

UNIVERSITY OF SOUTH BOHEMIA
FACULTY OF SCIENCE



**EVOLUTION OF HUMAN SOCIO-CULTURAL AND ECOLOGICAL TRAITS: A
PHYLOGENETIC (SUPERTREE) APPROACH**

Bc. Pavel Duda

Master thesis



Supervisor: prof. RNDr. Jan Zrzavý, CSc.

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

České Budějovice 2011

Duda, P., 2011. Evolution of human socio-cultural and ecological traits: a phylogenetic (supertree) approach. Master thesis in English, pp. 89, Faculty of Science, The University of South Bohemia, České Budějovice, Czech Republic.

Annotation:

Human species display complex intraspecies population structure and unparalleled behavioral and cultural diversity. In order to elucidate human population history and pattern of evolutionary change of socio-cultural and ecological traits, the first composite phylogenetic tree of 574 human populations (ethno-linguistic groups) was created on the basis of 129 recently published phylogenetic hypotheses based on genomic, genetic and linguistic data, utilizing supertree method matrix representation with parsimony. Subsequently, 56 selected socio-cultural and ecological characters based on ethnographic cross-cultural data were optimized on topology of obtained supertrees in order to reconstruct patterns of evolutionary change and states present in ancestral populations. The results are discussed in the light of recent studies of human phylogeography and cultural phylogenetic studies.

Declaration:

I declare that I worked out my Master thesis on my own and only using the cited literature.

I declare that in accordance with the Czech legal code § 47b law number 111/1998 Sb. in its valid version, I agree with electronic publication of my Master thesis, in its limited edition made by removing marked parts archived by the Faculty of Science, in the public access section of the STAG database run by the University of South Bohemia in České Budějovice on its web pages, while preserving my copyright to the text of this Master thesis. I further agree in accordance with the law number 111/1998 Sb. with online publication of the reviewers' and supervisor's assessments on the same portal as well as record of the defense procedure. I also agree with comparison of my Master thesis with the Theses.cz database run by National registry of theses and plagiarism - tracing system.

Prohlášení:

Prohlašuji, že svoji diplomovou 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 diplomové práce, a to v nezkrácené podobě – v úpravě vzniklé vypuštěním vyznačených částí archivovaných Přírodovědeckou fakultou - 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, 28.dubna 2011

.....
Bc. Pavel Duda

Acknowledgements:

Many thanks belong to my supervisor, prof. RNDr. Jan Zrzavý, CSc. for exemplary leadership. Particularly for enabling me to work on the topic of my personal interest while being always interested and helpful as far as his numerous duties allowed, for his useful comments, inspiring ideas, and thorough corrections of the manuscript. I also thank to RNDr. Jan Robovský, PhD. for the most useful recommendations concerning supertree construction and phylogenetic data management, to Dr. Brian McEvoy for providing data on genetic structure of Australian Aboriginals, and to doc. RNDr. Daniel Frynta, PhD., for initial consultations. I am obliged to Věra Řičánková for her assistance with obtaining cross-cultural datasets and the essential literature on cultural phylogenetics. The special thanks belong to Dr. Olaf Bininda-Emonds for useful consultation of the original topic of my master thesis and for discouraging me from it, which consequently led to the inception of my current research. Last but not least, I would like to thank my family and friends and to Jarmila Michálková.

CONTENTS

INTRODUCTION	1
Cultural evolution.....	1
<i>A short historical overview</i>	1
<i>Current state of the field of human phylogeography and cultural evolution</i>	4
Cultural phylogenetics.....	5
<i>In defence of cultural phylogenetics</i>	5
<i>Phylogenetic hypotheses (trees) of cultural phenomena</i>	8
<i>Phylogenetic comparative studies of cultural coevolution and adaptation</i>	9
<i>Some remarks on cultural phylogenies and current phylogenetic comparative studies</i>	10
Aim of this study.....	12
MATERIALS AND METHODS	13
Supertree.....	13
Input data.....	14
<i>Phylogenetic trees of languages and populations</i>	14
<i>Language phylogenetic networks</i>	15
<i>Phylogenetic trees of haplotypes</i>	16
<i>Population structure graphs (genome-wide SNP-based studies)</i>	17
Supertree-coding dataset.....	20
<i>Geographical and linguistic proximity</i>	21
Constructing and rooting the supertree.....	23
Character dataset.....	24
<i>Data integration</i>	24
Analyzing datasets and optimizing characters on phylogenies.....	27
RESULTS AND DISCUSSION	28
Phylogeny of the humankind: topology of the supertree(s).....	28
Taxa overview.....	31
Character evolution.....	37
<i>Case study 1: Coevolution of wealth transfers at marriage (bridewealth and dowry) and marital composition (mating system)</i>	38
<i>Case study 2: Evolution of postmarital residence (patri-, matri-, and ambilocality)</i> ...41	
<i>Case study 3: Evolution of socio-political complexity (jurisdictional hierarchy beyond</i>	

<i>local community</i>).....	44
<i>Case study 4: Male genital mutilation as an adaptation to sexual conflict</i>	46
CONCLUSIONS	49
REFERENCES	52
Table and figure legends.....	64
APPENDIX	74
List of socio-cultural and ecological characters.....	74
References.....	86
Figure legends.....	86

