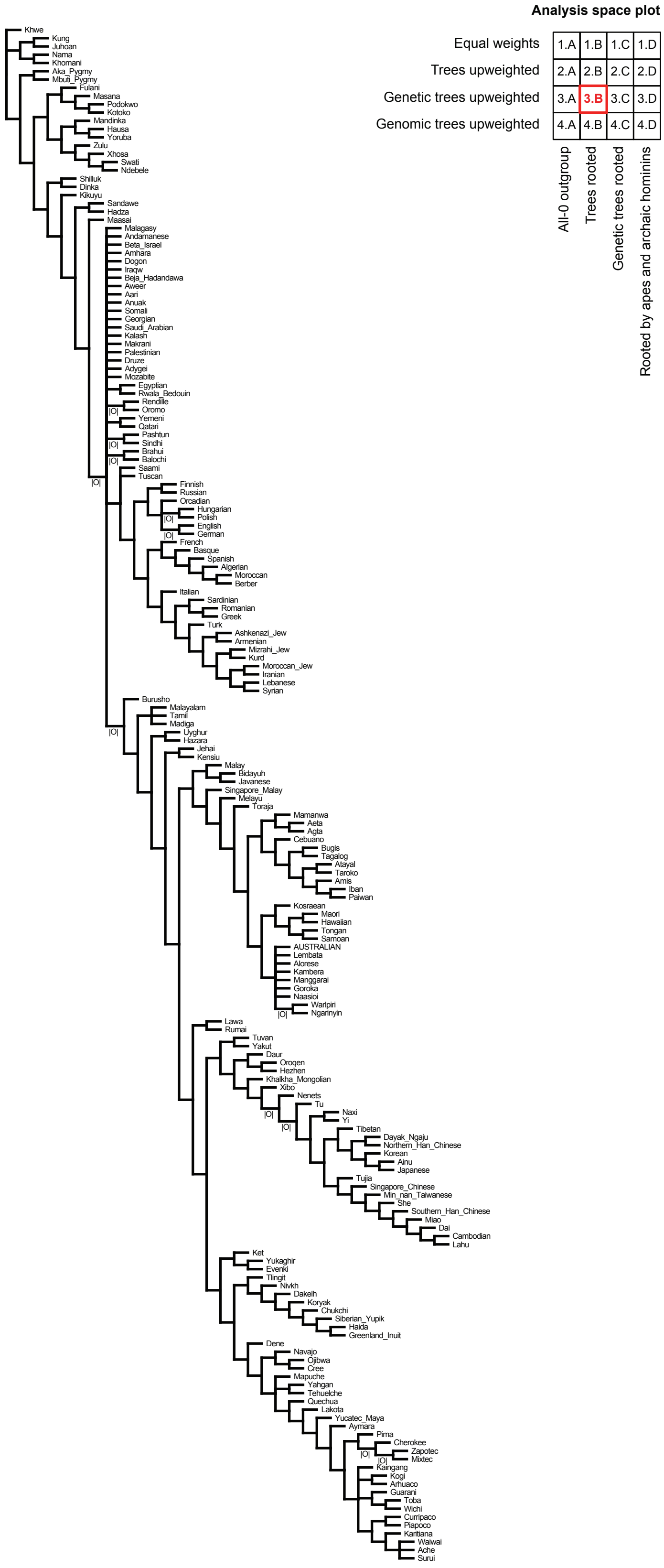
Supplementary Figure S8. Semistrict consensus supertree based on parameter set 2.D of the sensitivity analysis of the representative dataset. A symbol "|O|" indicates clades which would be collapsed into polytomies in a strict consensus supertree.



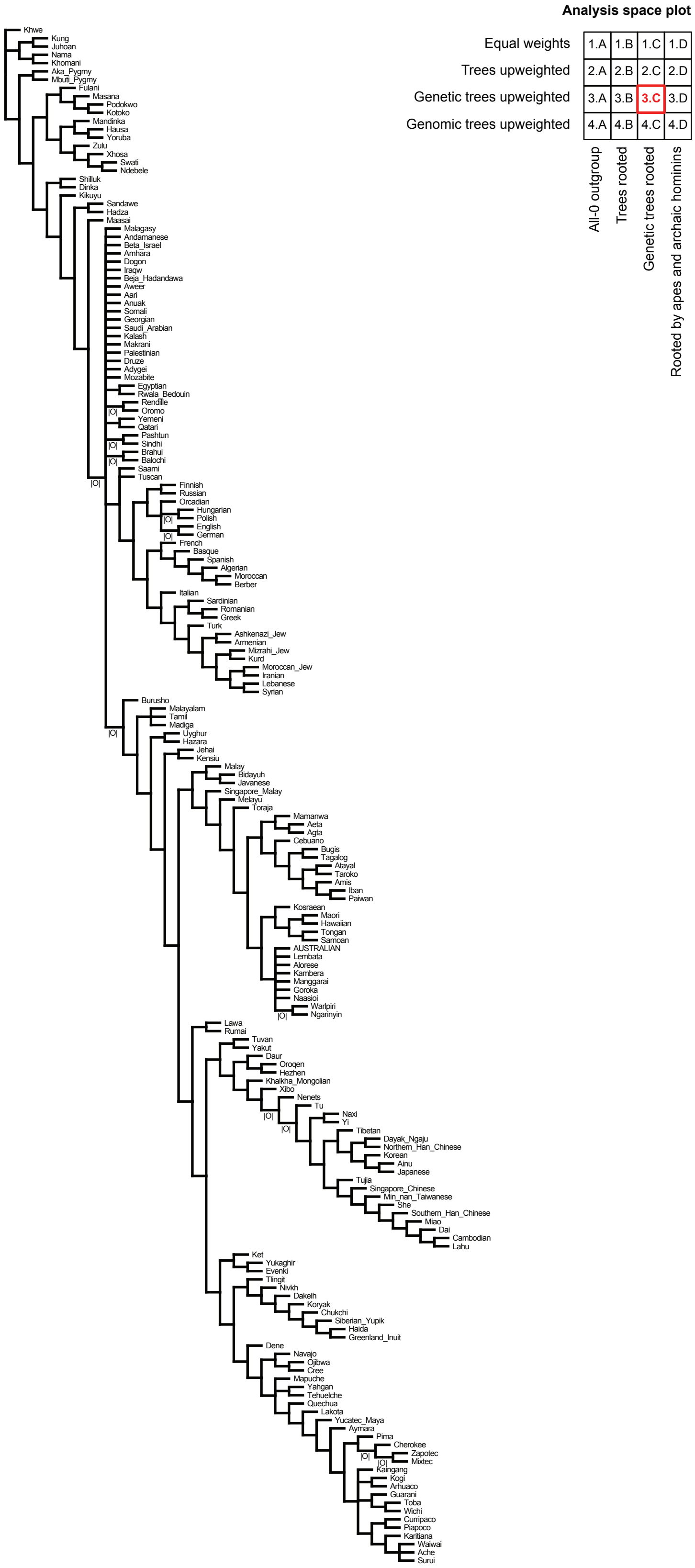
Supplementary Figure S9. Semistrict consensus supertree based on parameter set 3.A of the sensitivity analysis of the representative dataset. A symbol “|O|” indicates clades which would be collapsed into polytomies in a strict consensus supertree.



Supplementary Figure S10. Semistrict consensus supertree based on parameter set 3.B of the sensitivity analysis of the representative dataset. A symbol "|O|" indicates clades which would be collapsed into polytomies in a strict consensus supertree.

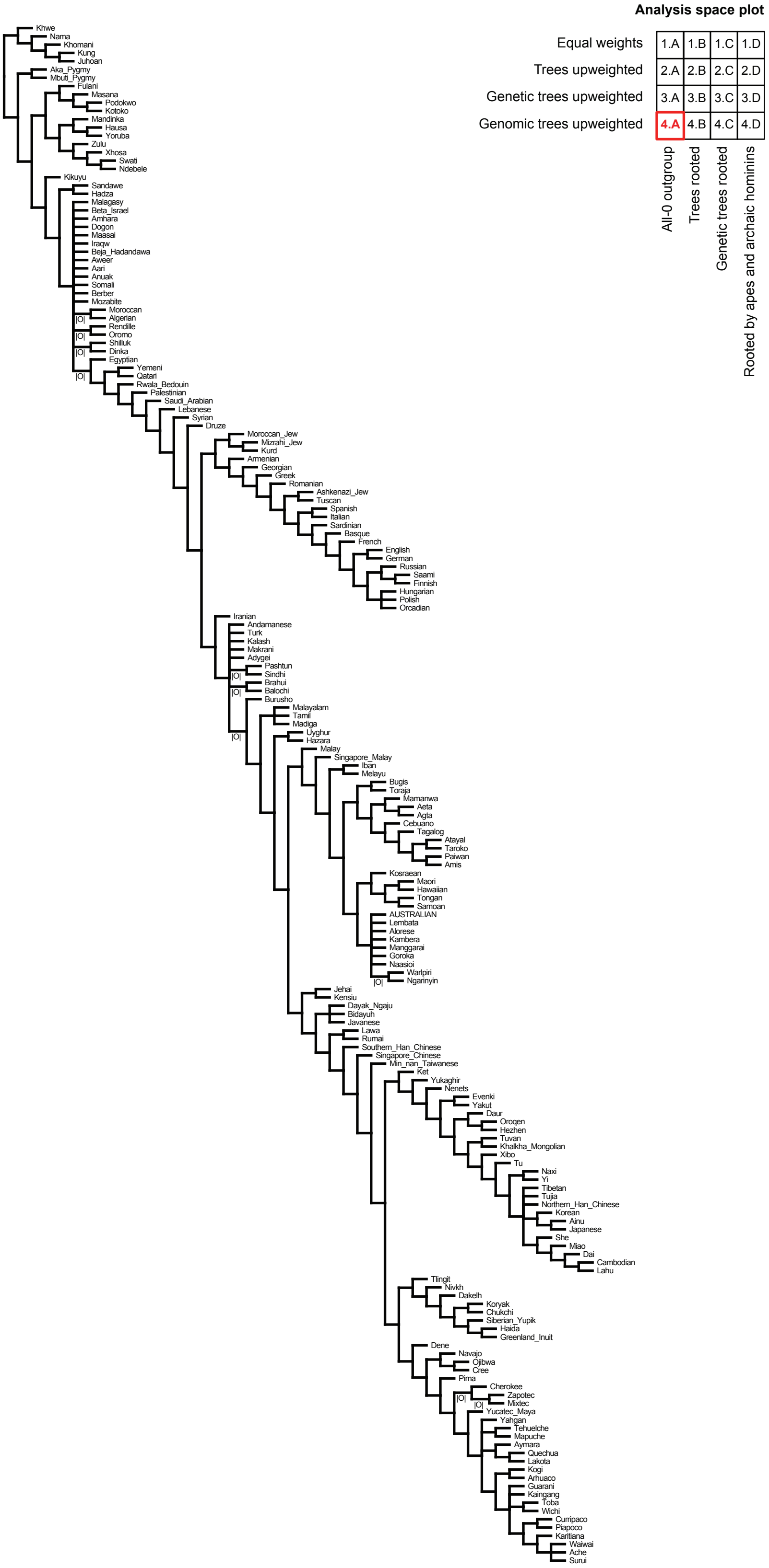


Supplementary Figure S11. Semistrict consensus supertree based on parameter set 3.C of the sensitivity analysis of the representative dataset. A symbol "|O|" indicates clades which would be collapsed into polytomies in a strict consensus supertree.

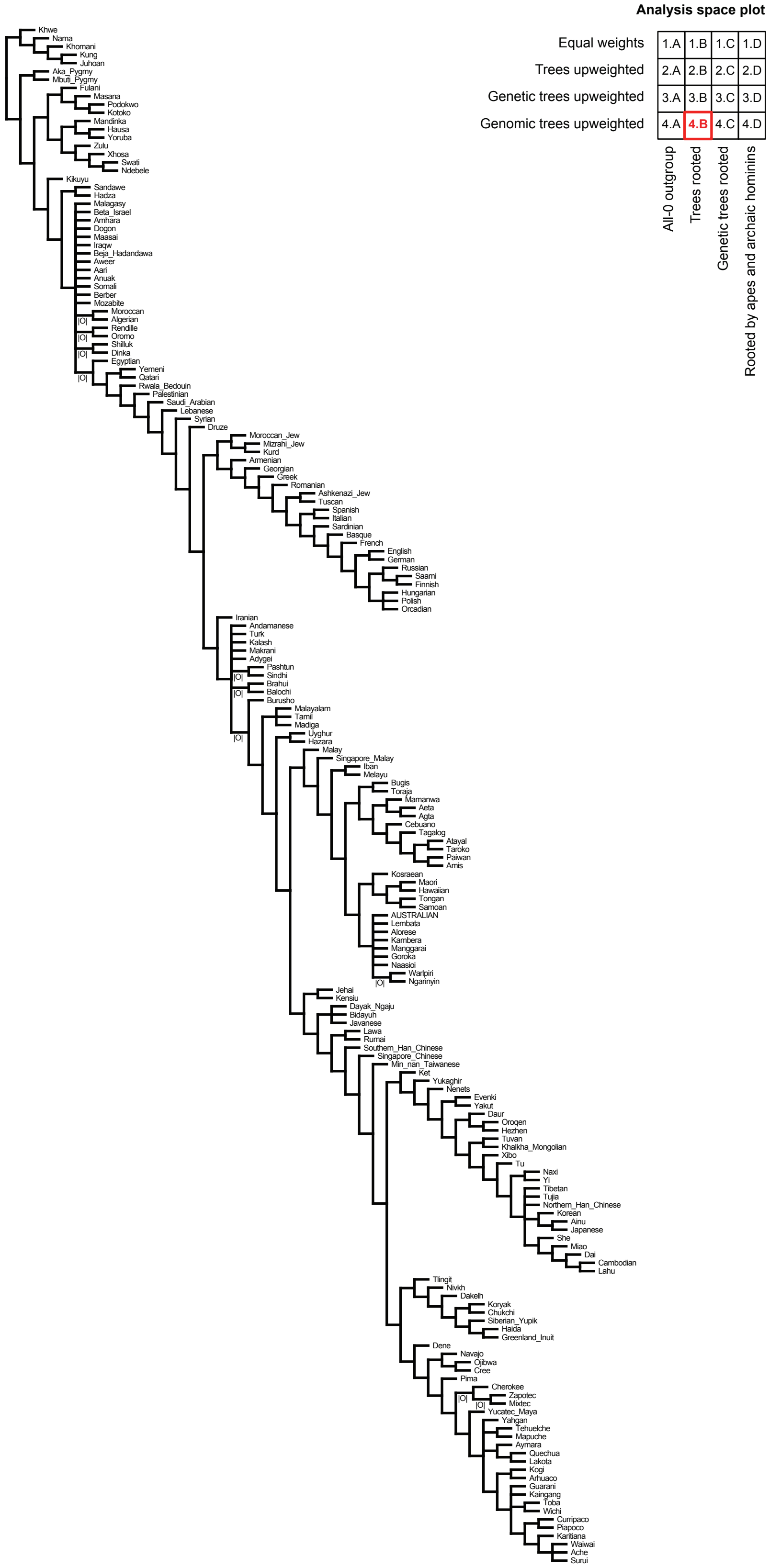


Supplementary Figure S12. Semistrict consensus supertree based on parameter set 3.D of the sensitivity analysis of the representative dataset. A symbol “|O|” indicates clades which would be collapsed into polytomies in a strict consensus supertree.

Supplementary Figure S13. Semistrict consensus supertree based on parameter set 4.A of the sensitivity analysis of the representative dataset. A symbol "|O|" indicates clades which would be collapsed into polytomies in a strict consensus supertree.



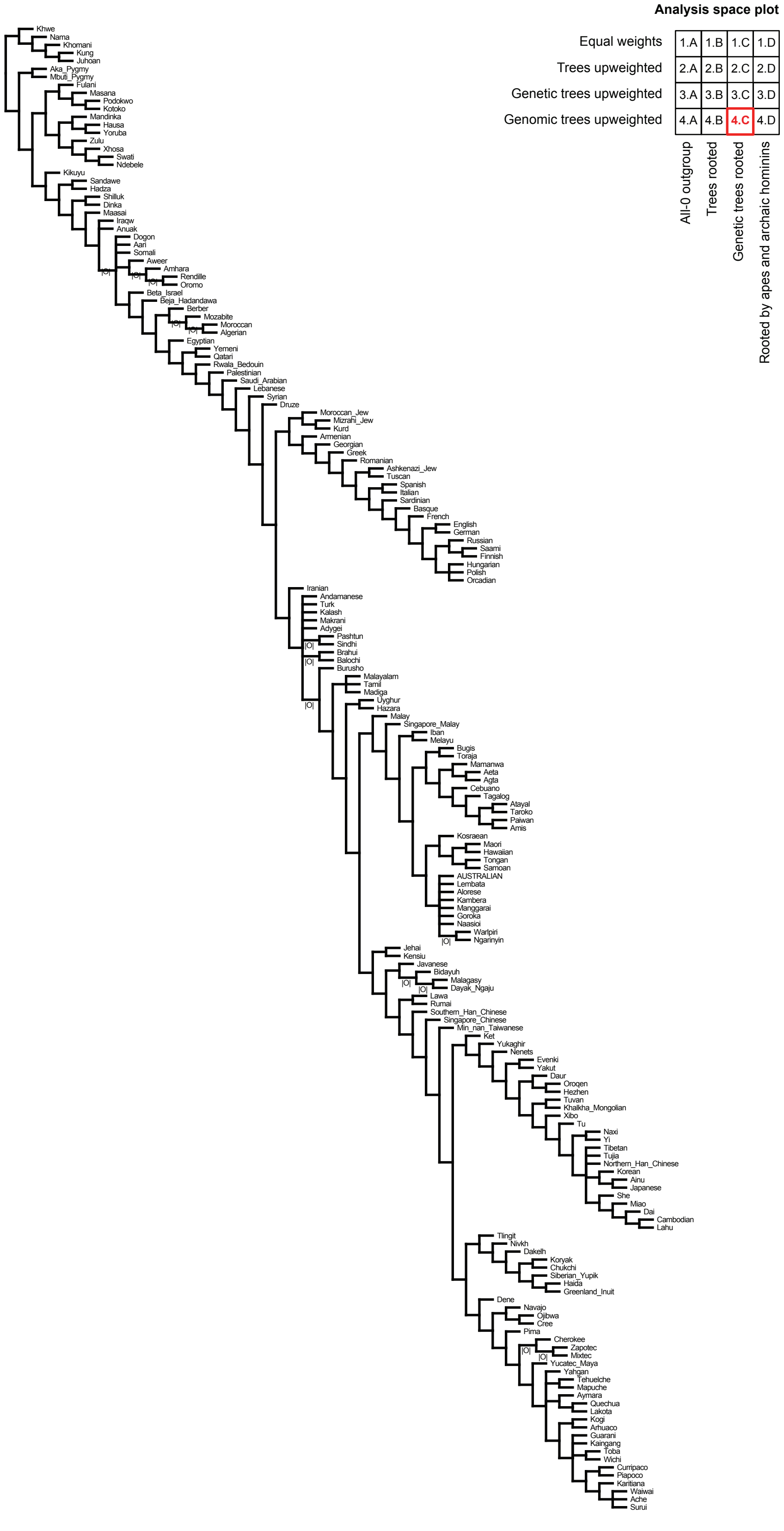
Supplementary Figure S14. Semistrict consensus supertree based on parameter set 4.B of the sensitivity analysis of the representative dataset. A symbol “|O|” indicates clades which would be collapsed into polytomies in a strict consensus supertree.