INTRODUCTION

Human species displays striking behavioral and cultural diversity both among and within individual populations (e.g. Brown et al., 2011). Human intraspecific diversity is, judging from the number of ethno-linguistic groups (languages), comparable to the species-level diversity of tetrapod classes: there is approximately 6,800 languages spoken world wide (Lewis, 2009; Greenhill et al., 2010), approximately 4,600 mammalian species (Bininda-Emonds et al., 2007), 10,000 species of birds, 8,300 species of reptiles, and 5,800 species of amphibians (Baillie et al., 2004). This diversity of “taxa” goes hand in hand with the outstanding range of ecological adaptations and cultural practices that enabled humans to inhabit virtually every environment on Earth.

Cultural evolution

A short historical overview

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. Biological evolution and cultural evolution are subject of long-standing comparisons and analogies. Several prominent figures of biological sciences had repeatedly pointed out that fundamental processes of biological evolution (such as cladogenesis, selection, drift, extinction, or mutation) have clear cultural analogues (e.g. Darwin, 1859; 1971; Huxley, 1942; Dawkins, 1976; see Mesoudi et al., 2006 for review). Some of these comparisons were drawn long before it was ascertained that “culture” is a phenomenon that is not entirely unique to humans (e.g. Whitten et al., 1999; Lycett et al., 2009; see Laland and Galef 2009 for the extensive review of the animal cultures). Darwin saw similarities between the evolution of biological species and the evolution of languages (Darwin, 1859; 1971), and nineteenth-century historical linguistics utilized the phylogenetic approach in a manner similar to systematic zoology.

It is not without significance that evolutionary thinking in linguistics has actually preceded the evolutionary thinking in biology. The first linguistic phylogenetic trees emerged decades before the first phylogenetic trees of biological species (see Atkinson and Gray, 2005; Borgerhoff Mulder et al., 2006; Mesoudi et al., 2006). The social scientists adopted and utilized evolutionary approach surprisingly early.

However, they have departed from it soon after. Independently on zoologists, cultural anthropologists realized the problem of statistical non-independence of taxa. This problem was recognized in 1889 by French anthropologist Francis Galton (and was since referred to as “Galton’s problem” by cultural anthropologists). It points out that similarity between cultures could be due to historical relationships (shared descent), or due to cultural borrowing. Without controlling for borrowing and shared descent one cannot make valid inferences regarding adaptive evolution (Eff, 2004).

Since then, cultural anthropologists attempt to control for Galton’s problem either by “taxon” sampling - the exclusion of closely related populations from the cultural datasets, the best known example being Murdock’s Standart Cross-Cultural Sample, (Murdock and White, 1969), or by the statistical removal of the inherited traits, both method causing a significant loss of valuable information contained in the data (Mesoudi et al., 2006). Also, these techniques might mitigate Galton’s problem, but not eliminate it (Eff, 2004).

Biological anthropology of the 20th century mostly focused on geographical and ecological correlates of human behavior, examining how social traits covary with geography, ecology, or also other social traits across cultures. Both science of cultural evolution and human phylogeography were not flourishing for methodological reasons and “historical and political” reasons (postmodern critics of anthropology, among other things, tend to be deeply suspicious of any attempt to infer evolutionary relationships as well as to make objective comparisons between cultures).

Early anthropologists, like Edward Tylor (1871), attempted to rank cultures along a continuum (to put them into linear order) from the most primitive to the most advanced. They shared the conviction that all cultures inevitably pass through certain stages. This notion of progress from savagery to civilization was – as having ethnocentric and racist connotation – later replaced by the principle of cultural relativism, established by Frans Boas and his students. Their notion that there can not be a relationship between culture and race became the central tenet of modern anthropology. The ultimate rejection of evolutionism in anthropology of the 20th century was merely caused by the fact that the proponents of this approach understood the evolution as a process of gradual progress and improvements. Such “ladder-like reasoning” presents, of course, a fundamental distortion of the process of evolution. 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 primitive-advanced continuum. No culture can be a priori considered primitive, or the lower grade of

the other, since the process of cladogenesis imply a tree-like pattern with many parallel branches (leading to the presence).

Another reason of stagnation of the science of cultural evolution was a long-standing theoretical issue concerning the dominant forces responsible for the cultural variation across cultures and its continuity over space and time. The key issue was the degree to which horizontal transmission (transfer of cultural information among contemporaneous cultures through intermarriage, trade, exchange, etc.) plays role in observed pattern of cultural variation of the world population. In other words: is the cultural evolution dominated by process of branching of cultural lineages analogous to divergence of biological clades, or by process of blending of two or more cultural lineages into one? This problem appeared in 20th and 21st century anthropology under various names: vertical vs. horizontal transmission; demic vs. cultural diffusion; cladogenesis vs. ethnogenesis; cladistic vs. rhizotic model of cultural evolution; branching vs. blending; family tree vs. entangled bank model; unilineal evolutionism vs. diffusionism etc. (Mesoudi et al., 2005; Borgerhoff Mulder et al., 2006; Collard et al., 2006; Lycet et al., 2009 etc.). The main idea behind all those terms is the widely accepted idea that human cultural evolution is much more complex and reticulated than biological evolution and that “cultural evolution is often far from tree-like” (e.g. Gould, 1987; Borgerhoff Mulder, 2001; Gray et al., 2007; Steele and Kandler, 2010).

The field of cultural evolution and cultural transmission was dominated by these two contradicting paradigms during the 20th century (Borgerhoff Mulder et al., 2006). This problem is by no means trivial and it is impossible to solve it by simply rejecting one of the opposing scientific convictions. The most appealing point of view would be that patterns of observed cultural variation should be considered trait by trait. Some cultural phenomena are result of long-term, vertical transmission (and probably descendants of very early cultural forms), while some were the product of extended borrowing (sometimes back and forth), and others represent true innovations (Borgerhoff Mulder et al., 2006). The long-standing conflict between “unilineal evolutionism” and “diffusionism” should not discourage evolutionary scientists from attempting to solve this conflict by testing the hypotheses that explicitly based on one of each paradigm. Galton’s problem and horizontal transmission certainly present a serious issue in cultural phylogenetics, but this should not serve as a reason against testing the hypothesis about cultural evolution explicitly and remaining in the state of “armchair speculations” (Greenhill et al., 2009).

The conflict between “unilineal evolutionism” and “diffusionism” has never been a struggle for domination over the field of science of human culture between evolutionary and social scientists. This was merely a conflict between the two factions of evolutionists about how explicit the analogy between nature and culture should be. One of the most prominent among critics of unilineal evolutionism was Stephen Jay Gould (1987) who wrote that “human cultural evolution proceeds along paths outstandingly different from the ways of genetic change.” The field of evolutionary science of human culture long remained in the state of drawing inspiring but rather vague comparisons between biological and cultural evolution (e.g. Dawkins, 1976; Dawkins, 1982; Blackmore, 1999) and lacked appropriate methodology that would allow testing explicit scientific hypotheses. The evolutionary science of human culture reached its maturity in the 80’s, thanks mostly to the work of Luigi Luca Cavalli-Sforza and Marcus Feldman (Cavalli-Sforza and Feldman, 1981). These researchers not only provided theoretical groundwork for analyzing culture in terms of modern evolutionary theory, but also developed rigorous mathematical treatments of cultural change inspired by population genetic models (Richerson and Boyd, 2005; Mesoudi et al., 2006).

Current state of the field of human phylogeography and cultural evolution

Increasingly better resolved phylogenies of human groups are published regularly. This includes works based on mitochondrial DNA for both partial (e.g. Gonder et al., 2007; Kong et al., 2010) and comprehensive human phylogenies (e.g. Ingman et al., 2000; Krause et al., 2010), or Y chromosome (e.g. de Filippo et al., 2011). Since the nineties, human phylogenies based on various autosomal genes were published (e.g. Cavalli-Sforza et al., 1994). Number of phylogenies is based on variation of the major histocompatibility complex – the human leukocyte antigen system (HLA) (e.g. Bannai et al., 2000; Garcia-Ortiz et al., 2006; Farjadian et al., 2009; Sulcebe et al., 2009; Arnaiz-Villena et al., 2010; Zhu et al., 2010). During the past decade, the human phylogenies based on entire individual genomes emerged (e.g. Rosenberg et al., 2002; Li et al., 2008; Tishkoff et al., 2009). Recently, these, increasingly larger and better resolved, genome-wide SNP-based studies are published regularly (Behar et al., 2010; Bryc et al., 2010; Huyghe et al., 2011; Rasmussen et al., 2010; Sikora et al., 2010; Xing et al., 2010; ; Xu et al., 2010; Chaubey et al., 2011; Henn et al., 2011; Listmann et al., 2011; Salmela et al., 2011; Wang et al., 2011 etc.).

The wide range of methods developed by evolutionary biology and phylogenetics has been utilized by either anthropologists or biologists themselves in order to tackle the nature of

cultural change empirically and quantitatively. The evolutionary history of human populations and cultures is explored using a variety of linguistic, archeological, and cultural datasets to test for detailed hypotheses about the historical patterns and adaptive functions of cultural evolution (Mace et al., 2005; Mace and Holden, 2005; Lipo et al., 2006, Mace and Jordan, 2011).

Some researchers focus on proximate mechanisms underlying cultural change in order to determine the appropriateness of using phylogenetic (tree-building) methods to study and to visualize cultural evolution, both through case studies (Guglielmino et al., 1995, Hewlett et al., 2002; Jordan and Shennan, 2003; Collard and Tehrani, 2005) and through simulation studies (Greenhill et al., 2009; Currie et al., 2010; Nunn et al., 2010). Some works examine how modes of transmission differ for various kinds of cultural traits (Guglielmino et al., 1995, Hewlett et al., 2002), others examine how modes of transmission of a single trait could be changed under different socio-ecological settings (Collard and Tehrani, 2005). Some studies examine how (or if) various degrees of vertical and horizontal transmission obscure cultural phylogenies (Greenhill et al., 2009, Currie et al., 2010). Nowadays, cultural phylogenetics is an emerging field, although the empirical branch of this field is still relatively small (Mace et al., 2005; Lipo et al., 2006; Mace and Jordan, 2011).