Analysis space plot

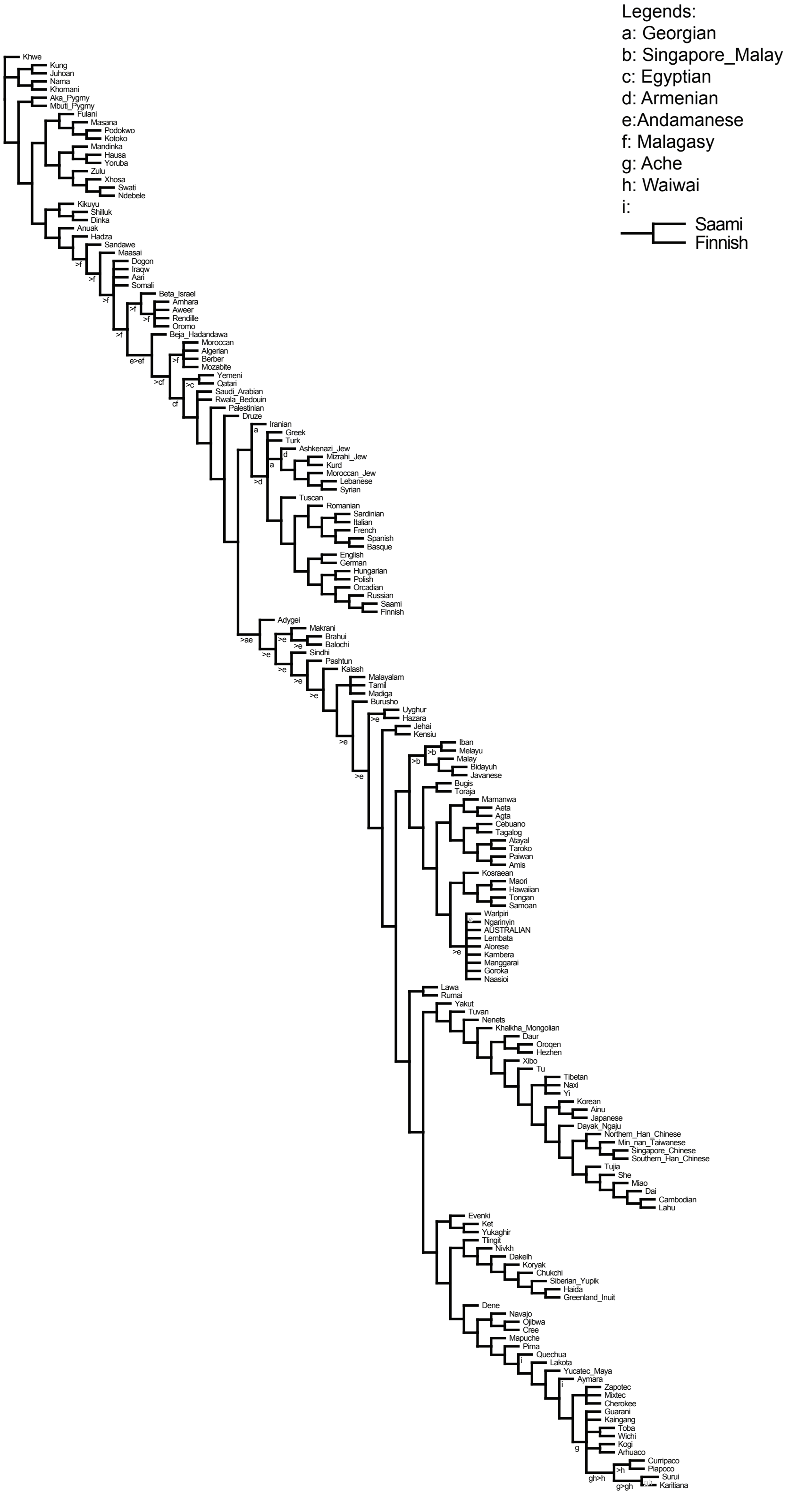
	All-0 outgroup	Trees rooted	Genetic trees rooted	Rooted by apes and archaic hominins
Equal weights	1.A	1.B	1.C	1.D
Trees upweighted	2.A	2.B	2.C	2.D
Genetic trees upweighted	3.A	3.B	3.C	3.D
Genomic trees upweighted	4.A	**4.B**	4.C	4.D

Supplementary Figure S15. Semistrict consensus supertree based on parameter set 4.C of the sensitivity analysis of the representative dataset. A symbol "|O|" indicates clades which would be collapsed into polytomies in a strict consensus supertree.

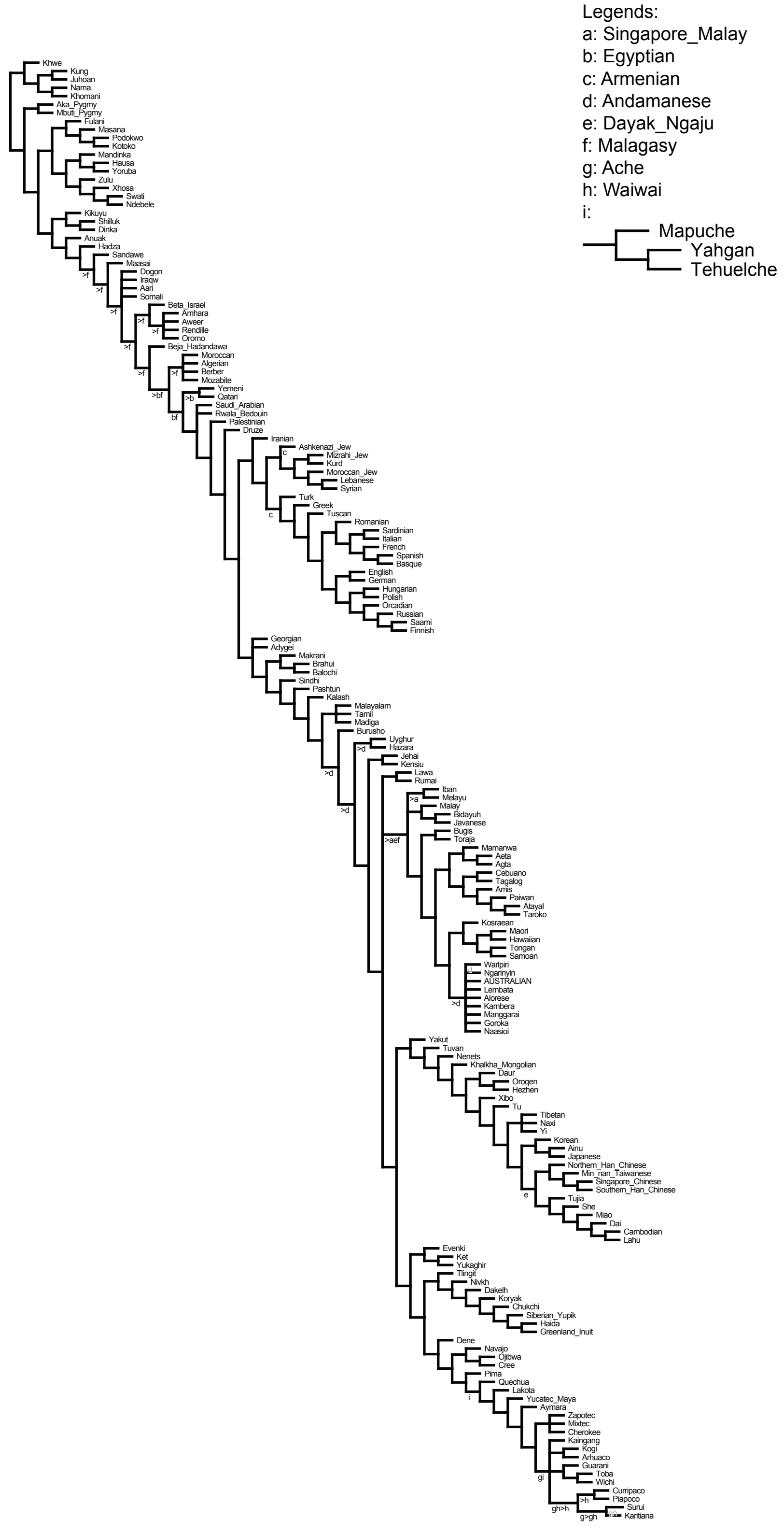


Supplementary Figure S16. Semistrict consensus supertree based on parameter set 4.D of the sensitivity analysis of the representative dataset. A symbol "|O|" indicates clades which would be collapsed into polytomies in a strict consensus supertree.

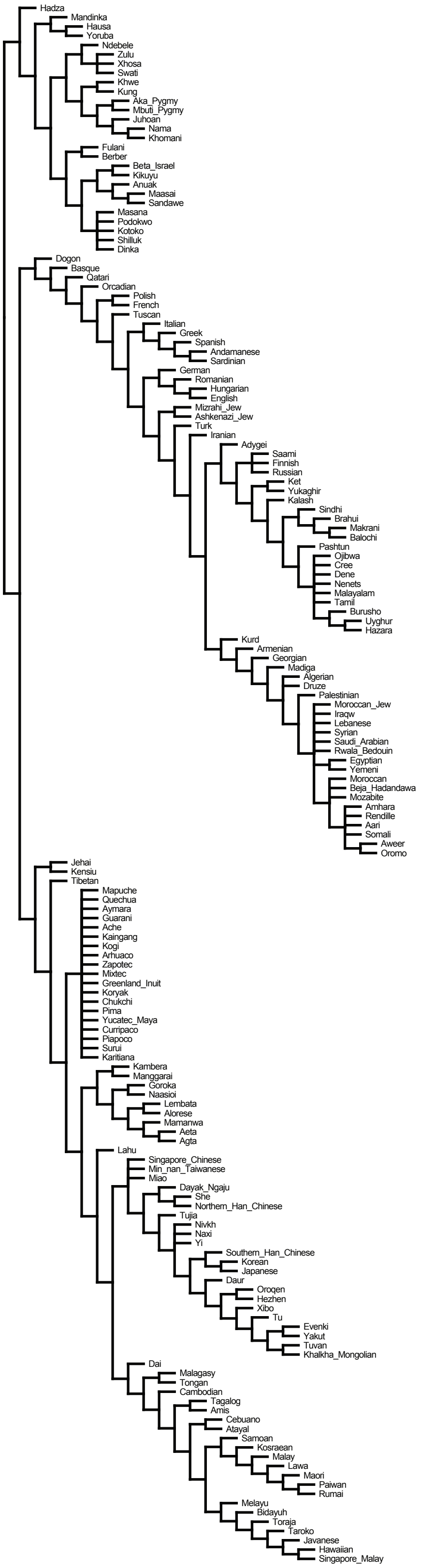
Supplementary Figure S17. Pruned strict consensus supertree based on parameter set 1.C of the representative dataset. Alternative positions of the identified wildcard taxa are indicated. A symbol “>” indicates wildcard taxa which would be placed within a polytomy, if included in the analysis.



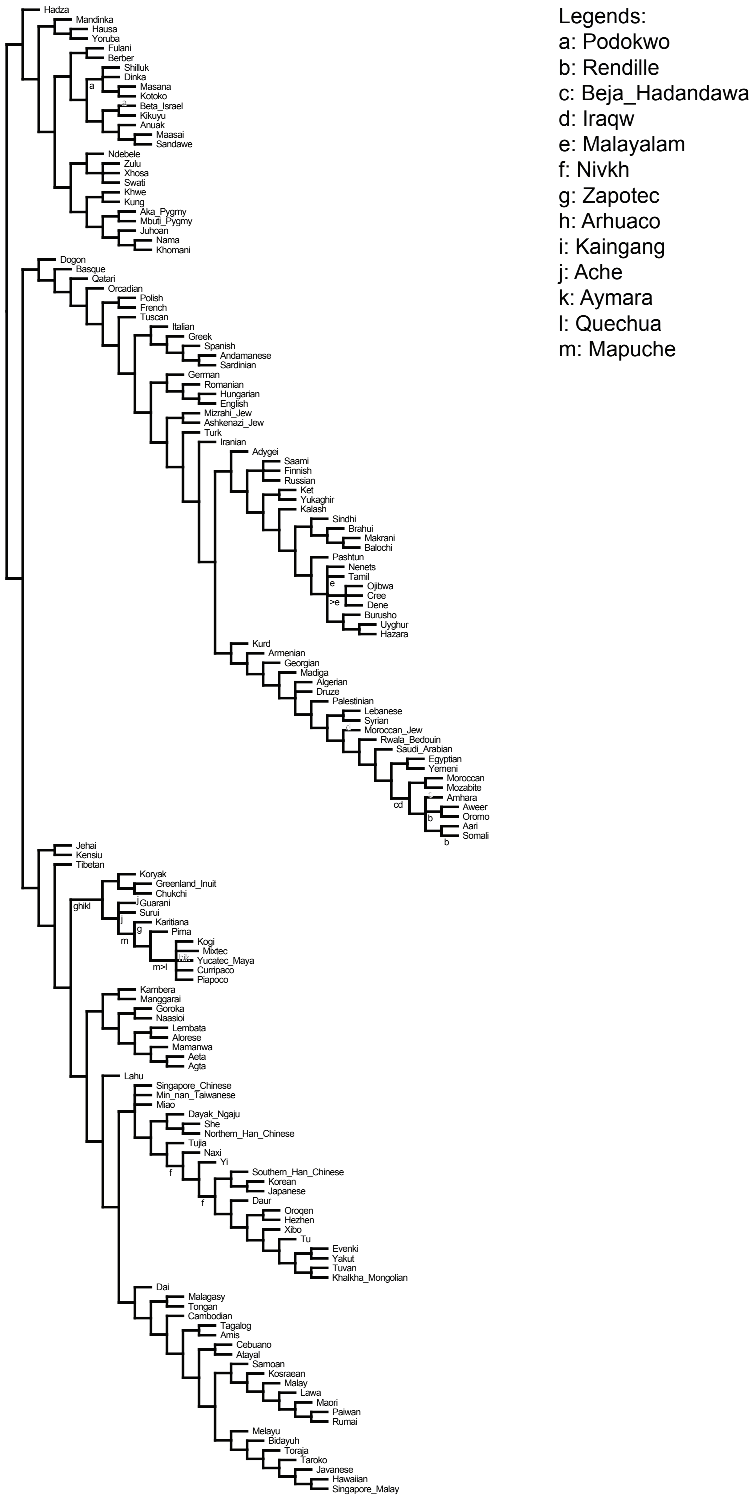
Supplementary Figure S18. Pruned strict consensus supertree based on parameter set 1.D of the sensitivity analysis of the representative dataset. Alternative positions of the identified wildcard taxa are indicated. A symbol ">" indicates wildcard taxa which would be placed within a polytomy, if included in the analysis.



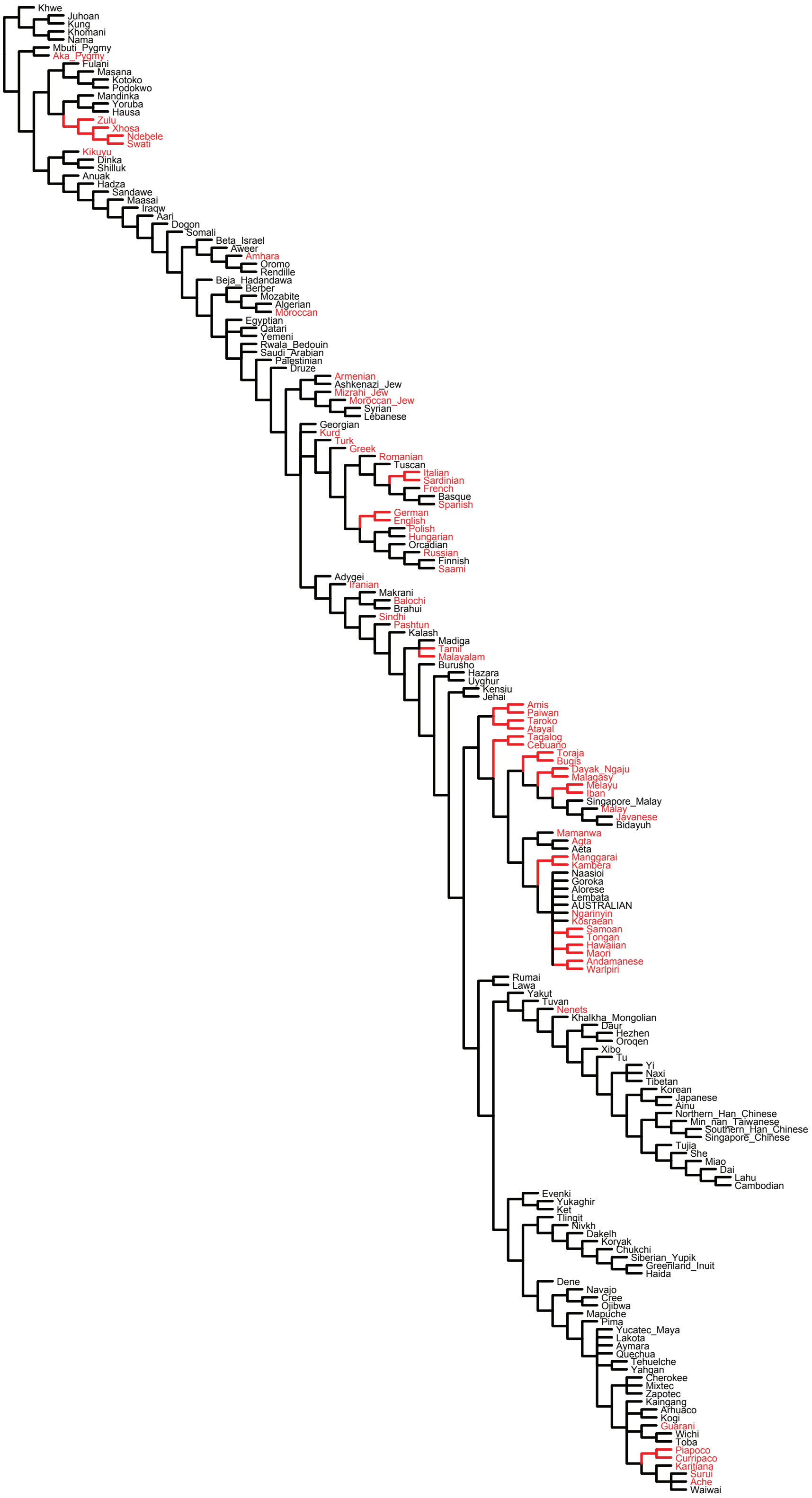
Supplementary Figure S19. Semistrict consensus supertree based on admixture plots.



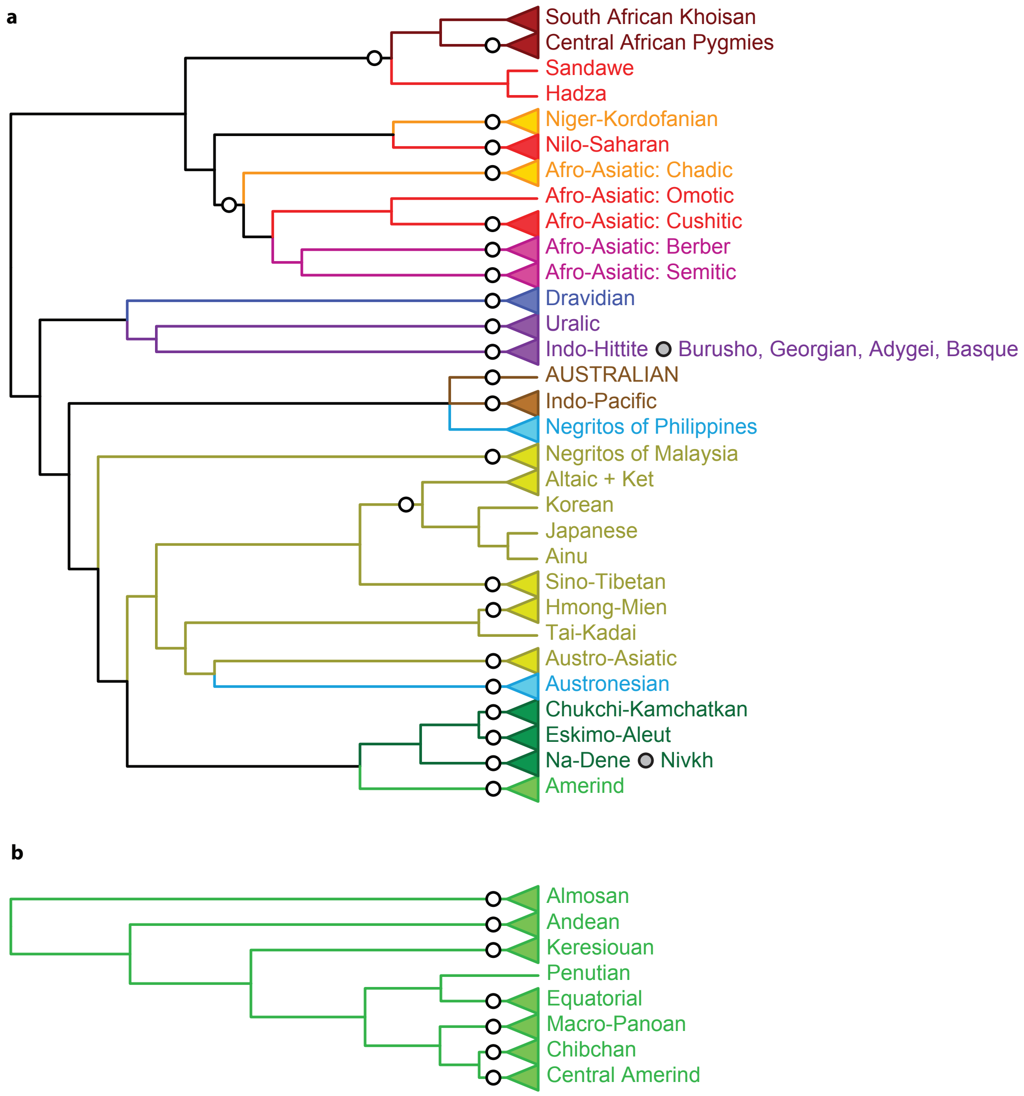
Supplementary Figure S20. Pruned strict consensus supertree based on admixture plots. Alternative positions of the identified wildcard taxa are indicated. A symbol ">" indicates wildcard taxa which would be placed within a polytomy, if included in the analysis.