Cultural phylogenetics

In defence of cultural phylogenetics

Despite the continuing validity of “diffusionists” objection that attempts to “Darwinize” human culture are controversial and might be misleading (e.g. Gould, 1987; Gray et al., 2007 Borgerhoff Mulder, 2001), there are compelling reasons to adopt the branching (tree-like) pattern of cultural evolution to treat human cultures (populations) as a species for the purpose of phylogenetic cross-cultural analyses.

- 1) Cultural phylogenetics is, in contrast to social sciences, “strongly empirical” (Holden and Mace, 2005; Mace et al., 2005). The “tree thinking” is not just a way of describing evolution but also a way of testing scientific hypothesis (Greenhill et al., 2009). While the tree model might be considered imperfect and simplistic (as indeed is, even for evolution of many biological species), it is generally far more realistic approximation than competing models that assume all cultures to be (un)related to each other equidistantly (Currie et al., 2010; Mace and Jordan, 2011). That holds true even in the

case when phylogenies are obscured by relatively large amount of horizontal transmission (Mace and Jordan, 2011).

- 2) Cultural change and biological change share the same fundamental properties of variation, selection and inheritance, and culture is a subject to phenomena that works for biological evolution (e.g. convergent evolution, functional constraint, punctuated equilibria, or even random, non-evolutionary processes such as drift). Convincing evidence was collated that human culture is a subject to Darwinian evolutionary processes (Mesoudi et al., 2004, 2006; Whitfield, 2008).
- 3) Placing the cultural anthropology within a unified evolutionary framework might be justified, since cultural anthropologists and evolutionary zoologists often seek answers to similar questions, facing similar problems. Therefore, it is desirable for anthropologists to utilize various tools, theories, and methods that have been developed by evolutionary biologists in order to answer the questions about the evolution of (human) culture (Mesoudi et al., 2006). Certainly, such biological culture does not necessarily imply “tree-likeness”.
- 4) Language, playing the role of the “model organism of cultural evolution” (Nettle, 2007), is tremendously similar to genomes in the way it is transmitted (Atkinson and Gray, 2005; Nettle, 2007; Mace and Jordan, 2011). Linguistic data were used successfully to obtain detailed and reliable population histories (e.g. Holden, 2002; Rexová et al., 2006; Pagel, et al., 2007; Kitchen et al., 2009; Walker and Ribeiro, 2011) and to answer questions regarding ancient populations movement that could not have been addressed by other methods (Gray and Jordan, 2000; Gray and Atkinson, 2003). Language and genetic phylogenies are often conspicuously similar (Cavalli-Sforza et al., 1988; Cavalli-Sforza et al., 1992) and spatial correlation of genetic and linguistic diversities has been documented numerously (e.g. Novembre et al., 2008; Hunley et al., 2008; Abdulla et al., 2009; Xu et al., 2010 Wang et al., 2011). Moreover, apart from these non-causal correlation between interpopulation genes and languages there might be also causal correlation present, since genetic factors predetermining linguistic features (particularly linguistic tone) were found in worldwide human population (Deddiu and Ladd, 2007).
- 5) It was shown that cultural (linguistic, archeological, and anthropological) datasets and biological (genetic, morphological, and behavioral) datasets are similarly “tree-like”, measured by indices of tree-dataset fitting (e.g. consistency index and retention index).

The observed variation within cultural datasets could most likely be the result of branching process (Collard et al., 2006; Lycet et al., 2009).

- 6) Although some cultural anthropologists are convinced that cultural traits are so labile that they show no phylogenetic signature, mapping cultural traits onto linguistic or genetic trees revealed that many cultural traits show a strong association with phylogeny. Many also appear to be historically conservative (e.g. Guglielmino et al., 1995; Hewlett et al., 2002; Mace and Holden, 2005; Fortunato et al., 2006; Fortunato, 2011a), although this does not apply generally to all kinds of traits and vertical cultural transmission is not always the best way to explain observed cultural diversity (Jordan and Shennan, 2003). Also, mode transmission of single trait might change under different socio-ecological settings (Collard and Tehrani, 2005).
- 7) It has been demonstrated through simulation studies that realistic levels of horizontal transmission of traits between closely related cultures is not too problematic for tree-building. Also, horizontal transmission does not produce systematic errors in the ancestral state estimation (Greenhill et al., 2009; Currie et al., 2010). The existence of horizontal transmission between cultures therefore does not invalidate phylogenetic approach to cultural and linguistic evolution.
- 8) The phylogenetic network-building algorithms were developed that allow to detect signals conflicting with a pure phylogenetic hypothesis (bifurcating tree) within the cultural dataset hence to assess whether the tree model provides an adequate representation of grouping of the data prior the use of the actual tree-building algorithm (Bryant and Moulton, 2004; Bryant et al., 2005).
- 9) Last but not least, it is worth noting that even the use of the terms “horizontal transmission” and “vertical transmission” are predicated on an assumption of an underlying tree-like model. These terms would in fact be meaningless unless we believed that branching process did indeed underpin our population history and cultural diversification (Mace and Jordan, 2011).

Phylogenetic hypotheses (trees) of cultural phenomena

The field of cultural phylogenetics comprises of two related approaches or sets of techniques: “building phylogenies” and “using phylogenies” (Mace and Jordan, 2011).

The first approach – the tree building - assesses the phylogenetic signal of a studied cultural trait (either material or non-material). It considers the extent to which the similarities and differences in the trait states across societies can be described by tree-like structure. The proponents of this approach are constructing phylogenies of cultural artifacts, operating with the assumption that observed similarities among cultural artifacts are function of common ancestry and that the artifact frequencies and phenotypes are a result of evolutionary forces in action (Dawkins, 1976, 1982).

This approach, first utilized by nineteenth century anthropologists to create lineages of cultural artifacts such as coins, stone tools and pottery (Mesoudi et al., 2006; Shennan, 2008) has experienced its renaissance recently in works of Darwinian anthropologists that adopt explicit evolutionary models and methods (e.g. Lipo et al., 2006, Shennan, 2008). The recent examples of this approach include cladistic analysis of Californian Indian basketry (Jordan and Shennan, 2003), phylogenetic analyses of eighteenth century Turkmen weaving design patterns (Collard and Tehrani, 2005), investigating the peopling of North America through cladistic analyses of continent-wide sample of early Paleoindian projectile points (Buchanan and Collard, 2007), explaining the absence of the Levallois and Acheuleen Paleolithic technological traditions via phylogenetic analysis of stone tool morphologies (Lycett, 2007), detecting phylogenetic signal of Neolithic plant economies through cladistic analysis of archeobotanical assemblages from various sites of West Eurasia, from the Near East to northwest Europe (Coward et al., 2008). This approach is not restricted only to human culture, since Lycet et al. (2009) carried out the cladistic analysis of wild chimpanzee cultures to find out that vertical inter-group transmission has been the dominant process also in chimpanzee cultural evolution.

Linguistic phylogenies, that fall within this category as well, are by far the most frequent type of cultural phylogenies (Nettle, 2007). Language, especially its basic vocabulary, is considered an excellent proxy for inferring human population history (Mace and Pagel, 1994, Atkinson and Gray, 2005; Greenhill et al., 2008; Pagel, 2009; Mace and Jordan, 2011), and number of anthropological studies have modelled population history and migration using language similarities as a cue of evolutionary relatedness (e.g. Gray and Jordan, 2000; Holden, 2002). Language trees are considered to have such a strong historical signal for at

least two reasons. Firstly, language is a neutral trait (i.e. the forms of words themselves have no fitness implications), and secondly, it should be highly conservative and fit the idea of “cultural core” (strong pressures, so called conformist bias, maintain the languages in distinct forms consistently; e.g. Mace and Jordan, 2011). This can be supported by the fact that language and genetic phylogenetic trees are often similar (Cavalli-Sforza et al., 1988; Cavalli-Sforza et al., 1992) and by frequent spatial correlation of genes and languages (e.g. Novembre et al., 2008; Hunley et al., 2008; Abdulla et al., 2009; Xu et al., 2010; Wang et al., 2011). See chapter Phylogenetic trees for the examples of recent linguistic phylogenetic studies.

Phylogenetic comparative studies of cultural coevolution and adaptation

Along with studies assessing phylogenetic signal of cultural traits, another type of studies that deal with cultural macroevolution emerged recently. These studies use phylogenetic comparative approach (Mace and Pagel 1994; e.g., Fortunato et al. 2006; Jordan et al. 2009; Walker and Hamilton, 2011). They reconstruct the pattern of evolutionary changes of various socio-cultural traits in the history of human populations. They optimize characters based on cross-cultural data on linguistic (or genetic) phylogenies in order to test for correlated evolution on the tree or to reconstruct states possessed by ancestral populations (taxa).

These studies are based heavily on the works of Guglielmino et al. (1995) and Hewlett et al., (2002) who carried out the analyses of cultural traits in sub-Saharan African societies included in *Ethnographic Atlas* (Murdock, 1967; White, 1999) to conclude that modes of transmission varies for different types of cultural traits. Most of the traits examined (namely those associated with social structure and kinship) were found to be associated with proxies for historical relatedness (language). In other words, these traits appear to be highly conservative and vertically transmitted. Distribution of other traits, such as religion and architecture, was best explained by geographical proximity, i.e., by cultural diffusion; however, the majority of traits seem to follow more than one explanatory model.