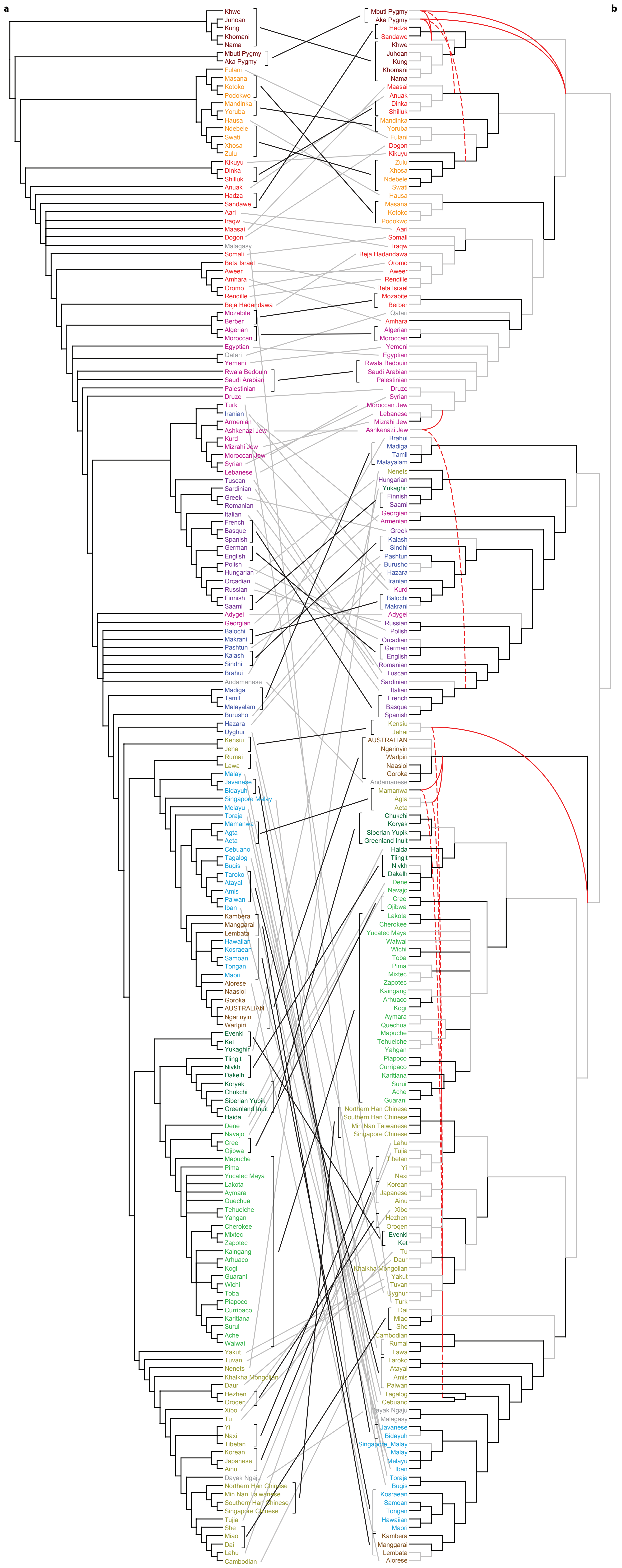
Supplementary Figure S21. Semistrict consensus supertree based on parameter set 1.A showing the taxa for which there are informative linguistic characters and the clades supported by linguistic characters (in red).



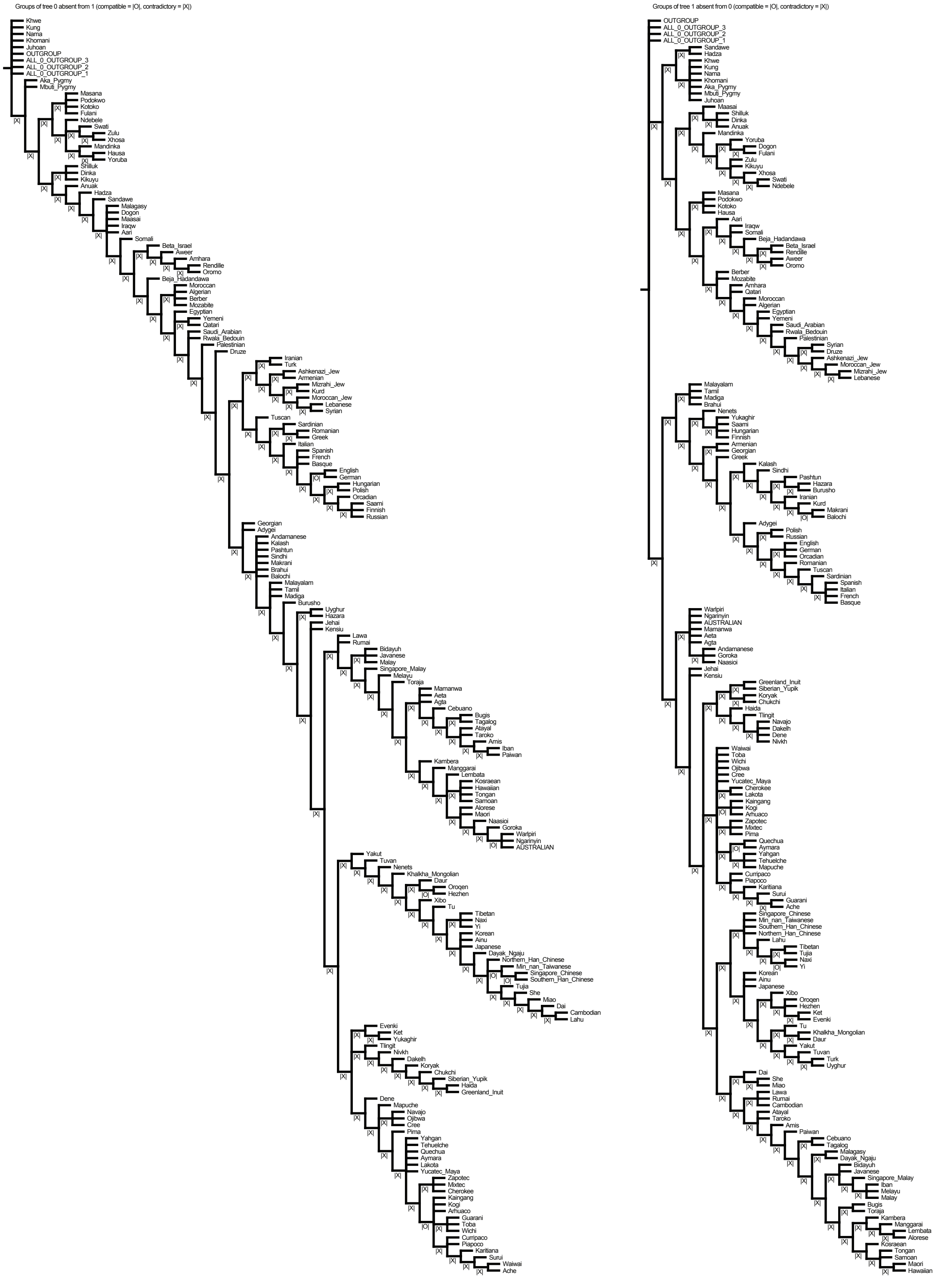
Supplementary Figure S22. (**a**) Supertree constrained by Greenberg-Ruhlen dataset based on linguistic classification from *Ethnologue*, Ruhlen, and Greenberg & Ruhlen. White circles indicate linguistic topological constraints. Grey circles indicate an unconstrained taxon or clade (usually a language isolate) recovered within a constrained one. (**b**) A section of the supertree constrained by linguistic classification of the Amerind macrofamily based on Greenberg & Ruhlen[2] but without a constraint on the monophyly of the Amerind itself.



Supplementary Figure S23. Tanglegram comparing supertrees based on a representative dataset. (**a**) The purely genetic supertree. (**b**) Galled network based on two supertrees constrained by linguistic classification based on Ruhlen[1]. Branches supported by linguistic information (i.e., informative characters based either on the linguistic classification datasets or on the linguistic source trees) are in black. Branches based solely on genetic information are in gray. Instances of recent language shifts in selected populations (C African Pygmies, "Negritos" of Malaysia and Philippines, and Ashkenazi Jews) are indicated by reticulations in red. Dashed red lines indicate the phylogenetic placement of the population based on its linguistic affiliation, solid red lines indicate the phylogenetic placement of the population based on genetic/genomic data. Brackets delimit corresponding monophyletic or paraphyletic groups in the two supertrees. Black lines connect corresponding (monophyletic or paraphyletic) groups, gray lines connect terminal taxa. Both trees are right-ladderized.



Supplementary Figure S24. (**a**) An "anticonsensus" tree showing groups present in the purely genetic supertree (Supplementary Figure S23a) that are absent from the supertree constrained by linguistic classification (Supplementary Figure S23b). (**b**) An "anticonsensus" tree showing groups in the supertree constrained by linguistic classification (Supplementary Figure S23b) that are absent from the genetic supertree (Supplementary Figure S23b). Polytomies indicate absence of conflict. A symbol "|O|" indicates compatible groups, a symbol "|X|" indicates contradictory groups.



SUPPLEMENTARY METHODS

Data collection and manipulation

Source trees. The source trees were searched on the electronic databases Web of Science, ScienceDirect, and Google Scholar. Complete bibliographies of all collected papers, as well as papers that cite them, were searched in order to find additional source trees.

Only the source trees that were inferred by formal phylogenetic methods and based on original analyses of real datasets (both genetic and linguistic) were included. To ensure accessibility of data and accountability of the supertree *sensu* Gatesy et al.[3] with respect to its source data, only trees published in peer-review sources (including web journals) were collected. The only exception was an unpublished extension[2] of the published article[4] and two conference papers[5,6]. Trees published exclusively as parts of doctoral dissertations were not included. The tree-like schemes derived solely from the formal linguistic classifications were not considered valid source trees. Trees and admixture plots that include, exclusively or predominantly, mixed ancestry populations such as African Americans, Mexican Mestizos, or Cape Coloreds[7,8] were not included, nor were the trees including only subpopulations of a single population[9,10].

The source trees were translated exactly as they appear in print. Standardization of taxonomic nomenclature and taxonomic level, as well as pruning of invalid taxa, was performed after the translation of a tree into matrix representation (MR).

Both rooted and unrooted trees of human populations or languages were utilized. Trees including individuals or individual genotypes of known ethnic population origin instead of ethnic populations[11–12] were utilized also. Replacement of individuals/individual genotypes with respective populations sometimes resulted in the MR implying that a terminal taxon is placed at multiple positions of a tree. These taxa were then fused together, which resulted in a less resolved tree.

The translation routine applies generally to all phylogenetic trees regardless of the method used for estimating phylogeny (i.e., UPGMA, neighbor-joining, maximum parsimony, maximum-likelihood, and Bayesian inference methods). The trees constructed by the program TreeMix[4] were also utilized, with the “admixture edges” omitted during translations to MR.

Admixture plots. In addition to the source trees, we utilized admixture plots for the first time as additional sources of data for the supertree construction. Admixture plots are graphical outputs of programs STRUCTURE[13], FRAPPE[14], and ADMIXTURE[15], developed to estimate individual ancestry and population structure on the basis of recombining genetic markers such as single nucleotide polymorphisms (SNPs), short tandem repeats (STRs), or sequence haplotypes. These programs have been utilized extensively in human population genetic studies. The STRUCTURE program has also been utilized to investigate the structure and history of human populations based on linguistic markers[16]. All the programs use iterative maximum-likelihood or Bayesian clustering algorithms that attribute individual genotypes of known ethnic population origin to K clusters such that Hardy–Weinberg equilibrium is maximized within the clusters. Individuals are given a membership coefficient for each cluster such that the estimated membership coefficient of each individual sums to 1 across K clusters.

The graphical outputs of these programs are plots indicating proportions of the individual genotypes attributable to K clusters by color. Although these programs do not model history explicitly, the inferred clusters can be interpreted post hoc as representing historical populations, and individuals or populations that are mixtures of different components as evidence of admixture between these populations[4]. Hierarchical information contained within the admixture plot was converted into matrix representation as follows. Each population was coded as "present" ("1") or absent ("0") based on proportions of individual genotypes attributable to each cluster. Limited attribution to a given cluster (less than ca. 10%) was neglected, and ambiguous sections of a plot (borderline proportions or different proportions in individuals within a single population) were scored as "unknown" ("?"). The resulting matrix of additive binary characters was analyzed by maximum parsimony to produce a phylogenetic tree corresponding roughly to population clustering implied by the admixture plot. The trees based on admixture plot often contained unresolved sections due to membership of some populations in several clusters, but they still preserve enough valuable branching information.

Controlling for data non-independence and duplication. It is common practice in human population history studies that both molecular and linguistic datasets are obtained from the literature or open databases and re-used in a novel analysis. The re-use of the old data is facilitated by on-line archiving molecular datasets including the Human Genome Diversity Project (HGDP) panel[17] or the International HapMap Project[18], and linguistic datasets available online, such as the Austronesian Basic Vocabulary Database (ABVD)[19].

Many molecular phylogenetic analyses of human populations are therefore extensions of previous studies. This applies especially to analyses based on genome-wide data and on human leukocyte antigen system (HLA). Also, analyses of the same linguistic groups are often based on identical sets of cognates. As a result, the same character information can contribute to more than one source tree. In all cases of data duplication, the overlap of character data between source studies means that the associated source trees are not independent of one another, a key assumption of phylogenetic analysis. Non-independence can arise among trees from different studies, as well as among trees presented within a single source study ("between-" and "within-study non-independence", respectively[20]). The consequence of data non-independence and the associated data duplication is that some data partitions are effectively upweighted and might affect the supertree topology disproportionally[3,21]. While the problem of data set non-independence cannot be eliminated entirely, it can be largely ameliorated using an appropriate source tree collection and exclusion protocol.

The most widely used protocol called "garbage in, garbage out"[20] is difficult to implement since there is no explicit dependence threshold between analyses, and its strict application neglects a huge amount of data and likely oversimplifies the existing phylogenetic controversies[22]. It is also likely to introduce bias in the supertree dataset due to exclusion and downweighting of matrices based on source trees derived from overlapping datasets. For these reasons, we used the so-called “less restricted” protocol for source tree retention and exclusion[22]. In this approach, studies derived from modifications of previous datasets (e.g., new scoring for certain taxa, addition of taxa or characters) were regarded as separate analyses. If a paper provides two or more different topologies, derived from modifications of the same dataset (e.g. addition/exclusion of few taxa, different scoring for some characters), these were included as separate source trees. However, secondary representations of a tree taken from another study were not considered valid source trees.

Standardization of taxonomic nomenclature and taxonomic level. Merging MRs of the source trees by terminal taxa requires prior standardization of taxonomic nomenclature. Standardizing taxonomic nomenclature throughout the source trees is a daunting task since there is no universally accepted taxonomic nomenclature of human populations. Numerous alternate names of human ethno-linguistic groups (ethnonyms) exist[23], reflecting differences between population and language names, between auto/endonyms and exonyms, political influences, or merely different spelling.

Taxonomic nomenclature was standardized using ISO 639-3 codes from *Ethnologue*[23], a reference work cataloging all of the world’s known living and recently extinct languages. Information on geographic range of a population in question, sampling location(s) of genotyped individual(s), language affiliation and ethnonyms was utilized in order to standardize taxonomy of the terminal taxa within and between source trees[23-29]. Data provided by Dediu & Ladd[30] were used to match groups of individuals included in HGDP panel[17] with the most likely corresponding ethno-linguistic groups.

Auto/endonyms were not systematically preferred over exonyms, since most of them are scarcely used and largely unknown to both the general public and expert audience. We avoided the use of exonyms that are nowadays considered derogatory (e.g., “Bushmen”, “Hottentot”, “Semang”, “Eskimo”, “Chipewyan”, “Mataco Indians”[23]).

The source trees often include taxa (or operational taxonomic units, OTUs) that are hierarchically nested one in another as different studies investigate human population structure and history on different levels. The taxa for molecular analyses are often less specified in comparison with taxa used in linguistic analyses. Nested taxa appear in published human population-level phylogenetic trees as well as in admixture plots. In some cases, taxa that are hierarchically inter-dependent appear within a single source tree (e.g., “S. Bantu” and “Xhosa” in Tishkoff et al.[31]).

The taxonomic level of the terminal taxa was standardized to correspond to ethno-linguistic groups listed in *Ethnologue*[23]. Lower-level taxa (i.e., subpopulations) took on the names of the corresponding higher-level taxa in *Ethnologue*. Higher-level taxa (i.e., above the level of groups listed in *Ethnologue*) were replaced by constituent lower-level taxa based on the information provided in the source study. When this information was unavailable, “type” population(s) were used in place of the higher taxon. They were selected with respect to the sampling location(s), language affiliation of the genotyped individual(s), and other available information (Supplementary Table S1).

Multiple taxa from a single source tree representing subpopulations were fused together to form a single taxon in the supertree dataset. Some subpopulations that are clearly distinguishable geographically and/or culturally and are sufficiently represented across the source trees (e.g., Cypriot Greek, Singapore Chinese) were retained in the supertree dataset. When two or more subpopulations were present in the source tree in too distant positions so that their fusion would lead to a great loss of the resolution of the tree, only the most representative subpopulation (with respect to the number of genotyped individuals and the sampling location) was retained in the supertree dataset.

Pruning invalid taxa. Matrix representations of the source trees with standardized taxonomic nomenclature and taxonomic level were merged into a single supertree dataset and further edited in Winclada ver. 1.00.08[32].

The dataset was trimmed to include only valid taxa. Decisions on what constitutes a valid taxon were guided by the appearance of the taxa in source trees, representation of these taxa across the source trees, and the ability to unambiguously associate these taxa with ethno-linguistic groups listed in *Ethnologue*[23].

Mixed ancestry populations (e.g., “African American”, “US Hispanic”, or “Cape Mixed Ancestry”) and colonial populations (e.g., Boer) were not considered valid taxa, but heavily admixed indigenous populations (e.g., Malagasy, Naasioi, Aleut, Yucatec Maya) were included. Extinct populations based solely on linguistic source trees were not considered, but the populations from genetic source trees that have recently lost their original languages (e.g., Zenu) were included. Creole languages (e.g., Haitian) and colonial forms of Indo-European languages (e.g., Afrikaans), based solely on linguistic source trees, were not considered valid, whereas indigenous populations speaking creole languages (e.g., Karretjie, Rabaul), present in genetic source trees, were retained. Loosely specified higher-level taxa in the source trees (e.g., “African”, “S.W. Asian”, or “Native North American”), for which constituent population(s) could not be identified nor type population(s) established, were excluded from the supertree dataset. The only exceptions were “Australian Aboriginals” and “Northern Australian Aboriginals” of unspecified ethnic population origin. Australia is poorly sampled and unspecified samples outnumber those that are specified in the source trees. All unspecified Australian and Northern Australian “Aboriginals” were fused together and analyzed as a single terminal taxon named “AUSTRALIAN”, together with the ethnically specified Australian Aboriginal populations. Taxa present in just one linguistic tree or in a single admixture plot were not included.

Population samples. The resultant supertree dataset (unpublished) included 973 valid taxa (human populations) and 5 great apes or archaic hominins that featured in the source trees (*Gorilla gorilla, Pan paniscus, Pan troglodytes,* Denisova hominin, *Homo neanderthalensis*). The representation of 973 taxa in the dataset is highly unequal: Cheyenne were scored only for five (0.05%) and Northern Han Chinese for 4,153 (44.28%) “characters” (“matrix elements”[33]) out of 9,380 phylogenetically informative characters.

In order to facilitate comprehension and visualization of the inferred human population relationship patterns, and to minimize the influence of unequal representation of individual taxa in the combined data matrix, we created two datasets based on restricted samples of the supertree dataset. The first dataset includes 186 populations and 5,717–5,987 phylogenetically informative characters, depending on outgroup selection ("representative dataset" hereinafter). This dataset included 52 populations from the HGDP panel[17,34]; 16 populations which were added to represent additional "ancestral components" identified on the basis of meta-analysis of 12 human genomic diversity projects by Shriner et al.[35]; and 16 more populations added to represent "ancestral components" that were admittedly missing in this meta-analysis[35] (i.e., Hadza, Sandawe, Fulani, Chadic, Cushitic, and Polynesian). Additional 102 populations were included in order to describe human diversity across world regions that are underrepresented in meta-analyses of human genomic diversity projects[35,36] (Island SE Asia, Melanesia, Australia, "Remote Oceania", Europe, Siberia, and North America). We included populations that seem to play a key role in understanding population expansions, as they could potentially represent sister or basal groups of the expanding population groups, e.g., populations of Caucasus (Armenian, Georgian, Adygei), Aboriginal Taiwanese (Amis, Paiwan, Taroko, Atayal), or E Siberians (Yukaghir, Ket, Nivkh). We also included language isolates and groups of debatable origin, e.g., Khwe, Andamanese, "Negritos" of Malaysia (Jehai and Kensiu) and Philippines (Agta, Aeta, and Mamanwa), Malagasy, Australian Aboriginals, Ainu, and Haida. The resulting dataset included both populations that are well represented across source trees, and based on overlapping genetic, genomic and linguistic trees and those that are poorly represented across source trees. Twenty-one populations in the dataset were not included in any source trees based on genome-wide data. These are mostly populations from Island South East Asia, Australia, and North America (Supplementary table S2). The least represented populations were Cherokee and Haida, both scored for 28 characters (0.47%), whereas the best represented Northern Han Chinese were scored for 3,272 characters (54.65%) out of 5,987 phylogenetically informative characters.

The second dataset was created in order to investigate the stability of principal phylogenetic groupings and to identify conflicts within the data, using only those populations that are best represented across the source trees. Its population sample consists of populations included in the HGDP panel[17,34] plus three additional populations: "AUSTRALIAN", Kosraean, and Samoan, representing Australia, Micronesia, and Polynesia, respectively ("HGDP dataset" hereinafter). This dataset included 52 populations and 2,874–3,070 phylogenetically informative characters, depending on outgroup selection. The least represented Kosraean were scored for 84 characters (5.99%), whereas the best represented Northern Han Chinese were scored for 2,250 characters (73.28%) out of 3,070 phylogenetically informative characters. Most populations in the dataset were scored for over 10% of phylogenetically informative characters.

Analyses

Phylogenetic analyses. Phylogenetic analyses were performed in TNT ver. 1.1[37]. The following tree-searching algorithm was used for each analysis:

```
xinact; hold 10000; xmult= level 10 fuse 5 drift 30 rss css xss rat 50;
bbreak= tbr safe; tsave trees.ctf; save; tsave /; comcomp*; tchoose /;
tsave supertree.ctf; save; tsave /; tchoose -.;
```

The dataset was analyzed under “new technology search” with search level 10 using sectorial, ratchet, and tree fusing searches, obtaining trees from a 10,000-replicate random addition sequence, treating gaps as missing data and all character changes as equal and nonadditive. The recovered most parsimonious trees (MPTs) were then subject to additional branch swapping using a tree-bisection and reconnection method. The datasets were analyzed without any topological constraints (i.e., without any assumptions on monophyly of any geographical region or linguistic group). Up to 10,000 MPTs and one semistrict consensus supertree was recovered in each of 16 parameter sets of the sensitivity analysis (see below). Tree length, consistency index (CI) and retention index (RI) for the resulting supertrees were calculated after removal of parsimony-uninformative characters.

Sensitivity analysis. One of the major drawbacks of the phylogenetic supertree method is the inapplicability of standard resampling or step-counting techniques for evaluating tree support (i.e., bootstrap and jackkife support, or Bremer support) and the general inability to measure uncertainty of inferred phylogenetic groupings (but see Bininda-Emonds[38] and Wilkinson et al.[39]).

To investigate stability of the inferred supertree topology, and to identify conflicts within the data (i.e., to identify alternative taxonomic groupings inherent to the data), we used a method inspired by sensitivity analysis *sensu* Wheeler[40]. Sensitivity analysis is carried out by selecting a set of parameters and examining the effect of these parameter sets on the stability of inferred tree topologies. The sensitivity analysis examines the influence of parameter sets on taxonomic groupings recovered in the analysis. The results are graphically represented using so called “analysis space plots”[40] or “sensitivity plots”[41]. In this way, a sensitivity analysis can discern between robust clades (those that appear under most or all parameter set combinations) and unstable ones (those that appear only under one or few parameter sets).

The supertree dataset was analyzed under 16 different sets of values for the analysis parameters. These 16 analyses consisted of combinations of four weighting schemes and four rooting options. Successive downweighting of data partitions and rooting by different outgroups leads to weakening the influence of respective data partitions on overall topology of the supertree. This allows to assess the support for various phylogenetic groupings across various types of data and to identify the causes of topological conflicts. The weighting schemes and rooting options were as follows:

Either (1) all data partitions were weighted equally, or (2) genetic and linguistic trees (5,066 phylogenetically informative characters) were upweighted by a factor of 1,000 relative to admixture plots; or (3) genetic trees (4,428 phylogenetically informative characters) were upweighted relative to linguistic trees and admixture plots; or (4) genomic trees data (1,704 phylogenetically informative characters) were upweighted relative to all the remaining data partitions; and either (A) all rooted source trees and admixture plots were treated as rooted (by inserting a hypothetical „all-0“ outgroup), or (B) only the trees were treated as rooted, or (C) only the genetic trees were treated as rooted, or (D) only source trees that featured real great ape and/or archaic hominin outgroups (*Gorilla gorilla*, *Pan paniscus*, *P. troglodytes*, Denisova hominin, *Homo neanderthalensis*) (a total of 28 source trees) were treated as rooted in this analysis.

When performing sensitivity analysis of HGDP dataset, the data partitions were both successively downweighted and successively deactivated to test whether the weighting scheme used for sensitivity analysis of the representative dataset (1:1,000) is sufficient for minimizing the effect of downweighted data partitions on the resulting supertree topologies.

Wildcard taxa identification. Taxa that possess a limited amount of informative characters can act as "wildcard" taxa[42]. Wildcard taxa can adopt multiple positions in optimal topologies, which lead to poorly resolved consensus trees with large polytomies that hamper the interpretation of the phylogenetic results.

Each set of MPTs recovered in the sensitivity analysis of the full dataset was analyzed using the *Iter*PCR script[43], implemented in TNT, to improve the resolution of the consensus tree by identifying wildcard taxa. Altogether, 24 wildcard taxa were identified (Supplementary Table S5). Alternative positions of the identified wildcard taxa were investigated using reduced strict consensus (*nelsen//*) in TNT that displays all alternative positions of the pruned taxa within a single consensus tree: (1) by comparing alternative positions of wildcards in each set of MPTs recovered under each parameter set (i.e., in a total of 16,000 trees); and (2) by comparing positions of wildcards across supertrees based on 16 parameter sets (i.e., in 16 consensus trees).

Four wildcards causing the greatest loss of supertree resolution were excluded from the dataset. The pruned version of the representative dataset (182 OTUs and 5,676–5,949 phylogenetically informative characters, depending on outgroup selection) was used for subsequent analyses. In the resulting supertrees based on sensitivity analysis (Fig. 1), the excluded wildcards were displayed in the basalmost position of all the positions they acquire when included in the dataset, but they were not taken into account when assessing node support and group support.

Linguistic classification. In order to measure congruence between the supertree and linguistic classification and to infer relationships of the language families and "macrofamilies", two datasets based on formal linguistic classifications were created to be optimized on the supertree and to serve as a "linguistic scaffold" (compare to "morphological scaffold" *sensu* Springer[44]). The inclusion of linguistic classification was necessary for inferring a well-resolved tree which shows the relationships of language families, considering the taxonomic coverage of published language phylogenetic trees is very limited, with most of them covering Indo-European, Austronesian, or Bantu language families.

The first classification-based dataset included 37 parsimoniously informative characters derived from *Ethnologue*[23] on the level of language families. The controversial higher-level groups within *Ethnologue* (i.e., Khoisan, Afro-Asiatic, Altaic, and Australian) were not used for constraining the supertree topology. The second dataset included an additional 26 parsimoniously informative characters based on linguistic classification by Ruhlen[1] and Greenberg & Ruhlen[2] on the highest levels of language macrofamilies and linguistic stocks. Highly controversial linguistic macrofamilies (e.g., “Eurasiatic”/“Nostratic”, “Dene-Caucasian”) were not included in this dataset (Supplementary Table S6). Consequently, several populations were unclassified in both these datasets, scored entirely using “?” (these included Adygei, Basque, Burusho and Ket classified as “Dene-Caucasian“, Ainu and Nivkh classified as “Eurasiatic“ by Ruhlen[1]). Both datasets were fully congruent (i.e., there was no hard conflict between them).

Hunter-gatherer populations speaking languages of neighboring agriculturalists, as a result of relatively recent language shifts, were scored as “unknown” (“?”) in the linguistic classification dataset in order to avoid inserting unnecessary dramatic conflicts between genetic and linguistic information. They included C African Pygmies who speak Nigero-Kordofanian or Nilo-Saharan languages[45], Jehai and Kensiu (the “Negrito” populations of Malaysia) who speak Austro-Asiatic languages[46,47], and Agta, Aeta and Mamanwa (the “Negritos” of Philippines) speaking Malayo-Polynesian Austronesian languages[48]. Similarly, Ashkenazi Jews, who traditionally speak Germanic Indo-European language (Yiddish), were also scored using “?” as to avoid conflict between grouping them within the C–E European and Middle East peoples. Alternatively, the above mentioned hunter-gatherer populations were scored according to their present-day language affiliation, and Ashkenazi Jews were classified as Indo-Europeans.

In order to compare genetic and linguistic information, the purely genetic supertree was inferred by analyzing data partitions based exclusively on genetic/genomic source trees and genomic admixture plots. The language-constrained supertrees were inferred by analyzing all data partitions (including language-based trees and admixture plots) together with *Ethnologue* and Greenberg-Ruhlen datasets, based on linguistic classification. The linguistic data partitions and the linguistic classification datasets were upweighted by a factor of 1,000 relative to the remaining data partitions. All language-constrained supertrees were inferred under parameter set 1.A of the sensitivity analysis (see above).

Measuring incongruence between supertrees. The resulting supertree topologies were compared using subtree prune and regraft (SPR) distance measure (*sprdiff*), using the following algorithm:

```
sprdiff: noviol; sprdiff 0 1 1000x100;
```

Topology of the supertree constrained by Greenberg–Ruhlen classification was compared with a purely genetic supertree using a “tanglegram”[49] computed in Dendroscope ver. 3.2.10[50]. Language-constrained supertrees, based on the two versions of the Greenberg–Ruhlen dataset (Supplementary Fig. S23), were reconciled, using galled networks[51] computed in Dendroscope to visualize language shifts of selected populations within a single tree. Purely genetic and language-constrained supertrees were additionally compared using an “anticonsensus” measure (*tcomp*) in TNT showing compatible and contradictory groups present in one tree but not in another, and vice versa (Supplementary Fig. S24).

In order to assess the support for proposed linguistic groupings (language macrofamilies, linguistic stocks, and language families), consistency index (CI) and retention index (RI) values were calculated in Mesquite ver. 3.02[52] for each character in the linguistic classification datasets optimized onto the purely genetic and combined supertree topologies based on parameter set 1.A of the sensitivity analysis. The resulting CI values were compared to the minimum possible CI values (for a binary character, $CI_{min} = 1/N$, where N taxa were scored positively for presence of a character), which made the values directly comparable for language families represented by different numbers of taxa.

Supplementary Methods References

1 Ruhlen, M. *Guide to the World's Languages: Classification*. Vol. 1 (Stanford University Press, 1991).

2 Greenberg, J. H. & Ruhlen, M. *An Amerind Etymological Dictionary*. (Stanford University, Department of Anthropological Sciences, 2007).

3 Gatesy, J., Matthee, C., DeSalle, R. & Hayashi, C. Resolution of a supertree/supermatrix paradox. *Systematic Biology* **51**, 652-664, doi:10.1080/10635150290102311 (2002).

4 Pickrell, J. K. & Pritchard, J. K. Inference of population splits and mixtures from genome-wide allele frequency data. *PLOS Genetics* **8**, doi:10.1371/journal.pgen.1002967 (2012).

5 Rama, T. & Singh, A. K. in *International Conference RANLP 2009* 355–359 (Borovets, Bulgaria, 2009).

6 Geisler, H. & List, J.-M. in *Arbeitstagung der Indogermanischen Gesellschaft 2009: Die Ausbreitung des Indogermanischen Thesen aus Sprachwissenschaft, Archäologie und Genetik* (Würzburg, Germany, 2009).

7 Bryc, K. *et al.* Genome-wide patterns of population structure and admixture among Hispanic/Latino populations. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 8954-8961, doi:10.1073/pnas.0914618107 (2010).

8 Moreno-Estrada, A. *et al.* Reconstructing the population genetic history of the Caribbean. *PLOS Genetics* **9**, 19, doi:10.1371/journal.pgen.1003925 (2013).

9 Lee, S. & Hasegawa, T. Evolution of the Ainu language in space and time. *PLOS ONE* **8**, 6, doi:10.1371/journal.pone.0062243 (2013).

10 Serva, M., Petroni, F., Volchenkov, D. & Wichmann, S. Malagasy dialects and the peopling of Madagascar. *Journal of the Royal Society Interface* **9**, 54-67, doi:10.1098/rsif.2011.0228 (2012).

11 Haber, M. *et al.* Genome-wide diversity in the levant reveals recent structuring by culture. *PLOS Genetics* **9**, 8, doi:10.1371/journal.pgen.1003316 (2013).

12 Lachance, J. *et al.* Evolutionary history and adaptation from high-coverage whole-genome sequences of diverse African hunter-gatherers. *Cell* **150**, 457-469, doi:10.1016/j.cell.2012.07.009 (2012).

13 Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945-959 (2000).

14 Tang, H., Peng, J., Wang, P. & Risch, N. J. Estimation of individual admixture: Analytical and study design considerations. *Genetic Epidemiology* **28**, 289-301, doi:10.1002/gepi.20064 (2005).

15 Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Research* **19**, 1655-1664, doi:10.1101/gr.094052.109 (2009).

16 Reesink, G., Singer, R. & Dunn, M. Explaining the linguistic diversity of Sahul using population models. *PLOS Biology* **7**, doi:10.1371/journal.pbio.1000241 (2009).

17 Cann, H. M. *et al.* A human genome diversity cell line panel. *Science* **296**, 261-262 (2002).

18 Gibbs, R. A. *et al.* The International HapMap Project. *Nature* **426**, 789-796, doi:10.1038/nature02168 (2003).

19 Greenhill, S. J., Blust, R. & Gray, R. D. The Austronesian Basic Vocabulary Database: From Bioinformatics to Lexomics. *Evolutionary Bioinformatics* **4**, 271-283 (2008).

20 Bininda-Emonds, O. R. *et al.* in *Phylogenetic Supertrees: Combining Information to Reveal the Tree of Life* (ed Olaf RP Bininda-Emonds) Ch. 12, 267-280 (Kluwer Academic Publishers, 2004).

21 Springer, M. S. & de Jong, W. W. Which mammalian supertree to bark up? *Science* **291**, 1709-+, doi:10.1126/science.1059434 (2001).

22 Bronzati, M., Montefeltro, F. C. & Langer, M. C. A species-level supertree of Crocodyliformes. *Historical Biology* **24**, 598-606, doi:10.1080/08912963.2012.662680 (2012).

23 Lewis, M., Simons, G. & Fennig, C. *Ethnologue: Languages of the World, Seventeenth Edition*. (SIL international Dallas, TX, 2013).

24 White, D. R. Focused ethnographic bibliography: Standard Cross-Cultural Sample. *Cross-Cultural Research* **23**, 1-145 (1989).

25 Gray, J. P. A corrected ethnographic atlas. *World Cultures* **10**, 24-85 (1999).

26 Maho, J. *NUGL online: the online version of the New Updated Guthrie List, a referential classification of the Bantu languages*, <goto.glocalnet.net/mahopapers/nuglonline.pdf> (2009).

27 Dryer, M. S. & Haspelmath, M. *The World Atlas of Language Structures Online*, <http://wals.info> (2013).

28 *The Joshua Project*, <www.joshuaproject.net> (

29 White, D. R. Pinpointing sheets for the Standard Cross-Cultural Sample: complete edition. *World Cultures eJournal* **17** (2009).

30 Dediu, D. & Ladd, D. R. Linguistic tone is related to the population frequency of the adaptive haplogroups of two brain size genes, ASPM and Microcephalin. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 10944-10949, doi:10.1073/pnas.0610848104 (2007).

31 Tishkoff, S. A. *et al.* The Genetic structure and history of Africans and African Americans. *Science* **324**, 1035-1044, doi:10.1126/science.1172257 (2009).

32 Nixon, K. C. WinClada v. 1.00.08 (Published by the author, Ithaca, NY, 2002).

33 Baum, B. R. & Ragan, M. A. Reply to A.G. Rodrigo's 'A comment on Baum's method for combining phylogenetic trees'. *Taxon* **42**, 637-640 (1993).

34 Li, J. Z. *et al.* Worldwide human relationships inferred from genome-wide patterns of variation. *Science* **319**, 1100-1104, doi:10.1126/science.1153717 (2008).

35 Shriner, D., Tekola-Ayele, F., Adeyemo, A. & Rotimi, C. N. Genome-wide genotype and sequence-based reconstruction of the 140,000 year history of modern human ancestry. *Scientific Reports* **4**, doi:10.1038/srep06055 (2014).

36 Pemberton, T. J., DeGiorgio, M. & Rosenberg, N. A. Population structure in a comprehensive genomic data set on human microsatellite variation. *G3-Genes Genomes Genetics* **3**, 891-907, doi:10.1534/g3.113.005728 (2013).

37 Goloboff, P. A., Farris, J. S. & Nixon, K. C. TNT, a free program for phylogenetic analysis. *Cladistics* **24**, 774-786 (2008).

38 Bininda-Emonds, O. R. P. Novel versus unsupported clades: assessing the qualitative support for clades in MRP supertrees. *Systematic Biology* **52**, 839-848, doi:10.1080/10635150390252242 (2003).

39 Wilkinson, M., Pisani, D., Cotton, J. A. & Corfe, I. Measuring support and finding unsupported relationships in supertrees. *Systematic Biology* **54**, 823-831, doi:10.1080/10635150590950362 (2005).

40 Wheeler, W. C. Sequence alignment, parameter sensitivity, and the phylogenetic analysis of molecular data. *Systematic Biology* **44**, 321-331, doi:10.2307/2413595 (1995).

41 Schulmeister, S., Wheeler, W. C. & Carpenter, J. M. Simultaneous analysis of the basal lineages of Hymenoptera (Insecta) using sensitivity analysis. *Cladistics-Int. J. Willi Hennig Soc.* **18**, 455-484, doi:10.1016/s0748-3007(02)00100-7 (2002).