Since then, there has been a rapid increase of phylogenetic methods to test hypotheses about evolutionary history of cultural traits (by optimization of traits onto independently constructed phylogeny of cultures). Examples of these works include investigating of the coevolution of pastoralism and lactose digestion capability in adults (Holden and Mace, 1997), the coevolution of mode of subsistence (hunting-gathering and agriculture) and fertility among the world populations through (Sellen and Mace, 1997), phylogenetic cross-cultural analysis

of the association of sexual dimorphism in stature and sexual division of labor (Holden and Mace, 1999), investigating the evolution of cattle-keeping (pastoralism) in relation to descent rules (matrilineality and patrilineality) in Bantu-speaking societies of sub-Saharan Africa (Holden and Mace, 2003), evolution of wealth transactions associated with marriage (bridewealth and dowry) in Indo-European-speaking societies of western Eurasia (Fortunato et al., 2006), post-marital residence rules (matrilocal and patrilocal) and descent rules in ancestral Austronesian and Indo-European societies (Jordan and Mace, 2007; Jordan et al., 2009; Fortunato and Jordan, 2010), evolution of socio-political complexity (number of levels of political authority beyond local community) in Bantu and Austronesian societies (Walker and Hamilton, 2010), evolution of long-house architecture in Native American hunter-fisher-gatherers of the Pacific northwest coast (Jordan and O'Neill, 2010), (co)evolution of conception beliefs (partible paternity) and post-marital residence among indigenous societies of lowland South America (Walker et al., 2010), or the pattern of change in marital composition (monogamy and polygyny) and postmarital residence (neo-, uxori-, and virilocality) in the history of Indo-European-speaking societies (Fortunato, 2011a; 2011b).

Some remarks on cultural phylogenies and current phylogenetic comparative studies

The recent studies of cultural coevolution and adaptation based on optimization of cultural characters on language phylogenies are intriguing and demonstrate that evolutionary methods can bring significant contributions to the cultural science. However, the method they use might be fundamentally flawed. These studies optimize socio-cultural characters mostly on linguistic (cognate-based) phylogenies. Only a few studies using this approach mapped socio-cultural characters on genetic phylogenies (Holden and Mace, 1997; Sellen and Mace, 1997); however, the phylogenies they used were rather poorly resolved and contained limited sample of human cultures (taxa).

Modern language phylogenies are detailed, fine-scaled, and congruent with archeological and historical evidence. Language and genetic trees are often similar, both reflecting the same underlying population history (Cavalli-Sforza et al., 1988; Cavalli-Sforza et al., 1992). Moreover, the majority of recent studies shows that genes, languages and geography are intercorrelated (e.g. Novembre et al., 2008; Hunley et al., 2008; Abdulla et al., 2009; Xu et al., 2010 Wang et al., 2011). There is vast evidence that linguistic data provide a good source of phylogenies (e.g. Mace and Pagel, 1994; Nettle, 2007; Greenhill et al., 2008; Mace and Jordan, 2011), although some contradictory evidence does exist (e.g. Borgerhoff Mulder,

2001; Gray et al., 2007; Steele and Kandler, 2010). By all means, language is transmitted not only vertically but also horizontally and the approach these studies use – to simply synonymize “language” and “people” phylogeny - is daring and susceptible to type I errors. Works that use this approach of mapping cultural traits on language phylogenies (rather than optimizing culture on “people” phylogeny) are optimizing culture on another culture. This factor might be the one to explain the astonishingly positive results of these studies (correlation of culture and culture is much less mesmerizing than coevolution of culture and genes).

This approach also allows number of errors to emerge due to switched causality. Consider for example the case of western Pygmies of Central Africa (e.g. Baka of Cameroon, Gabon and Congo) speaking Nigero-Kordofanian languages (Adamawa-Ubangi group; Lewis, 2009). Linguistically, Baka are closely related to populations who practice mostly extensive agriculture. By inference from linguistic phylogenies, the lifestyle of Baka who are forest-dwelling hunter-gatherers would be interpreted as a reversal (therefore an apomorphy). However, the language could be recently adopted and the population could belong to an entirely different (hunter-gatherer) group, so that hunting and gathering present retention of the ancestral state (a plesiomorphy). That is, in fact, the case of Pygmies who appear to be genetically related to South-African Khoe-Sans more closely than to West-African Nigero-Kordofanians (e.g. Tishkoff et al., 2009; Verdu et al., 2009; Sikora et al., 2011; Henn et al., 2011). With some exaggeration and simplification we could say that mapping the evolution of cattle-keeping on languages phylogeny (Holden and Mace, 2003) is similar to mapping the evolution of cattle-keeping (or any other cultural trait) onto phylogeny of human influenza viruses.

It is worth noting that these works (and this apply to both constructing and using phylogenies) are concerned over and over again with cultural phylogeny of Indo-European, Austronesian and Bantu speakers. Of course, inferring population and cultural history of certain groups is more approachable with regard to statistical methods and computational input required. The reason why population history of the Indo-European speakers is frequently reconstructed and fairly often used in phylogenetic comparative approach studies lies not only in its accessibility but in the nature of its history. History of Indo-European-speaking populations is “strikingly tree-like” (Rexová et al., 2001, Bryant et al., 2005) and it is thought to represent Neolithic expansion ((Diamond and Bellwood, 2003; Fortunato et al., 2010). The same (expansion into more or less empty spaces with limited possibility of the cultural borrowing) implies for

population history of Austronesian or Bantu speakers which can be reconstructed with precision. Evolutionary histories of other populations or languages are considerably more difficult to reconstruct using standard methods (Reesink et al. 2009).