42 Nixon, K. C. & Wheeler, Q. D. in *Extinction and Phylogeny* (eds M.J. Novacek & Q.D. Wheeler) 119-143 (Columbia University Press, 1993).

43 Pol, D. & Escapa, I. H. Unstable taxa in cladistic analysis: identification and the assessment of relevant characters. *Cladistics* **25**, 515-527, doi:10.1111/j.1096-0031.2009.00258.x (2009).

44 Springer, M. S., Teeling, E. C., Madsen, O., Stanhope, M. J. & de Jong, W. W. Integrated fossil and molecular data reconstruct bat echolocation. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 6241-6246, doi:10.1073/pnas.111551998 (2001).

45 Bahuchet, S. Changing language, remaining pygmy. *Human Biology* **84**, 11-43 (2012).

46 Burenhult, N., Kruspe, N. & Dunn, M. in *Dynamics of Human Diversity: The Case of Mainland Southeast Asia* (ed N.J. Enfield) Ch. 11, 257-275 (Pacific Linguistics, 2011).

47 Dunn, M., Kruspe, N. & Burenhult, N. Time and place in the prehistory of the Aslian languages. *Human Biology* **85**, 383-399 (2013).

48 Reid, L. A. Who are the Philippine negritos? Evidence from language. *Human Biology* **85**, 329-358 (2013).

49 Planet, P. J. Tree disagreement: measuring and testing incongruence in phylogenies. *Journal of Biomedical Informatics* **39**, 86-102, doi:10.1016/j.jbi.2005.08.008 (2006).

50 Huson, D. H. & Scornavacca, C. Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. *Systematic Biology* **61**, 1061-1067, doi:10.1093/sysbio/sys062 (2012).

51 Huson, D. H., Rupp, R., Berry, V., Gambette, P. & Paul, C. Computing galled networks from real data. *Bioinformatics* **25**, I85-I93, doi:10.1093/bioinformatics/btp217 (2009).

52 Maddison, W. P. & Maddison, D. R. Mesquite: a modular system for evolutionary analysis v. 3.02 (2015).

Appendix II

Supplementary information for Chapter II

Hunter-Gatherers and the Origins of Religion
Hervey C. Peoples, Pavel Duda, and Frank W. Marlowe

Electronic Supplementary Materials (ESM)
Human Nature 27(3), doi: 10.1007/s12110-016-9260-0

Contents

Methods

Supertree dataset manipulations

Some sample populations are underrepresented or absent from the supertree dataset. Many hunter-gatherer populations live in areas that are difficult to access and some of them, e.g., Vedda, Botocudo, and Yahgan, are nowadays extinct or near-extinct. Taxa possessing insufficient amount of character information can act as so called 'wildcard' taxa *sensu* (Nixon and Wheeler 1993). These taxa adopt multiple positions in optimal topologies leading to poorly resolved consensus trees with large polytomies that hamper the interpretation and further utilization of the phylogenetic results. Population absent from the supertree dataset were replaced by more inclusive (i.e., higher level) taxon or by genetically closely related taxon (population) present in the dataset that was used as a proxy for the population in question (ESM Table A2). Positions of four North American populations – Kaska, Eyak, Twana, and Yurok – were based solely on linguistic classification (ESM Table A2). Six populations not present in the source trees for which sufficient linguistic classification was not available – Ingalik, Micmac, E. Pomo, Yokuts (Lake), Klamath, and Kutenai – were excluded from the dataset. The analysis of most parsimonious trees (MPTs) using the IterPCR script (Pol and Escapa 2009) implemented in TNT (Goloboff et al. 2008) identified these populations as wildcard taxa, supporting their exclusion.

Some hunter-gatherer populations act as wildcard taxa because of conflicting information implied by the source trees. This conflict is often caused by recent genetic admixture with (often distantly related) immigrant populations. For example Aleut of southwestern Alaska who have undergone a pervasive admixture with Russian colonizers and Scandinavian and English fishermen since the Russian contact in 1741. As a consequence, underlying patterns of genetic structure of the Aleut population are obscured (Rubicz et al. 2010a, b). Genetic admixture between hunter-gatherers and agriculturalists is often sex-biased as documented for Central African Pygmies and Bantu immigrants (Batini et al. 2011; Quintana-Murci et al. 2008; Verdu et al. 2009), causing conflicts between source trees based on maternally, paternally, and biparentally inherited genetic markers. Contacts between hunter-gatherers and agriculturalists are often followed by some degree of cultural assimilation that can include language shifts (Bahuchet 2012).

In order to overcome the problem of the lack of data and the conflicting signals caused by recent genetic admixture and language shifts in some hunter-gatherer populations in the study sample, the characters based on linguistic classifications were up-weighted by a factor of 100 to serve as a topological constraint or 'scaffold'. This 'linguistic scaffold' (compare to 'molecular scaffold' *sensu* Springer et al. (2001)) constrains the topology for a subset of populations for which linguistic affiliation can be determined (i.e., those scored for characters). Language isolates according to Ethnologue classification (Lewis et al. 2013) were scored entirely using '?'. Languages of American hunter-gatherers classified were classified as Amerindian (Ruhlen 1991) merely to ensure their presence within the American clade.

Hunter-gatherer populations speaking languages of agriculturalists as a result of relatively recent language shift were scored entirely using '?' in the scaffold tree. This allowed these populations to adopt a position on the MRP supertree based on contributing source trees alone. These populations include Mbuti and Aka Pygmies who speak Niger- Kordofanian and Nilo-Saharan languages (Bahuchet 2012), Vedda of Sri Lanka who speak Indo-European language (Dharmadasa 1974), Semang, the Negritos of Malaysia who speak Austro-Asiatic (Aslian) language (Burenhult et al. 2011; Dunn et al. 2013), and Agta, the Negritos of Philippines who speak Malayo-Polynesian (Austronesian) language (Reid 2013).

The linguistic scaffold tree included 20 phylogenetically informative characters for the 33 populations in the study sample. Note that this linguistic scaffold implied relatively few internal groupings (clades based on linguistic classification), particularly among the Old World hunter-gatherers (ESM Fig. A1).

References

Bahuchet, S. (2012). Changing Language, Remaining Pygmy. *Human Biology, 84*(1), 11–43.

Batini, C., Lopes, J., Behar, D. M., Calafell, F., Jorde, L. B., van der Veen, L., et al. (2011). Insights into the Demographic History of African Pygmies from Complete Mitochondrial Genomes. *Molecular Biology and Evolution, 28*(2), 1099–1110.

Burenhult, N., Kruspe, N., & Dunn, M. (2011). Language history and culture groups among Austroasiatic-speaking foragers of the Malay Peninsula. In N. J. Enfield (Ed.), *Dynamics of Human Diversity: The Case of Mainland Southeast Asia* (pp. 257–275). Canberra: Pacific Linguistics.

Dharmadasa, K. (1974). The creolization of an aboriginal language: The case of Vedda in Sri Lanka (Ceylon). *Anthropological Linguistics, 16*(2), 79–106.

Dunn, M., Kruspe, N., & Burenhult, N. (2013). Time and Place in the Prehistory of the Aslian Languages. *Human Biology, 85*(1–3), 383–399.

Goloboff, P. A., Farris, J. S., & Nixon, K. C. (2008). TNT, a free program for phylogenetic analysis. *Cladistics, 24*(5), 774–786.

Greenberg, J. H., & Ruhlen, M. (2007). *An Amerind Etymological Dictionary*. Stanford, CA: Stanford University Press.

Lewis, M., Simons, G., & Fennig, C. (2013). *Ethnologue: Languages of the world.* (Seventeenth ed.). Dallas, TX: SIL International. Available online at *http://www.ethnologue.com*

Nixon, K. C., & Wheeler, Q. D. (1993). Extinction and the origin of species. In M. J. Novacek, & Q. D. Wheeler (Eds.), *Extinction and Phylogeny* (pp. 119–143). New York: Columbia University Press.

Pol, D., & Escapa, I. H. (2009). Unstable taxa in cladistic analysis: identification and the assessment of relevant characters. *Cladistics, 25*(5), 515–527.

Quintana-Murci, L., Quach, H., Harmant, C., Luca, F., Massonnet, B., et al. (2008). Maternal traces of deep common ancestry and asymmetric gene flow between Pygmy hunter-gatherers and Bantu-speaking farmers. *Proceedings of the National Academy of Sciences, 105*(5), 1596–1601.

Reid, L. A. (2013). Who Are the Philippine Negritos? Evidence from Language. *Human Biology,* 85(1–3), 329–358.

Rubicz, R., Melton, P. E., Spitsyn, V., Sun, G. Y., Deka, R., & Crawford, M. H. (2010a). Genetic Structure of Native Circumpolar Populations Based on Autosomal, Mitochondrial, and Y Chromosome DNA Markers. *American Journal of Physical Anthropology, 143*(1), 62–74.

Rubicz, R., Zlojutro, M., Sun, G., Spitsyn, V., Deka, R., Young, K. L., et al. (2010b). Genetic Architecture of a Small, Recently Aggregated Aleut Population: Bering Island, Russia. *Human Biology, 82*(5–6), 719–736.

Ruhlen, M. (1991). *A guide to the world's languages: classification* (Vol. 1). Stanford, CA: Stanford University Press.

Springer, M. S., Teeling, E. C., Madsen, O., Stanhope, M. J., & de Jong, W. W. (2001). Integrated fossil and molecular data reconstruct bat echolocation. *Proceedings of the National Academy of Sciences of the United States of America, 98*(11), 6241–6246.

Verdu, P., Austerlitz, F., Estoup, A., Vitalis, R., Georges, M., Théry, S., et al. (2009). Origins and Genetic Diversity of Pygmy Hunter-Gatherers from Western Central Africa. *Current Biology, 19*(4), 312–318.

Table A1a. Sample of hunter-gatherer societies: matrix of religiosity characters

SCCS/ **EA** Number	Society Name	References	Animism	Afterlife	Shamanism	Ancestor Worship	High Gods	Active Ancestor Worship	Active High Gods
2	!Kung	1–7	1	1	1	0	1	0	1
—	G/wi	8, 9	1	1	1	0	1	0	1
9	Hadza	1, 10	1	0	0	0	0	0	0
301	Sandawe	11	1	1	0	1	0	0	0
13	Mbuti	1–3, 12, 13	1	0	0	0	0	0	0
—	Aka	14, 15	1	0	0	0	0	0	0
77	Semang	1–4,16, 17	1	1	1	0	1	0	1
79	Andamanese	1, 4, 18	1	1	1	0	0	0	0
80	Vedda	1, 4, 19, 20	1	1	1	1	0	1	0
86	Badjau Tawi	21, 22	1	1	1	0	0	0	0
—	Agta	23, 24	1	0	1	0	1	0	0
90	Tiwi	1, 25–27	1	1	0	1	1	0	0
91	Aranda	1, 4, 28–30	1	1	1	1	0	0	0
1177	Walbiri	31	1	1	1	1	0	0	0
118	Ainu	1, 32–35	1	1	1	1	0	1	0
119	Gilyak	1, 36, 37	1	1	1	1	1	1	0
120	Yukaghir	38	1	1	1	1	1	1	0
123	Aleut	1, 39–41	1	1	1	1	1	1	0
124	Copper Eskimo	1, 28, 42–46	1	1	1	1	0	1	0
125	Montagnais	1–3, 47–49	1	1	1	0	0	0	0
127	Salteaux	1, 4, 50, 51	1	1	1	1	0	0	0
128	Slave	1, 52	1	0	1	0	1	0	0
129	Kaska	1–3, 53, 54	1	1	1	0	0	0	0
130	Eyak	1, 55	1	1	1	0	0	0	0
131	Haida	1, 56–61	1	1	1	0	0	0	0
132	Bellacoola	1, 62, 63	1	1	1	1	1	0	0
133	Twana	1–3, 64, 65	1	1	1	0	0	0	0
134	Yurok	1, 28, 64, 66	1	1	1	0	0	0	0
162	Warrau	1, 67, 68	1	1	1	1	1	1	1
173	Siriono	1–3, 69	1	0	0	0	1	0	0
178	Botocudo	4, 70, 71	1	0	0	0	0	0	0
180	Aweikoma	1, 4, 72–74	1	1	1	1	0	1	0
186	Yahgan	1, 28, 75	1	1	1	1	1	0	1

Present = 1, absent = 0

Table A1b. References for 33 hunter-gatherer societies religiosity characters

1. Murdock G. P., & White, D. R. (1980). Standard cross-cultural sample. In H. Barry & A. Schlegel (Eds.), *Cross-cultural samples and codes* (pp. 3–43). Pittsburgh: University of Pittsburgh Press.
2. Winkelman, M. (1990). Shamans and other "magico-religious" healers: a cross-cultural study of their origins, nature, and social transformations. *Ethos*, *18*(3), 308–352.
3. Winkelman, M. & White, D. (1987). A cross-cultural study of magico-religious practitioners and trance states: data base. In D. Levinson and R. Wagner (Eds.), *Human relations area files research series in quantitative cross-cultural data* (Vol. 3). New Haven: HRAF Press.
4. Sheils, D. (1975). Toward a unified theory of ancestor worship: a cross-cultural study. *Social Forces*, *54*(2), 427–440.
5. Marshall, L. (1962). !Kung bushman religious beliefs*. Africa: Journal of the International African Institute*, *32*(3), 221–252.
6. Marshall, L. (1969). The medicine dance of the !Kung Bushmen. *Africa: Journal of the International African Institute*, *39*(4), 347–381.
7. Marshall, L. (1976*). The !Kung of Nyae Nyae*. Cambridge, MA: Harvard University Press.
8. Barnard, A. (1992). *Hunters and herders of Southern Africa: a comparative ethnography of the Khoisan peoples.* Cambridge, UK: Cambridge University Press.
9. Silberbauer, G. B. (1981). *Hunter and habitat in the central Kalahari Desert*. Cambridge: Cambridge University Press.
10. Marlowe, F. W. (2010). *The Hadza: hunter-gatherers of Tanzania*. Berkeley: University of California Press.
11. Raa, E. T. (1969). The moon as a symbol of life and fertility in Sandawe thought. *Africa*: *Journal of the International African Institute*, *39*(1), 24–53.
12. Turnbull, C. M. (1965). The Mbuti Pygmies: an ethnographic survey. *Anthropological Papers of the AMNH*, *50*, 139-282.
13. Turnbull, C. M. (1965). *Wayward servants: the two worlds of the African Pygmies*. Garden City, NY: The Natural History Press.
14. Hewlett, B. S. (1993*). Intimate fathers: the nature and context of Aka Pygmy paternal infant care*. Ann Arbor: University of Michigan Press.
15. Sawada, M. (2001). Rethinking methods and concepts of anthropological studies on African Pygmies' world view: the creator-god and the dead. *African Study Monographs, 27* (supplementary issue), 29–42.
16. Endicott, K. M. (1979). *Batek Negrito religion: the world-view and rituals of a hunting and gathering people of Peninsular Malaysia*. Oxford: Oxford University Press.
17. Schebesta, P., & Schutze, F. (1957). *The Negritos of Asia*. Wien-Mödling: St.-Gabriel- Verlag.
18. Radcliffe-Brown, A. R. (1922). *The Andaman Islanders: a study in social anthropology*. Cambridge: Cambridge University Press.
19. Bailey, J. (1863). An account of the wild tribes of the Veddahs of Ceylon: their habits, customs, and superstitions. *Transactions Ethnological Society of London*, *2*, 278–320.
20. Seligman, C. G., & Seligman, B. Z. (1911). *The Veddas*. Cambridge: Cambridge University Press.
21. Nimmo, H. A. (1972). *Badjau of the Philipines*. New Haven: HRAF Press.
22. Nimmo, H. A. (1965). Social organization of the Tawi-Tawi Badjaw. *Ethnology*, *4*(4), 421–439.
23. Headland, T. N. (1987). Kinship and social behavior among Agta Negrito hunter- gatherers. *Ethnology*, *6*(4), 261–280.
24. Rahmann, R., & Maceda, M. N. (1955). Notes on the Negritos of Northern Negros. *Anthropos, 50*(4/6), 810–836.
25. Goodale, J. C. (1971). *Tiwi wives: a study of the women of Melville Island, North Australia.* Seattle: University of Washington Press.

26. Goodale, J. C. (1999). The Tiwi of Melville and Bathurst Islands, north Australia. In R. B. Lee, & R. H. Daly (Eds.), *Cambridge encyclopedia of hunters and gatherers* (pp. 353–362). Cambridge, UK: Cambridge University Press.
27. Hart, C. W. M. (1930). *The Tiwi of Melville and Bathurst Islands*. Sydney: University of Sydney.
28. Swanson, G .E. (1960). *The birth of the gods: the origin of primitive belief.* Ann Arbor: University of Michigan Press.
29. Spencer, S. B., & Gillen, F. J. (1938 [1899]). *The native tribes of Central Australia.* London: Macmillan.
30. Penniman, T. K. (1929). The Arunta religion. *The Sociological Review*, 21, 10–37.
31. Meggitt, M. J. (1965). *Desert people: a study of the Walbiri Aborigines of Central Australia.* Chicago: University of Chicago Press.
32. Watanabe, H. (1964). *The Ainu: a study of ecology and the system of social solidarity between man and nature in relation to group structure*. Tokyo: University of Tokyo.
33. Batchelor, J. (1927. *Ainu life and lore: Echoes of a departing race*. Tokyo: Kyobunkwan.
34. Ohnuki-Tierney, E. (1973). The shamanism of the Ainu of the northwest coast of Southern Sakhalin. *Ethnology*, *12*(1), 15–29.
35. Munro, N. G., Seligman, B. Z., & Watanabe, H. (1963). *Ainu creed and cult.* New York: Columbia University Press.
36. Black, L. (1973). The Nivkh (Gilyak) of Sakhalin and the Lower Amur. *Arctic Anthropology, 10*(1)*, 1–112.*
37. Seeland, N., & Schütze, F. (1882). The Gilyaks: an ethnographic sketch. *Russische Revue*, *21*, 97-130, 222–254.
38. Jochelson, W. (1926). The Yukaghir and the Yukaghirized Tungus. In F. Boas (Ed.), *The Jesup North Pacific expedition* (Vol. 9, pp. 135–342). Leiden, NL: Brill.
39. Sarychev, G. A. (1806*). Account of a voyage of discovery to the north-east of Siberia, the Frozen Ocean, and the North-East Sea* (Vol. 2). London: Printed for R. Phillips by J. G. Barnard.
40. Lantis, M. (1984). Aleut. In D. Damas (Ed.), *Arctic* (pp. 161–184). Washington, D.C: Smithsonian Institution.
41. Jones, D. M. (1970. *A study of social and economic problems in Unalaska, an Aleut village*. Ann Arbor: University Microfilms. .
42. Jenness, D. (1922). *The life of the Copper Eskimos*. Ottawa: F.A. Acland.
43. Jenness, D. (1959). *The people of the twilight*. Chicago: University of Chicago Press.
44. Damas, D. (1972). The Copper Eskimo. In M. G. Bicchieri (Ed.), *Hunters and gatherers today: a socioeconomic study of eleven such cultures in the Twentieth Century* (pp. 3– 50). New York: Holt, Rinehart and Winston..
45. De Coccola, R., King, P., & Houston, J. (1986). *The incredible Eskimo: life among the barren land Eskimo*. Surrey, B.C.: Hancock House.
46. Jenness, D. (1917). The Copper Eskimos. *Geographical Review*, *4*, 81–91.
47. Desbarats, P. (1969). *What they used to tell about: Indian legends from Labrador*. Toronto: McClelland and Stewart Limited.
48. Lane, K. S. (1952). The Montagnais Indians, 1600–1640. *Kroeber Anthropological Society papers, 7*, 1-62.
49. Speck, F. G. (1935). *Naskapi: the savage hunters of the Labrador Peninsula*. Norman: University of Oklahoma Press.
50. Hallowell, A. I. (2002). Ojibwa ontology, behavior, and world view. In. G. Harvey (Ed.), *Readings in indigenous religion* (pp. 17–49). London: Continuum.
51. Dunning, R. W. (1959). *Social and economic change among the northern Ojibwa*. Toronto: University of Toronto Press.
52. MacNeish, J. H. (1954). Folk beliefs of a Slave Indian band. *Journal of American Folklore,*

67(264), 185–198.

53. MacNeish, J. H., & Teit, J. A. (1956). Field notes on the Tahltan and Kaska Indians: 1912–1915. *Anthropologica*, *3*(1), 40–171.
54. Honigmann, J. J. (1949). *The Kaska Indians: culture and ethos of Kaska society*. New Haven: Yale University Press.
55. Birket-Smith, K., & De Laguna, F. (1938). *The Eyak Indians of the Copper River Delta, Alaska.* Copenhagen: Levin & Munksgaard.
56. Tylor, E. B. (1899). On two British Columbian house-posts with totemic carvings, in the Pitt-Rivers Museum, Oxford. *The Journal of the Anthropological Institute of Great Britain and Ireland, 28*(1/2), 136–137.
57. Stevenson, I. (1975). The belief and cases related to reincarnation among the Haida. *Journal of Anthropological Research*, *31*(4), 364–375.
58. Murdock, G. P. (1934). Kinship and social behavior among the Haida. *American Anthropologist*, *36*(3), 355–385.
59. Kan, S. (1986). The 19th-Century Tlingit potlatch: a new perspective. *American Ethnologist, 13*(2), 191–212.
60. Blackman, M. B. (1973). Totems to tombstones: culture change as viewed through the Haida mortuary complex, 1877-1971. *Ethnology*, *12*(1), 47–56.
61. Blackman, M. B. (1977). Ethnohistoric changes in the Haida potlatch complex. *Arctic Anthropology*, *14*(1), 39–53.
62. McIlwraith, T. F. (1948). The Bella Coola Indians (Vol. 1).Toronto: University of Toronto Press.
63. McIlwraith, T. F. (1948). The Bella Coola Indians (Vol. 2). Toronto: University of Toronto Press.
64. Elmendorf, W. W., & Kroeber, A. L. (1960). The structure of Twana culture with comparative notes on the structure of Yurok culture. *Washington State University Research Studies*, *28* (3, Monographic Supp. 2), 1–576.
65. Elmendorf, W. W. (1948). The cultural setting of the Twana Secret Society. *American Anthropologist*, *50*(4), 625–633.
66. Kroeber, A. L. (1925). The Yurok. In *Handbook of the Indians of California* (pp. 1–97). Bureau of American Ethnology (Bulletin No.78). Washington, DC: Smithsonian Institute.
67. Turrado Moreno, A., & Muirden, S. (1945). *Ethnography of the Guarauno Indians*. Caracas: Vargas.
68. Olsen, D. A. (1973*). Music and shamanism of the Winikina-Warao Indians: Songs for curing and other theurgy* (Vol. 1). Ann Arbor: University Microfilms International.
69. Holmberg, A. R. (1950). *Nomads of the Long Bow: The Siriono of Eastern Bolivia.* Washington, DC: Smithsonian Institution.
70. Keane, A. H. (1884). On the Botocudos. *Journal of the Anthropological Institute of Great Britain and Ireland*, *13*, 199–213.
71. Nimuendaju, C. (1946). Social organization and beliefs of the Botocudo of Eastern Brazil. *Southwestern Journal of Anthropology, 2*(1), 93–115.
72. Henry, J. (1964). *Jungle people: a Kaingang tribe of the highlands of Brazil.* New York: Vintage.
73. Hicks, D. (1966). The Kaingang and the Aweikóma: a cultural contrast. *Anthropos*, *61*(2), 839–846.
74. Métraux, A. (1947). Social organization of the Kaingang and Aweikóma according to C. Nimuendajú's unpublished data. *American Anthropologist*, *49*(1), 148–151.
75. Cooper, J. M. (1946). The Yahgan. In J. H. Stewart (Ed.), *Handbook of South American Indians, the marginal tribes* (Vol.1, pp. 81–106). Washington: U.S. Gov't.Print. Ofc.

Table A2. Hunter-gatherer populations in the study sample, their taxonomic nomenclature, representation across source trees, and proxies selected for hunter-gatherer populations absent in source trees.

Name[a]	ISO 639-3[b]	Alternative name [b]	Classification[c]	Lewis et al. (2013) classification[d]
!Kung	ktz	Ju|'hoan	Khoisan	Khoisan, Southern Africa
G/wi	gwj	|Gwi	Khoisan	Khoisan, Southern Africa
Hadza	hts	Hatsa	Khoisan	Khoisan, Hatsa
Sandawe	sad	Sandawe	Khoisan	Khoisan, Sandawe
Mbuti	efe	Efe	? (Niger-Kordofanian)	? (Nilo-Saharan)
Aka	axk	Yaka	? (Nilo-Saharan)	? (Niger-Congo)
Semang	jhi	Jehai	? (Austric, Austro-Asiatic)	? (Austro-Asiatic, Mon-Khmer)
Agta	agt	Agta	? (Austric, Austronesian)	?Austronesian, Malayo-Polynesian
Andamanese[1]	abj	Aka-Bea	Indo-Pacific, Andaman Is.	? (Andamanese)
Vedda[2,†]	ved	Veddah	? (Indo-Hittite)	? (Indo-European)
Badjau Tawi	bdl	Bajau	Austric, Austronesian	Austronesian, Malayo-Polynesian
Tiwi	tiw	Tiwi	Australian	Australian
Aranda[3,†]	axl	Aranda	Australian, Pama-Nyungan	Australian, Pama-Nyungan
Walbiri	wbp	Warlpiri	Australian, Pama-Nyungan	Australian, Pama-Nyungan
Ainu	ain	Ainu	? (Eurasiatic)	? (language isolate)
Gilyak	niv	Gilyak	? (Eurasiatic)	? (language isolate)
Yukaghir	niv	Yukaghir	? (Altaic)	? (language isolate)
Aleut	ale	Aleut	Eskimo-Aleut	Eskimo-Aleut
Copper Eskimo[4]	ikt	Inuinnaqtun	Eskimo-Aleut	Eskimo-Aleut
Montagnais	moe	Montagnais	Amerind, Almosan	Algic, Algonquian
Saulteaux[5]	ojw	Ojibwa, Western	Amerind, Almosan	Algic, Algonquian
Slave[6]	xls	Slavey, South	Na-Dene	Eyak-Athabaskan, Athabaskan
Kaska[7,†]	kkz	Kaska	Na-Dene	Eyak-Athabaskan, Athabaskan
Eyak[8,†]	eya	Eyak	Na-Dene	Eyak-Athabaskan
Haida[9]	hdn	Haida, Northern	Na-Dene	? (Haida)
Bellacoola	blc	Bella Coola	Amerind, Almosan	Salish
Twana[10,†]	twa	Twana	Amerind, Almosan	Salish
Yurok[11,†]	yur	Yurok	Amerind, Almosan	Algic
Warrau	wba	Warao	Amerind, ? (Paezan)	? (language isolate)
Siriono	srq	Siriono	Amerind, Equatorial	Tupian
Botocudo[12]	xok	Xokleng	Amerind, Macro-Ge	Jean
Aweikoma[13]	xok	Xokleng	Amerind, Macro-Ge	Jean
Yahgan	yag	Yámana	Amerind, ? (Andean)	? (language isolate)

† Population not present in source trees.

a. SCCS/eHRAF, b. *Ethnologue,* c. Greenberg and Ruhlen 2007; Ruhlen 1991, 2007, d. *Ethnologue*

1. More inclusive taxon Andamanese (oon, anq, gac) used as a proxy for Andamanese
2. Sinhalese (sin), a population speaking Sinhalese-Maldivian language related to Veddah (Lewis et al. 2013) and identified as genetically closest relative of Vedda (ref.) used as proxy for Vedda.
3. Arrernte (aer), a closely related population speaking Arandic language (Lewis et al. 2013) used as a proxy for Aranda.
4. More inclusive taxon Canadian Inuit (ikt, ike) used as a proxy for Copper Eskimo.
5. More inclusive taxon Ojibwa (ojc, ojg, ojb, ojs) used as a proxy for Saulteaux.
6. More inclusive taxon Slave (den (scs, xsl)) used as a proxy for Slave.
7. Kaska position based solely on linguistic classification
8. Eyak position based solely on linguistic classification
9. More inclusive taxon Haida (hdn, hax) used as a proxy for Haida.
10. Twana position based solely on linguistic classification.
11. Yurok position based solely on linguistic classification.
12. More inclusive taxon Kaingang (xok, zkp) used as a proxy for Botocudo.
13. More inclusive taxon Kaingang (xok, zkp) used as a proxy for Aweikoma.

Figure A1. Linguistic scaffold for supertree typology

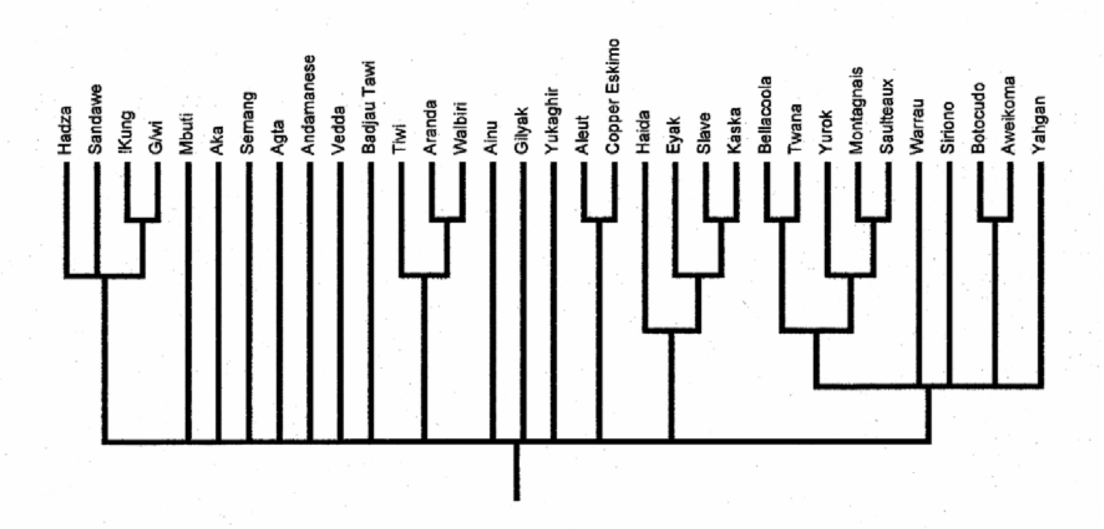


Figure A2. Supertree topology used in the study

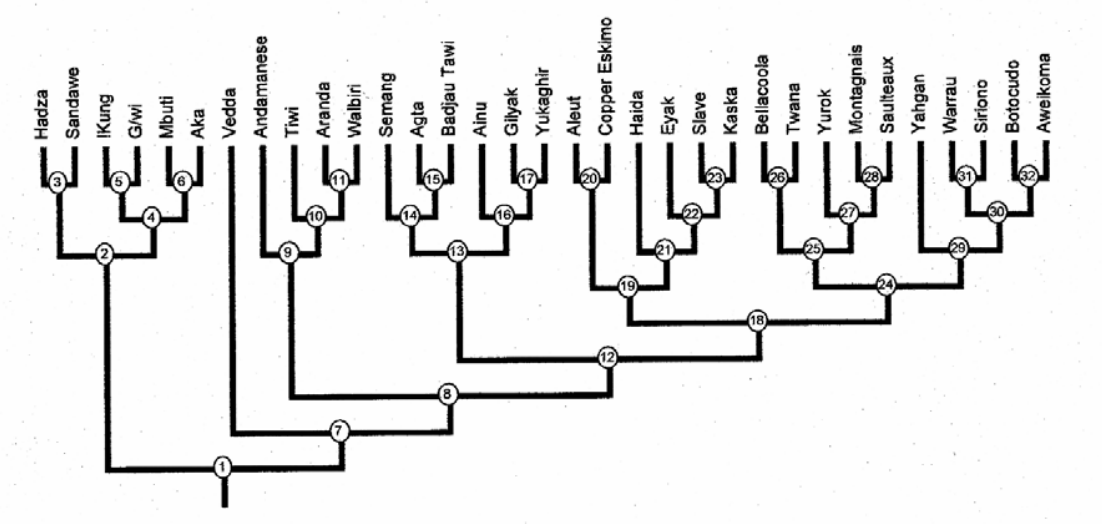


Table A3a. Divergence table of hunter-gatherer populations in the study sample

Node	Branching event	Divergence Dates
1	South African Khoisan+Pygmies+East African Khoisan divergence	An oldest divergence event in Africa
2	South African Khoisan+Pygmies/East African Khoisan divergence	
3	Hadza/Sandawe	
4	South African Khoisan/Pygmies	
5	!Kung/G/wi	North/Central South African Khoisan divergence
6	Mbuti/Aka	Eastern/Western Pygmies divergence
7	Vedda divergence	Out of Africa expansion into East Asia via ‘Southern route’
8	Andamanese + Australian Aboriginals divergence	Initial colonization of East Asia
9	Andamanese/Australian aboriginals	Initial colonization of Sahul
10	Tiwi/Aranda + Walbiri	Pama-Nyungan/non-Pama-Nyungan languages divergence
11	Aranda/Walbiri	Pama-Nyungan languages dispersal
12	East Asians/Native Americans	Colonization of East Asia via ‘Northern route’
13	South/North East Asians	
14	Semang/Philippinese	
15	Agta/Badjau Tawi	
16	Ainu/Gilyak + Yukaghir	
17	Gilyak/Yukaghir	
18	Eskimo-Aleut + Na-Dene + Amerindian divergence	Asian and Native American lineages divergence
19	Eskimo-Aleut/Na-Dene	
20	Aleut/Copper Eskimo	
21	Haida/Eyak-Athabaskan	Na-Dene languages dispersal
22	Eyak/Athabaskan	Eyak-Athabaskan languages dispersal
23	Kaska/Slave	Athabaskan languages dispersal
24	Amerindian divergence	Initial colonization of North America
25	Salish/Algic	Almosan languages dispersal
26	Bellacoola/Twana	Salishan languages dispersal
27	Yurok/Algonquian	Algic languages dispersal
28	Montagnais/Saulteaux	Algonquian languages dispersal
29	Yahgan divergence	Initial colonization of the Southern tip of South America
30	Tupi/Macro-Ge	Occupation of Lowland South America
31	Warrau/Siriono	
32	Botocudo/Aweikoma	Jean languages dispersal

Node	Time estimate (kya) - shallow divergence dates	Type of data
1	90	genetic (mtDNA, Y-chromosome, autosomal, genome-wide SNP)
2	60	genetic (mtDNA, genome-wide SNP data)
3	23	genetic (mtDNA; genome-wide SNP data)
4	51	genetic (mtDNA, genome-wide SNP data)
5	2	genetic (mtDNA)
6	**18**	genetic (mtDNA, autosomal data, genome-wide data)
7	**65**	genetic (mtDNA, autosomal data, genome-wide SNP data)
8	55.2	genetic (mtDNA)
9	50	archeological
10		genetic(mtDNA, genome-wide SNP data)
11	3 (expansion of Pama-Nyungan languages)	linguistic and archeological
12	?	
13	?	
14	30 (divergence of haplogroups in Austronesian speaking groups, Malaysia)	genetic (mtDNA + genome-wide SNP data)
15	4 (Austronesian entry into the Philippines)	archeological and linguistic
16	12 (disappearance of last land bridges between Japan/continental Asia)	archeological
17	?	
18	20	genetic (mtDNA) and archeological
19	8 (Na-Dene migration to America)	linguistic and archeological
20	5	linguistic and archeology
21	5 (appearance of coastal art, aesthetic styles associated with NW Coast)	archeological
22	3 (proto-Athabascan origin in the subarctic region of North America)	linguistic and archeological
23	2	linguistic
24	11.5 (expansion of a Beringian source population)	genetic, linguistic, archeological and paleoclimatology
25	?	
26	2 (unbroken ancestor/descendant relationships in American SW)	genetic (mtDNA)
27	4	linguistic
28	3 (dispersal of the Algonquian family)	linguistic
29	8 (disappearance of last land bridges of the Straits of Magellan)	archeological
30	6 (Macro-Ge dispersal)	linguistic
31	3 (beginning of the Tupi-Guarani expansion)	archeological
32	1.8 (expansion of Southern Jê languages through Southern Brazil)	archeological

Node	Reference
1	Knight et al 2003; Zhivotovsky et al. 2004; Gonder et al. 2007; Behar et al. 2008; Veeramah et al. 2012; Shriner et al. 2014
2	Tishkoff et al. 2007; Schlebush et al. 2012
3	Tishkoff et al. 2007
4	Tishkoff et al. 2007; Schlebush et al. 2012; Shriner et al. 2014
5	Barbieri et al. 2014
6	Chen et al. 2000; Destro-Bisol et al. 2004; Batini et al. 2007; Quintana-Murci et al. 2008; Patin et al. 2009; Verdu et al. 2009; Tishkoff et al. 2009
7	Macaulay et al. 2005; Liu et al. 2006; Gronau et al. 2011; Rasmussen et al. 2011; Fu et al. 2013b
8	Kumar et al. 2009
9	Bowler et al. 2003; O'Connell and Allen 2004; Summerhayes et al. 2010
10	Redd and Stoneking 1999; Pugach et al. 2013
11	McConvell 1996; Evans and McConvell 1998
12	
13	
14	Jinam et al. 2012
15	Pawley 2002; Reid 2013
16	Hammer et al. 2006
17	
18	Torroni et al. 1992; Saillard et al. 2000; Zlojutro et al. 2006; Goebel et al. 2008
19	Greenberg 1986
20	Greenberg 1986; Holman et al. 2011; Davis and Knecht 2010
21	Schurr et al. 2012
22	Schurr et al. 2012
23	Holman et al. 2011
24	Greenberg 1987; Kemp et al. 2007; Tamm et al. 2007; Achilli et al. 2013
25	
26	Eshleman et al. 2004
27	Golla 2007
28	Golla 2007
29	Borrero and McEwan 1997; McCulloch et al. 1997
30	Urban 1992; Callegari-Jacques et al. 2011
31	Walker et al. 2012
32	de Souza 2011

Node	Time estimate (kya) Deeper divergence dates
1	140
2	97.6
3	
4	51 (the Click Speaker ancestral component divergence)
5	35.3
6	27
7	93.5
8	87 (the Melanesian ancestral component divergence)
9	60
10	50
11	6 (expansion of Pama-Nyungan languages)
12	53 (the Native American ancestral component divergence)
13	40 (occupation of East Asia by populations ancestral to present day East Asians)
14	?
15	36 (divergence times for Australian + Papuan aboriginals + Phillippine negritos)
16	19.4 (start of spread of Y-chromosomal lineage D associated with Jomon culture)
17	?
18	30 (beginning of the radiation of Amerindian-specific mtDNA lineages)
19	12.1 (Beringian sublineage of mtDNA haplogroup A coalescence)
20	6 (Aleut-specific A and D mtDNA sublineages coalescence)
21	8 (Na-Dene portion of the HaeIII np 663 mtDNA lineage coalescence)
22	
23	2
24	16.5 (first migration from Beringia to the Americas)
25	?
26	
27	
28	
29	14.6 (earliest occupation of Mante Verde site, Chile)
30	11
31	4 (early occupation of the Orinoco river delta)
32	3 (separation of Southern Jê languages from the Northern and Central branches)

Node	Type of data
1	genetic (mtDNA; Y-chromosome, autosomal, genome-wide SNP data)
2	genetic (genome-wide SNP data)
3	genetic (mtDNA)
4	genetic (mtDNA, genome-wide SNP data)
5	genetic (mtDNA, genome-wide SNP data)
6	genetic (mtDNA)
7	genetic (genome-wide SNP data)
8	genetic (genome-wide SNP data)
9	genetic (mtDNA)
10	archeological
11	linguistic and archeological
12	genetic (genome-wide SNP data)
13	archeological and genetic (mtDNA)
14	
15	genetic (genome-wide SNP data)
16	genetic (Y-chromosome)
17	
18	genetic (mtDNA, Y-chromosome)and archeological
19	genetic (mtDNA)
20	genetic (mtDNA) and archeological
21	genetic (mtDNA)
22	linguistic
23	linguistic
24	genetic (mtDNA, Y-chromosome) and archeological
25	
26	linguistic
27	linguistic
28	linguistic
29	archeological
30	genetic and archeological
31	archeological
32	linguistic

Node	Reference
1	Chen et al. 2000; Knight et al. 2003; Zhivotovsky et al. 2004; Behar et al. 2008; Gronau et al. 2011; Veeramah et al. 2012
2	Schlebush et al. 2012
3	Gonder et al. 2007
4	Tishkoff et al. 2007; Schlebush et al. 2012; Shriner et al. 2014
5	Gonder et al. 2007, Tishkoff et al. 2007; Behar et al. 2008; Schlebush et al. 2012
6	Batini et al. 2011
7	Shriner et al. 2014
8	Shriner et al. 2014
9	Macaulay et al. 2005; Hill et al. 2006
10	O'Connell and Allen 2004
11	McConvell 1996; Evans and McConvell 1998
12	Shriner et al. 2014
13	Fu et al. 2013a
14	
15	Pugach et al. 2013
16	Hammer et al. 2006
17	
18	Torroni et al. 1992; Saillard et al. 2000; Zlojutro et al. 2006; Goebel et al. 2008
19	Achilli et al. 2013
20	Rubicz et al. 2003; Zlojutro et al. 2006; Davis and Knecht 2010
21	Torroni et al. 1992
22	Holman et al. 2011
23	Holman et al. 2011
24	Goebel et al. 2008
25	
26	Holman et al. 2011
27	Holman et al. 2011
28	Holman et al. 2011
29	Dillehay et al. 2008
30	Rothhammer and Dillehay 2009
31	Gasson 2002
32	Urban 1992; Callegari-Jacques et al. 2011

Table A3b. References for divergence dates for hunter-gatherer populations in the study sample

Achilli, A., Perego, U. A., Lancioni, H., Olivieri, A., Gandini, F., Kashani, B. H., et al. (2013). Reconciling migration models to the Americas with the variation of North American native mitogenomes. *Proceedings of the National Academy of Sciences of the United States of America, 110*(35), 14308-14313, doi:10.1073/pnas.1306290110.