The studies that keep examining cultural macroevolutionary processes almost exclusively on the background of well-resolved, well-studied language/population expansions, are limited: they are capable to present only cultural evolution of more or less derived monophyletic groups and only partial image of macroevolution of the cultural traits in question. Linguistic data are apparently unable to provide for the well-resolved deep global phylogeny of human ethno-linguistic groups (see Greenhill et al., 2010). The analyses limited to evolution of single higher taxa (e.g. Indo-European speakers) therefore lack reliable outgroups, and the attempts to reconstruct their ancestral states are seriously flawed.

Advantages and drawbacks of genetic/genomic phylogenies mirror those of linguistic phylogenies. Genetic/genomic phylogenies can provide reliable information about large-scale clustering of human “higher taxa”. They sufficiently illustrate inter-group relationships but often fail to deliver well-resolved phylogenies of closely related neighboring populations (see e.g. Salmela et al., 2011). Such populations are actually prone to gene transfer more than to the language transfer (Holden and Mace 2011).

Aim of this study

The reasoning behind our approach is incorporating both genetic and linguistic evidence into a single dataset (utilizing philosophical principle of total evidence) in order to obtain comprehensive phylogeny of human ethno-linguistic groups, with both higher- and lower-taxa relationships resolved.

The aim of this study was

- to construct such composite phylogenetic hypothesis (on the level of ethno-linguistic groups) based on number of recently published biological as well as cultural phylogenies, using the supertree approach (the “matrix representation with parsimony” method),
- to identify suitable social, cultural and ecological traits of various human ethno-linguistic groups and create a dataset describing the intergroup variability of these traits and integrate these data in the cladistic character matrix,

- to use various methods to reconstruct the tree topology and to optimize states of selected socio-cultural characters onto obtained tree topology, in order to reconstruct evolutionary history of these traits, and
- to compare my results to the results the recent studies of cultural coevolution and adaptation which optimized cultural traits onto partial phylogenies based on linguistic data.

MATERIALS AND METHODS

Supertree

Composite trees are now routinely produced for evolutionary analyses using supertree approach, which use existing tree topologies as their input data to create unique composite phylogenies (Bininda-Emonds, 2004). Phylogenetic supertree is a method that uses existing phylogenetic topologies as their input data to create rather unique composite phylogenies (Bininda-Emonds, 2004). Supertree results from combining many smaller, even if only partially overlapping phylogenetic trees, into a single, more comprehensive tree.

This method has since been used to provide some of the largest, most comprehensive phylogenies for diverse groups at various taxonomic levels, e.g. mammalian species (Bininda-Emonds et al., 2007), dinosaur genera (Pisani et al. 2002), or hexapod orders (Davis et al., 2010).

To construct phylogenetic supertree of human ethno-linguistic groups based on various types of input data (and in this case also various types of graphic representations of phylogeny) a supertree method standard matrix representation with parsimony (hereinafter “MRP”) was used (Baum, 1992; Ragan, 1992). MRP represented a universally applicable method that could combine even incompatible sets of source trees using existing phylogenetic software (Bininda-Emonds, 2004).

This method has several unique features among consensus methods that make it suitable for our goals: Firstly, it utilizes the topology of source trees, not the original data. Therefore, trees derived from different types of data (here molecular sequences, lexical data) and analyzed by different clustering techniques (e.g., maximum parsimony, maximum likelihood, neighbor joining, neighbor-net, split-decomposition, or Bayesian and maximum likelihood clustering algorithms like STRUCTURE and FRAPPE) can be combined. Secondly, source patterns are

evaluated on a more or less equal basis, so that the phylogenetic signal from datasets with a smaller number of characters is not swamped by those with a larger number. This method is less sensitive to conflict among source trees than are most conventional consensus techniques so that resolution is not necessarily lost as increasing numbers of conflicting trees are analyzed. Lastly and perhaps most importantly, the great advantage of this method is that trees which overlap only partially (due to different set of terminal taxa) can be successfully combined (Bininda-Emonds, et al., 1998).

The MRP supertree is done by collating phylogenetic hypothesis and translating those into “partitions” (partial matrices) using additive binary coding (see Fig. 1a). These partitions are merged (completely or partially, depending on the degree of taxa overlap in partitions) and the composite MRP matrix used for subsequent maximum parsimony analysis is created.

This study used various graphic representations of shared descent, relatedness, or phylogeny, based on various types of data, as sources of input data for constructing partitions for MRP supertree. Some of them were not used for this purpose before. Therefore, various sources of input data should be discussed one by one.

Input data

Phylogenetic trees of languages and populations

Modern linguists are reconstructing language histories and, by inference, population histories using the toolkit of phylogenetics on the basis of linguistic, most frequently lexical data (and especially of the core vocabulary) but also morphological and phonological ones (Dunn et al., 2005). During the past decade these tree-building methods and various linguistic datasets have been used to investigate the population expansions of various ethno-linguistic groups, especially the three large and well-known linguistic families, the Indo-European (e.g. Gray and Atkinson, 2003; Rexová et al. 2003; Pagel, et al., 2007), African Bantu (e.g. Holden, 2002; Rexová et al., 2006), and Austronesian (e.g. Gray and Jordan, 2000; Gray et al., 2009), but were applied also to other groups such as Arawak of lowland South America (Walker and Ribeiro, 2011), Semitic languages (Kitchen et al., 2009), Melanesians and Papuans (Dunn et al., 2005; Hunley et al., 2008) etc.