Barbieri, C., Guldemann, T., Naumann, C., Gerlach, L., Berthold, F., Nakagawa, H., et al. (2014). Unraveling the Complex Maternal History of Southern African Khoisan Populations. *American Journal of Physical Anthropology, 153*(3), 435-448, doi:10.1002/ajpa.22441.

Batini, C., Coia, V., Battaggia, C., Rocha, J., Pilkington, M. M., Spedini, G., et al. (2007). Phylogeography of the human mitochondrial L1c haplogroup: Genetic signatures of the prehistory of Central Africa. [Article]. *Molecular Phylogenetics and Evolution, 43*(2), 635-644, doi:10.1016/j.ympev.2006.09.014.

Behar, D. M., Villems, R., Soodyall, H., Blue-Smith, J., Pereira, L., Metspalu, E., et al. (2008). The dawn of human matrilineal diversity. *American Journal of Human Genetics, 82*(5), 1130-1140, doi:10.1016/j.ajhg.2008.04.002.

Borrero, L. A., & McEwan, C. (1997). The peopling of Patagonia: The first human occupation. In *Patagonia. Natural History, Prehistory and Ethnography at the uttermost end of the Earth* (pp. 32-45). London: British Museum Press.

Bowler, J. M., Johnston, H., Olley, J. M., Prescott, J. R., Roberts, R. G., Shawcross, W., et al. (2003). New ages for human occupation and climatic change at Lake Mungo, Australia. *Nature, 421*(6925), 837-840, doi:10.1038/nature01383.

Callegari-Jacques, S. M., Tarazona-Santos, E. M., Gilman, R. H., Herrera, P., Cabrera, L., dos Santos, S. E. B., et al. (2011). Autosome STRs in Native South America-Testing Models of Association With Geography and Language. *American Journal of Physical Anthropology, 145*(3), 371-381, doi:10.1002/ajpa.21505.

Davis, R. S., & Knecht, R. A. (2010). Continuity and Change in the Eastern Aleutian Archaeological Sequence. *Human Biology, 82*(5-6), 507-524.

de Souza, J. G. (2011). Linguistics, archaeology, and the histories of language spread: the case of the Southern Jê languages, Brazil. *Cadernos de Etnolingüística (ISSN 1946-7095), 3*(2).

Destro-Bisol, G., Coia, V., Boschi, I., Verginelli, F., Caglià, A., Pascali, V., et al. (2004). The analysis of variation of mtDNA hypervariable region 1 suggests that Eastern and Western Pygmies diverged before the Bantu expansion. *The American Naturalist, 163*(2), 212-226.

Eshleman, J. A., Malhi, R. S., Johnson, J. R., Kaestle, F. A., Lorenz, J., & Smith, D. G. (2004). Mitochondrial DNA and prehistoric settlements: Native migrations on the western edge of North America. *Human Biology, 76*(1), 55-75, doi:10.1353/hub.2004.0019.

Evans, N., & McConvell, P. (1998). The enigma of Pama-Nyungan expansion in Australia. In R. Blench, & M. Springs (Eds.), *Archaeology and language II: Archeological Data and Linguistic Hypotheses* (pp. 174-192). London: Routledge.

Fu, Q. M., Meyer, M., Gao, X., Stenzel, U., Burbano, H. A., Kelso, J., et al. (2013). DNA analysis of an early modern human from Tianyuan Cave, China. [Article]. *Proceedings of the National Academy of Sciences of the United States of America, 110*(6), 2223-2227, doi:10.1073/pnas.1221359110.

Fu, Q. M., Mittnik, A., Johnson, P. L. F., Bos, K., Lari, M., Bollongino, R., et al. (2013). A Revised Timescale for Human Evolution Based on Ancient Mitochondrial Genomes. *Current Biology, 23*(7), 553-559, doi:10.1016/j.cub.2013.02.044.

Goebel, T., Waters, M. R., & O'Rourke, D. H. (2008). The Late Pleistocene dispersal of modern humans in the Americas. *Science, 319*(5869), 1497-1502, doi:10.1126/science.1153569.

Golla, V. (2007). Linguistic prehistory. In T. L. Jones, & K. A. Clar (Eds.), *California prehistory:*

Colonization, culture, and complexity (pp. 71-82). Plymouth: Altamira Press.

Gonder, M. K., Mortensen, H. M., Reed, F. A., de Sousa, A., & Tishkoff, S. A. (2007). Whole-mtDNA genome sequence analysis of ancient African lineages. [Article]. *Molecular Biology and Evolution, 24*(3), 757-768, doi:10.1093/molbev/msl209.

Greenberg, J. H. (1987). *Language in the Americas*: Stanford University Press.

Gronau, I., Hubisz, M. J., Gulko, B., Danko, C. G., & Siepel, A. (2011). Bayesian inference of ancient human demography from individual genome sequences. *Nature Genetics, 43*(10), 1031-U1151, doi:10.1038/ng.937.

Hammer, M. F., Karafet, T. M., Park, H., Omoto, K., Harihara, S., Stoneking, M., et al. (2006). Dual origins of the Japanese: common ground for hunter-gatherer and farmer Y chromosomes. [Article]. *Journal of Human Genetics, 51*(1), 47-58, doi:10.1007/s10038-005-0322-0.

Holman, E. W., Brown, C. H., Wichmann, S., Muller, A., Velupillai, V., Hammarstrom, H., et al. (2011). Automated Dating of the World's Language Families Based on Lexical Similarity. *Current Anthropology, 52*(6), 841-875, doi:10.1086/662127.

Chen, Y. S., Olckers, A., Schurr, T. G., Kogelnik, A. M., Huoponen, K., & Wallace, D. C. (2000). mtDNA variation in the South African Kung and Khwe - and their genetic relationships to other African populations. *American Journal of Human Genetics, 66*(4), 1362-1383, doi:10.1086/302848.

Jinam, T. A., Hong, L. C., Phipps, M. E., Stoneking, M., Ameen, M., Edo, J., et al. (2012). Evolutionary History of Continental Southeast Asians: "Early Train" Hypothesis Based on Genetic Analysis of Mitochondrial and Autosomal DNA Data. *Molecular Biology and Evolution, 29*(11), 3513-3527, doi:10.1093/molbev/mss169.

Kemp, B. M., Malhi, R. S., McDonough, J., Bolnick, D. A., Eshleman, J. A., Rickards, O., et al. (2007). Genetic analysis of early holocene skeletal remains from Alaska and its implications for the settlement of the Americas. *American Journal of Physical Anthropology, 132*(4), 605-621, doi:10.1002/ajpa.20543.

Knight, A., Underhill, P. A., Mortensen, H. M., Zhivotovsky, L. A., Lin, A. A., Henn, B. M., et al. (2003). African Y chromosome and mtDNA divergence provides insight into the history of click languages. *Current Biology, 13*(6), 464-473, doi:10.1016/s0960-9822(03)00130-1.

Kumar, S., Ravuri, R. R., Koneru, P., Urade, B. P., Sarkar, B. N., Chandrasekar, A., et al. (2009). Reconstructing Indian-Australian phylogenetic link. *Bmc Evolutionary Biology, 9*, doi:10.1186/1471-2148-9-173.

Liu, H., Prugnolle, F., Manica, A., & Balloux, F. (2006). A geographically explicit genetic model of worldwide human-settlement history. *American Journal of Human Genetics, 79*(2), 230-237, doi:10.1086/505436.

Macaulay, V., Hill, C., Achilli, A., Rengo, C., Clarke, D., Meehan, W., et al. (2005). Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Science, 308*(5724), 1034-1036, doi:10.1126/science.1109792.

McConvell, P. (1996). Backtracking to Babel: the chronology of Pama-Nyungan expansion in Australia. *Archaeology in Oceania*, 125-144.

McCulloch RD, C. C., Rabassa J, Currant, AP (1997). The natural setting: The glacial and post-glacial environmental history of Fuego-Patagonia. In B. L. McEwan C, Prieto A (Ed.), *Patagonia. Natural history, prehistory and ethnography at the uttermost end of the earth* (pp. 12-31). London: British Museum 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*(6), 835-853, doi:10.1016/j.jas.2003.11.005.

Patin, E., Laval, G., Barreiro, L. B., Salas, A., Semino, O., Santachiara-Benerecetti, S., et al. (2009).

Inferring the Demographic History of African Farmers and Pygmy Hunter-Gatherers Using a Multilocus Resequencing Data Set. *Plos Genetics, 5*(4), doi:10.1371/journal.pgen.1000448.

Pawley, A. (2002). The Austronesian dispersal: Languages, technologies and people. In P. Bellwood, & C. Renfrew (Eds.), *Examining the Farming/Language Dispersal Hypothesis* (pp. 251-274). Cambridge: McDonald Institute for Archaeological Research.

Pugach, I., Delfin, F., Gunnarsdottir, E., Kayser, M., & Stoneking, M. (2013). Genome-wide data substantiate Holocene gene flow from India to Australia. *Proceedings of the National Academy of Sciences of the United States of America, 110*(5), 1803-1808, doi:10.1073/pnas.1211927110.

Quintana-Murci, L., Quach, H., Harmant, C., Luca, F., Massonnet, B., Patin, E., et al. (2008). Maternal traces of deep common ancestry and asymmetric gene flow between Pygmy hunter–gatherers and Bantu-speaking farmers. *Proceedings of the National Academy of Sciences, 105*(5), 1596-1601.

Rasmussen, M., Guo, X. S., Wang, Y., Lohmueller, K. E., Rasmussen, S., Albrechtsen, A., et al. (2011). An Aboriginal Australian Genome Reveals Separate Human Dispersals into Asia. *Science, 334*(6052), 94-98, doi:10.1126/science.1211177.

Redd, A. J., & Stoneking, M. (1999). Peopling of Sahul: mtDNA variation in Aboriginal Australian and Papua New Guinean populations. *American Journal of Human Genetics, 65*(3), 808-828, doi:10.1086/302533.

Reid, L. A. (2013). Who Are the Philippine Negritos? Evidence from Language. *Human Biology, 85*(1-3), 329-358.

Saillard, J., Forster, P., Lynnerup, N., Bandelt, H. J., & Norby, S. (2000). mtDNA variation among Greenland Eskimos: The edge of the Beringian expansion. *American Journal of Human Genetics, 67*(3), 718-726, doi:10.1086/303038.

Shriner, D., Tekola-Ayele, F., Adeyemo, A., & Rotimi, C. N. (2014). Genome-wide genotype and sequence-based reconstruction of the 140,000 year history of modern human ancestry. *Scientific Reports, 4*, doi:10.1038/srep06055.

Schlebusch, C. M., Skoglund, P., Sjodin, P., Gattepaille, L. M., Hernandez, D., Jay, F., et al. (2012). Genomic Variation in Seven Khoe-San Groups Reveals Adaptation and Complex African History. [Article]. *Science, 338*(6105), 374-379, doi:10.1126/science.1227721.

Schurr, T. G., Dulik, M. C., Owings, A. C., Zhadanov, S. I., Gaieski, J. B., Vilar, M. G., et al. (2012). Clan, language, and migration history has shaped genetic diversity in Haida and Tlingit populations from Southeast Alaska. *American Journal of Physical Anthropology, 148*(3), 422-435, doi:10.1002/ajpa.22068.

Summerhayes, G. R., Leavesley, M., Fairbairn, A., Mandui, H., Field, J., Ford, A., et al. (2010). Human Adaptation and Plant Use in Highland New Guinea 49,000 to 44,000 Years Ago. *Science, 330*(6000), 78-81, doi:10.1126/science.1193130.

Tamm, E., Kivisild, T., Reidla, M., Metspalu, M., Smith, D. G., Mulligan, C. J., et al. (2007). Beringian Standstill and Spread of Native American Founders. *Plos One, 2*(9), doi:10.1371/journal.pone.0000829.

Tishkoff, S. A., Gonder, M. K., Henn, B. M., Mortensen, H., Knight, A., Gignoux, C., et al. (2007). History of click-speaking Populations of Africa inferred from mtDNA and Y chromosome genetic variation. [Article]. *Molecular Biology and Evolution, 24*(10), 2180-2195, doi:10.1093/molbev/msm155.

Tishkoff, S. A., Reed, F. A., Friedlaender, F. R., Ehret, C., Ranciaro, A., Froment, A., et al. (2009). The Genetic Structure and History of Africans and African Americans. *Science, 324*(5930), 1035-1044, doi:10.1126/science.1172257.

Torroni, A., Schurr, T. G., Yang, C. C., Szathmary, E. J. E., Williams, R. C., Schanfield, M. S., et al. (1992). Native American mitochondrial DNA analysis indicates that the Amerind and the

Nadene populations were founded by two independent migrations. *Genetics, 130*(1), 153-162.

Urban, G. (1992). A história da cultura brasileira segundo as línguas nativas. In M. Carneiro da Cunha (Ed.), *História dos Índios no Brasil* (pp. 87-102). São Paulo: Companhia das Letras.

Veeramah, K. R., Wegmann, D., Woerner, A., Mendez, F. L., Watkins, J. C., Destro-Bisol, G., et al. (2012). An Early Divergence of KhoeSan Ancestors from Those of Other Modern Humans Is Supported by an ABC-Based Analysis of Autosomal Resequencing Data. *Molecular Biology and Evolution, 29*(2), 617-630, doi:10.1093/molbev/msr212.

Verdu, P., Austerlitz, F., Estoup, A., Vitalis, R., Georges, M., Thery, S., et al. (2009). Origins and Genetic Diversity of Pygmy Hunter-Gatherers from Western Central Africa. *Current Biology, 19*(4), 312-318, doi:10.1016/j.cub.2008.12.049.

Walker, R. S., Wichmann, S., Mailund, T., & Atkisson, C. J. (2012). Cultural Phylogenetics of the Tupi Language Family in Lowland South America. *Plos One, 7*(4), doi:10.1371/journal.pone.0035025.

Zhivotovsky, L. A., Underhill, P. A., Cinnioglu, C., Kayser, M., Morar, B., Kivisild, T., et al. (2004). The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time. *American Journal of Human Genetics, 74*(1), 50-61, doi:10.1086/380911.

Zlojutro, M., Rubicz, R., Devor, E. J., Spitsyn, V. A., Makarov, S. V., Wilson, K., et al. (2006). Genetic structure of the Aleuts and circumpolar populations based on mitochondrial DNA sequences: A synthesis. *American Journal of Physical Anthropology, 129*(3), 446-464, doi:10.1002/ajpa.20287.

Table A4a. Ancestral reconstruction for three characters of hunter-gatherer religiosity in all nodes: Animism, Afterlife, Shamanism

Ancestral node	Reconstruction method	Time calibration	Animism	Belief in an Afterlife	Shamanism
1	Parsimony ancestral states	-	1	0.5	0
	Likelihood ancestral states	shallow divergences	**0.98935669***	0.50525806	0.55770544
		deep divergences	**0.98855061***	0.54626414	0.57603722
2	Parsimony ancestral states	-	1	0.5	0
	Likelihood ancestral states	shallow divergences	**0.99978439***	0.49614074	0.34514788
		deep divergences	**0.9992759***	0.51593342	0.42783392
3	Parsimony ancestral states	-	1	0.5	0
	Likelihood ancestral states	shallow divergences	**0.99995378***	0.49974141	0.06760492*
		deep divergences	**0.99995883***	0.50208167	0.05310023*
4	Parsimony ancestral states	-	1	0.5	0
	Likelihood ancestral states	shallow divergences	**0.99992715***	0.49164308	0.35377935
		deep divergences	**0.99998478***	0.51717751	0.49593014
5	Parsimony ancestral states	-	1	1	1
	Likelihood ancestral states	shallow divergences	**0.99999957***	**0.99469716***	**0.99922111***
		deep divergences	**0.99998875***	0.72794605	0.81660599
6	Parsimony ancestral states	-	1	0	0
	Likelihood ancestral states	shallow divergences	**0.99997597***	0.250521	0.04693761*
		deep divergences	**0.9999904***	0.18028348	0.09180091*
7	Parsimony ancestral states	-	1	1	1
	Likelihood ancestral states	shallow divergences	**0.99961897***	0.53856939	0.83548137
		deep divergences	**0.99985014***	0.70772543	0.87487299
8	Parsimony ancestral states	-	1	1	1
	Likelihood ancestral states	shallow divergences	**0.9999881***	0.57241813	**0.91262316***
		deep divergences	**0.99998102***	0.74213059	**0.90140422***

Asterisk (*) indicates significant result $p \leq 0.05$.

Ancestral node	**Reconstruction Method**	**Time Calibration**	**Animism**	**Belief in an Afterlife**	**Shamanism**
9	Parsimony ancestral states	-	1	1	1
	Likelihood ancestral states	shallow divergences	**0.99996997***	0.56556156	**0.90538627***
		deep divergences	**0.99998954***	0.78460594	0.87452778
10	Parsimony ancestral states	-	1	1	1
	Likelihood ancestral states	shallow divergences	**0.99999949***	**0.99165112***	0.84383663
		deep divergences	**0.9999868***	0.8038125	0.83073051
11	Parsimony ancestral states	-	1	1	1
	Likelihood ancestral states	shallow divergences	**0.99999997***	**0.99769778***	**0.98743635***
		deep divergences	**0.99999915***	**0.99340609***	**0.99847948***
12	Parsimony ancestral states	-	1	1	1
	Likelihood ancestral states	shallow divergences	**0.99999893***	0.73089051	**0.99469604***
		deep divergences	**0.99999445***	**0.89199822***	**0.98944864***
13	Parsimony ancestral states	-	1	1	1
	Likelihood ancestral states	shallow divergences	**0.99999965***	0.73103535	**0.99593307***
		deep divergences	**0.99999992***	**0.9155599***	**0.99909357***
14	Parsimony ancestral states	-	1	1	1
	Likelihood ancestral states	shallow divergences	**0.99999485***	0.7047243	**0.99353478***
		deep divergences	**0.99999994***	**0.91351938***	**0.99911158***
15	Parsimony ancestral states	-	1	1	**1**
	Likelihood ancestral states	shallow divergences	**0.9999991***	0.53063856	**0.99931817***
		deep divergences	**0.99999752***	0.87068323	**0.99765799***
16	Parsimony ancestral states	-	1	1	1
	Likelihood ancestral states	shallow divergences	**0.9999967***	**0.91450125***	**0.9973549***
		deep divergences	**0.99999794***	**0.97538817***	**0.99837651***

Ancestral node	**Reconstruction Method**	**Time Calibration**	**Animism**	**Belief in an Afterlife**	**Shamanism**
17	Parsimony ancestral states	-	1	1	1
	Likelihood ancestral states	shallow divergences	**0.99999951***	**0.97574001***	**0.99948763***
		deep divergences	**0.9999995***	**0.99381645***	**0.99954537***
18	Parsimony ancestral states	-	1	1	1
	Likelihood ancestral states	shallow divergences	**0.99999677***	0.87608822	**0.9969949***
		deep divergences	**0.99999701***	**0.9633506***	**0.99689702***
19	Parsimony ancestral states	-	1	1	1
	Likelihood ancestral states	shallow divergences	**0.99999976***	**0.98712943***	**0.99977702***
		deep divergences	**0.99999976***	**0.99699529***	**0.99977844***
20	Parsimony ancestral states	-	1	1	1
	Likelihood ancestral states	shallow divergences	**0.99999984***	**0.99353513***	**0.99986176***
		deep divergences	**0.99999988***	**0.99899329***	**0.99990679***
21	Parsimony ancestral states	-	1	1	1
	Likelihood ancestral states	shallow divergences	**0.99999993***	**0.99353513***	**0.99994319***
		deep divergences	**0.99999993***	**0.99852242***	**0.99993989***
22	Parsimony ancestral states	-	1	1	1
	Likelihood ancestral states	shallow divergences	**0.99999999***	**0.98827912***	**0.99998903***
		deep divergences	**0.99999998***	**0.99557049***	**0.9999853***
23	Parsimony ancestral states	-	1	1	1
	Likelihood ancestral states	shallow divergences	**0.99999999***	**0.95388015***	**0.99999317***
		deep divergences	**1***	**0.96353441***	**0.99999643***
24	Parsimony ancestral states	-	1	1	1
	Likelihood ancestral states	shallow divergences	**0.99999953***	**0.94280473***	**0.99861319***
		deep divergences	**0.99999985***	**0.98661895***	**0.99885499***

Ancestral node	Reconstruction Method	Time Calibration	Animism	Belief in an Afterlife	Shamanism
25	Parsimony ancestral states	-	1	1	1
	Likelihood ancestral states	shallow divergences	**0.99999998***	**0.99907201***	**0.9999851***
		deep divergences	**0.99999999***	**0.99989791***	**0.99999276***
26	Parsimony ancestral states	-	1	1	1
	Likelihood ancestral states	shallow divergences	**0.99999998***	**0.99948244***	**0.99998451***
		deep divergences	**0.99999998***	**0.99987743***	**0.99998721***
27	Parsimony ancestral states	-	1	1	1
	Likelihood ancestral states	shallow divergences	**1***	**0.99959627***	**0.9999955***
		deep divergences	**1***	**0.99994808***	**0.99999633***
28	Parsimony ancestral states	-	1	1	1
	Likelihood ancestral states	shallow divergences	**0.99999998***	**0.99946611***	**0.99998532***
		deep divergences	**0.99999999***	**0.99991261***	**0.99999068***
29	Parsimony ancestral states	-	1	1	1
	Likelihood ancestral states	shallow divergences	**0.99999987***	**0.92760674***	**0.99526214***
		deep divergences	**0.99999994***	**0.98201006***	**0.99753155***
30	Parsimony ancestral states	-	1	1	1
	Likelihood ancestral states	shallow divergences	**0.99999994***	0.86947973	**0.97269425***
		deep divergences	**0.99999989***	**0.93207902***	**0.97216753***
31	Parsimony ancestral states	-	1	1	1
	Likelihood ancestral states	shallow divergences	**0.99999994***	0.79676298	**0.94075587***
		deep divergences	**0.99999994***	0.84931385	**0.9264907***
32	Parsimony ancestral states	-	1	1	1
	Likelihood ancestral states	shallow divergences	**0.99999997***	0.77185503	**0.92859314***
		deep divergences	**0.99999996***	0.83886236	**0.92033659***

Table A4b. Ancestral reconstruction for four characters of hunter-gatherer religiosity in all nodes: Ancestor Worship, High Gods, Active High Gods and Ancestor Worship

Ancestral node	Reconstruction Method	Time calibration	Ancestor Worship	High Gods	Active High Gods	Active Ancestor Worship
1	Parsimony	-	0	0	0	0
	Likelihood Ancestral states	shallow divergences	0.5	0.5	0.04649112*	0.43429948
		deep divergences	0.5	0.5	0.07916184*	0.10739427*
2	Parsimony	-	0	0	0	0
	Likelihood ancestral states	shallow divergences	0.49999984	0.5	0.06306996*	0.3534881
		deep divergences	0.5	0.49999999	0.12071412	0.06484999*
3	Parsimony	-	0	0	0	0
	Likelihood ancestral states	shallow divergences	0.5	0.49967258	0.01172006*	0.18041384
		deep divergences	0.5	0.49676052	0.01412709*	0.01828653*
4	Parsimony	-	0	0	0	0
	Likelihood ancestral states	shallow divergences	0.49999815	0.5	0.09924286*	0.33153467
		deep divergences	0.5	0.5	0.29764904	0.01194196*
5	Parsimony	-	0	1	1	0
	Likelihood ancestral states	shallow divergences	0.06488181*	0.89889098*	**0.99946059***	0.00201257*
		deep divergences	0.5	0.50133933	0.80457262	0.00837643*
6	Parsimony	-	0	0	0	0
	Likelihood ancestral states	shallow divergence	0.49221734	0.49812636	0.00958869*	0.12433432
		deep divergences	0.49999989	0.4936424	0.03991104*	0.00594483*
7	Parsimony	-	0.5	0	0	0
	Likelihood Ancestral states	shallow divergences	0.50000002	0.5	0.0147429*	0.40874244
		deep divergences	0.5	0.49999995	0.01351448*	0.08523942*
8	Parsimony	-	0.5	0	0	0
	Likelihood ancestral states	shallow divergences	0.50000011	0.49999999	0.00509984*	0.34535215
		deep divergences	0.5	0.49999988	0.00723788*	0.05094172*

Ancestral node	Reconstruction Method	Time Calibration	Ancestor Worship	High Gods	Active High Gods	Active Ancestor Worship
9	Parsimony	-	0.5	0	0	0
	Likelihood ancestral states	shallow divergences	0.50000086	0.49999997	0.00751864*	0.33788734
		deep divergences	0.5	0.49998812	0.00506884*	0.01994129*
10	Parsimony	-	0.5	0	0	0
	Likelihood ancestral states	shallow divergences	0.86638714	0.4003842	0.00008568*	0.002961*
		deep divergences	0.5	0.49996439	0.00511664*	0.01621562*
11	Parsimony	-	0.5	0	0	0
	Likelihood ancestral states	shallow divergences	**0.91601139***	0.23857514	0.00000794*	0.00056653*
		deep divergences	0.52822414	0.20606209	0.00020569*	0.00045198*
12	Parsimony	-	0.5	0	0	0
	Likelihood ancestral states	shallow divergences	0.4999709	0.49999631	0.00269126*	0.2670984
		deep divergences	0.5	0.4999766	0.0022686*	0.01833662*
13	Parsimony	-	0.5	0.5	0	0
	Likelihood ancestral states	shallow divergences	0.49998523	0.50000421	0.00385904*	0.27570418
		deep divergences	0.5	0.50053666	0.00237663*	0.02320494*
14	Parsimony	-	0	0.5	0	0
	Likelihood ancestral states	shallow divergences	0.49965116	0.50001396	0.02166731*	0.24326767
		deep divergences	0.5	0.50053968	0.00285221*	0.02030816*
15	Parsimony	-	0	0.5	0	0
	Likelihood ancestral states	shallow divergences	0.19529316	0.5	0.00023072*	0.00718567*
		deep divergences	0.5	0.50027495	0.00235248*	0.01497359*
16	Parsimony	-	1	0.5	0	1
	Likelihood ancestral states	shallow divergences	0.55725585	0.50737492	0.00062315*	**0.93293674***
		deep divergences	0.50001471	0.51254573	0.00058618*	**0.94440009***

Ancestral node	Reconstruction Method	Time Calibration	Ancestor Worship	High Gods	Active High Gods	Active Ancestor Worship
17	Parsimony	-	1	1	0	1
	Likelihood ancestral states	shallow divergences	0.69400272	0.62055337	0.00009844*	**0.98948634***
		deep divergences	0.50313132	0.65638219	0.00013518*	**0.99662563***
18	Parsimony	-	0.5	0	0	0
	Likelihood ancestral states	shallow divergences	0.49826689	0.49850052	0.00103311*	0.15734521
		deep divergences	0.49999998	0.49472749	0.00124285*	0.01505229
19	Parsimony	-	0.5	0	0	0
	Likelihood ancestral states	shallow divergences	0.45586828	0.44045262	0.00004553*	0.28690215
		deep divergences	0.49932472	0.40121434	0.00006578*	0.0820227*
20	Parsimony	-	1	0	0	1
	Likelihood ancestral states	shallow divergences	0.68882479	0.47909209	0.00002802*	**0.88555905***
		deep divergences	0.52818629	0.46849228	0.00002801*	**0.96743862***
21	Parsimony	-	0	0	0	0
	Likelihood ancestral states	shallow divergences	0.19239451	0.33042693	0.00001137*	0.02137913*
		deep divergences	0.48374344	0.28716776	0.00001737*	0.00370175*
22	Parsimony	-	0	0	0	0
	Likelihood ancestral states	shallow divergences	0.05982445	0.2874466	0.00000222*	0.00085507*
		deep divergences	0.38447096	0.23094688	0.00000443*	0.00006444*
23	Parsimony	-	0.5	0	0	0
	Likelihood ancestral states	shallow divergences	0.02906697	0.35004854	0.00000142*	0.0001209*
		deep divergences	0.21364036	0.32182248	0.00000109*	0.00000285*
24	Parsimony	-	0.5	0	0	0
	Likelihood ancestral states	shallow divergences	0.49982502	0.48841483	0.0027186*	0.03498002*
		deep divergences	0.5	0.47689201	0.00319949*	0.00165965*

Ancestral node	Reconstruction Method	Time Calibration	Ancestor Worship	High Gods	Active High Gods	Active Ancestor Worship
25	Parsimony	-	0.5	0	0	0
	Likelihood ancestral states	shallow divergences	0.35041263	0.23064584	0.0000053*	0.00025072*
		deep divergences	0.48629034	0.15911451	0.0000039*	0.00000516*
26	Parsimony	-	0.5	0	0	0
	Likelihood ancestral states	shallow divergences	0.42170835	0.38340443	0.00000333*	0.00012641*
		deep divergences	0.49640135	0.27635687	0.00000399*	0.00000806*
27	Parsimony	-	0.5	0	0	0
	Likelihood ancestral states	shallow divergences	0.3326565	0.19004611	0.00000095*	0.00007796*
		deep divergences	0.48098824	0.12137504	0.0000012*	0.0000023*
28	Parsimony	-	0.5	0	0	0
	Likelihood ancestral states	shallow divergences	0.37222357	0.13970302	0.00000312*	0.00012309*
		deep divergences	0.49485905	0.05293082	0.00000288*	0.00000587*
29	Parsimony	-	0.5	0.5	0	0
	Likelihood ancestral states	shallow divergences	0.54906057	0.53114357	0.02124681*	0.03821882*
		deep divergences	0.50007616	0.50257617	0.01054609*	0.00253994*
30	Parsimony ancestral states	-	0.5	0.5	0	0
	Likelihood ancestral states	shallow divergences	0.52860332	0.5263287	0.01481972*	0.0788802*
		deep divergences	0.50000896	0.50746724	0.00797786*	0.02437226*
31	Parsimony ancestral states	-	0.5	1	0	0
	Likelihood ancestral states	shallow divergences	0.51273333	0.79383241	0.03472997*	0.13325871
		deep divergences	0.50000014	0.87539264	0.04073876*	0.06411901*
32	Parsimony ancestral states	-	0.5	0	0	0
	Likelihood ancestral states	shallow divergences	0.50921197	0.09371939*	0.00008399*	0.15299041
		deep divergences	0.50000008	0.07598965*	0.00005246*	0.06951924*

Table A5. Pagel's test for correlated discrete character evolution

Character X	Character Y		Independent model	Dependent model
Animism	Belief in an Afterlife	q12(alpha1)	0.017922678171600308	2.955006068003479E-7
		q13(alpha2)	3.5141401263485466	3.876039800319336E-6
		q21(beta1)	7.547566959924306E-4	4.876833291438565
		q31(beta2)	1.0812738648986855	0.001330532761444173
		q24	-	6.963364043052747
		q34	-	11.770106926545987
		q42	-	1.916250063082768E-5
		q43	-	2.753450428794025
		log Likelihood (L)	-20.611133199198772	-18.5037576532465
		log Likelihood difference (LD)	2.107375545952273	
		p-value	**0.041000000000000036***	
Animism	Shamanism	q12(alpha1)	0.017922679679795207	1.0453676822495886E-4
		q13(alpha2)	2.611746557985332	5.466212448651201E-7
		q21(beta1)	7.547530435731249E-4	3.2146092549083307
		q31(beta2)	0.8036310168494801	9.801534639612092E-4
		q24	-	2.986507679741015
		q34	-	2.3082172206318075
		q42	-	3.878578851844866E-5
		q43	-	0.6022309534477134
		log Likelihood (L)	-20.61113662702108	-18.43901665082429
		log Likelihood difference (LD)	2.1721199761967895	
		p-value	**0.029000000000000026***	
Animism	High Gods	q12(alpha1)	0.01792268529659327	1.0615647497622018E-5
		q13(alpha2)	0.16814429612889825	5.003935597094861E-7
		q21(beta1)	7.547560491754113E-4	0.9826259441965896
		q31(beta2)	0.2728853651313188	2.4115585431978422E-4
		q24	-	0.9668616658801443
		q34	-	0.4816886228320609
		q42	-	2.5009443596585324E-5
		q43	-	0.7121275394333404
		log Likelihood (L)	-24.58771065166488	-23.506430016941867
		log Likelihood difference (LD)	1.081280634723015	
		p-value	0.17900000000000005	

Character X	**Character Y**		**Independent model**	**Dependent model**
Belef in an Afterlife	Shamanism	q12(alpha1)	5.218879100103742	3.769966946383435E-5
		q13(alpha2)	9.1998518117755	0.6567008570002618
		q21(beta1)	1.6058090787256674	1.2789864728129487
		q31(beta2)	2.8307237237801117	0.7640944397067033
		q24	-	1.275241216945288
		q34	-	1.1884735867468457
		q42	-	0.21069875170004448
		q43	-	4.1732939719157696E-7
		log Likelihood (L)	-35.71413666880457	-28.747538310792255
		log Likelihood difference (LD)	6.966598358012316	
		p-value	**0.0***	
Belef in an Afterlife	Ancestor Worship	q12(alpha1)	4.062588573214016	0.07729929525467856
		q13(alpha2)	0.1316880632191157	0.3423011923963714
		q21(beta1)	1.2500272153269685	9.849092744435119
		q31(beta2)	0.15935550941025572	0.11805877538869874
		q24	-	1.859174871172054
		q34	-	0.13335933543247983
		q42	-	0.11193154654817793
		q43	-	9.717776444793614E-8
		log Likelihood (L)	-40.48177159569678	-34.65735223063223
		log Likelihood difference (LD)	5.82441936506455	
		p-value	**0.0030000000000000027***	
Belef in an Afterlife	High Gods	q12(alpha1)	4.4941435818152	1.5833328218838048
		q13(alpha2)	0.16814436626199697	2.428529741694923
		q21(beta1)	1.3828132507988233	3.1337798343938976
		q31(beta2)	0.2728854530140183	0.6503667749932317
		q24	-	7.829647431342137E-6
		q34	-	0.11160233996231926
		q42	-	0.17082008637351465
		q43	-	3.035481021877122E-8
		log Likelihood (L)	-39.69071419127651	-39.337955277123875
		log Likelihood difference (LD)	0.3527589141526377	
		p-value	0.736	

Character X	Character Y		Independent model	Dependent model
Shamanism	Ancestor Worship	q12(alpha1)	4.395323925248568	41.625709147923466
		q13(alpha2)	0.13168786236526236	0.13779603960001932
		q21(beta1)	1.3524075349737754	101.14497024117185
		q31(beta2)	0.15935523841411559	2.919637565200422E-7
		q24	-	0.5929928616459182
		q34	-	0.08712683229206851
		q42	-	0.1676741087519177
		q43	-	7.131724941334274E-7
		log Likelihood (L)	-40.48177159146289	-36.602667272039206
		log Likelihood difference (LD)	3.879104319423682	
		p-value	**0.010000000000000009***	
Shamanism	Active Ancestor Worship	q12(alpha1)	4.254852997372292	0.21941655975829275
		q13(alpha2)	0.024119413532695626	1.7896916645488917E-5
		q21(beta1)	1.3091859333853737	3.6826969861781986
		q31(beta2)	0.07192919182033762	0.03730735853886139
		q24	-	11.81473513324769
		q34	-	1.7747220847551834E-6
		q42	-	0.4163554402784801
		q43	-	0.059289818102207396
		log Likelihood (L)	-34.73547880446629	-28.717411054843915
		log Likelihood difference (LD)	6.018067749622375	
		p-value	**0.0010000000000000009***	
Ancestor Worship	High Gods	q12(alpha1)	0.13168795974909628	0.12272137671597863
		q13(alpha2)	0.16814428896352349	0.34750012782276024
		q21(beta1)	0.1593553811356608	0.2770546994789072
		q31(beta2)	0.2728853524995929	0.5496547163764938
		q24	-	6.226891899527741E-6
		q34	-	5.4257204773097305
		q42	-	2.236504233470032E-5
		q43	-	6.214980189961206
		log Likelihood (L)	-44.45834911234516	-43.560892189874494
		log Likelihood difference (LD)	0.897456922470667	
		p-value	0.529	

Character X	Character Y		Independent model	Dependent model
Ancestor Worship	Active Ancestor Worship	q12(alpha1)	0.13168794114432877	0.04236040487972608
		q13(alpha2)	0.024119429548873753	0.04301070470687921
		q21(beta1)	0.15935535472867446	7.93553170358179
		q31(beta2)	0.07192921083907078	0.12185613637053422
		q24	-	15.919166612679389
		q34	-	1.0147909229550855E-5
		q42	-	0.17943526424508463
		q43	-	4.3823101954721057E-7
		log Likelihood (L)	-39.5031137244075	-30.182930342847683
		log Likelihood difference (LD)	9.320183381559815	
		p-value	**0.0***	
High Gods	Active High Gods	q12(alpha1)	0.1681442651727595	3.642066233186282E-7
		q13(alpha2)	0.018183457243443422	0.25937024676865095
		q21(beta1)	0.2728853051596625	14.92329103446345
		q31(beta2)	0.09893275509479003	0.6885715270306392
		q24	-	3.919324295428966
		q34	-	0.06872501587988517
		q42	-	1.6855097188720912E-6
		q43	-	0.1061500550737619
		log Likelihood (L)	-33.579043578951854	-27.93444889583701
		log Likelihood difference (LD)	5.644594683114843	
		p-value	**0.0010000000000000009***	
Active High Gods	Active Ancestor Worship	q12(alpha1)	0.018183458135025643	0.03218896757368417
		q13(alpha2)	0.024119427628359044	0.00624367320658855
		q21(beta1)	0.09893274533510152	0.07723439624163879
		q31(beta2)	0.07192920929065762	0.04597475587572726
		q24	-	7.7471500624074565
		q34	-	6.597699131357802E-7
		q42	-	54.09955636099969
		q43	-	0.11583308097744709
		log Likelihood (L)	-28.623808191014163	-27.726982419315043
		log Likelihood difference (LD)	0.8968257716991204	
		p-value	0.5860000000000001	

e-mail: dudapa01@gmail.com

Human population history and evolution of culture: A phylogenetic approach
Ph.D. Thesis, 2017

Printed in the Czech Republic

University of South Bohemia in České Budějovice
Faculty of Science
Branišovská 1760
CZ-37005 České Budějovice, Czech Republic

Phone: +420 387 776 201
www.prf.jcu.cz, e-mail: sekret-fpr@prf.jcu.cz