Population geneticists are creating genetic trees of human populations using various markers (HLA-A, HLA-B, red cell enzyme systems, serum protein systems, STRPs, pseudocholinesterase-1, color blindness, etc.) and various tree construction methods (most

frequently neighbor-joining tree on the basis of genetic distances). Numerous human phylogenies based on autosomal genes include, for example, the peopling of America (Tsuneto et al., 2003; Garcia-Ortiz et al., 2006; Arnaiz-Villena et al., 2007; 2010), genetic origins of the Japanese (Omoto and Saitou, 1997), genetic relationships of the populations in China (Chu et al., 1998), population genetic studies of indigenous Taiwanese (Jin et al., 1999), or of Iranian ethnic groups (Farjadian et al., 2009). There also exist genetic phylogenies concerning the whole world (e.g. Cavalli-Sforza et al., 1988; Cavalli-Sforza et al., 1994; Ayub et al., 2003). Standard phylogenetic trees can be, of course, based on not only autosomal genes but also on Y-chromosomal, mtDNA, or genome-wide patterns of variation.

The translation routine applies generally to trees containing information about successive branching regardless the method used for estimating phylogeny (UPGMA, neighbor joining, maximum parsimony, maximum likelihood, Bayesian MCMC algorithm etc.).

Language phylogenetic networks

Recently, there has been a rapid increase of application of phylogenetic networks in various evolutionary studies.

Phylogenetic network is, as opposed to the phylogenetic tree, the depiction of evolutionary history of the set of taxa, where the taxa are represented by nodes and their evolutionary relationships are represented by edges (Huson and Bryant, 2006). Specific type of phylogenetic network is the reticulated network, graphic depiction of evolutionary history that represent (visualize) more complex evolutionary scenarios that can not be accurately represented by a phylogenetic tree (such as hybridization, horizontal gene transfer, recombination, host-parasite coevolution, and of course, cultural transmission). Agglomerative methods for the construction of phylogenetic networks such as NeighborNet (Bryant and Moulton, 2004; Bryant et al., 2005) were utilized by evolutionary anthropologists in order to cope with the criticism of diffusionists (see Introduction) who repeatedly point out that use of explicit phylogenetic methods to make inference about history of human populations or evolution of cultural phenomena is invalid, since patterns of (not only) linguistic and cultural diversity might be strikingly reticulate. Since NeighborNet constructs reticulated networks rather than trees it is most useful to for initial analyses of any cultural (phylogenetic) dataset, for assessing the degree to which a tree structure provides an adequate representation of such dataset. It is capable of showing evidence of signals conflicting with a pure phylogenetic hypothesis, i.e. a tree (Bryant et al., 2005).

For this and other reasons, reticulated networks were used in number of recently published studies on human population history to show the amount of conflict in the phylogenetic data (by virtually showing the alternative tree topologies present in the whole sample within a single figure), to demonstrate that the degree of horizontal transmission in order to justify presenting the data in form of phylogenetic tree, or to show the particular cases of extensive cultural borrowing (e.g. creolization of language). The use of networking techniques in evolutionary anthropology includes studies of population history of various ethno-linguistic groups like African Bantu-speakers (Holden and Gray, 2006), populations of Sahul (Reesink et al., 2009), Arawak-speaking societies of lowland South America (Walker and Ribeiro, 2011) and even languages of the entire world (Greenhill et al., 2010).

Phylogenetic network is translated into matrix of binary additive characters in a way similar to translating an unrooted phylogenetic tree. Basal reticulations present in reticulated phylogenetic networks, “box-like sections” sensu Bryant et al. (2005), were treated as unresolved sections of phylogeny and coded as polytomies (see Fig. 1b).

Phylogenetic trees of haplotypes

Mitochondrial (mt) DNA has been a useful tool in our understanding of human evolution, owing to characteristics such as high copy number, lack of recombination, high substitution rate, and maternal mode of inheritance (Oven and Kayser, 2009). Similarly, the Y chromosome is suitable for investigating recent human evolution from a male perspective (Jobling and Tyler-Smith, 2003).

The studies of origins of human populations and population movement through mtDNA include numerous studies dealing with the evolutionary relationships of various human groups such as East Asians (Horai et al., 1996; Kong et al., 2010), click-speaking Africans (Tishkoff et al., 2007), sub-Saharan Africans (Gonder, et al., 2007), indigenous Taiwanese (Tajima et al., 2003), Native North Americans (Eshleman et al., 2004) and also the whole mankind (Ingman, M., et al., 2000; Krause et al., 2010). The same imply for studies of human evolution through analysis of Y-chromosomal variation (e.g. Wells et al., 2001; Semino et al., 2002; Tajima et al., 2004; B  r   et al., 2009; de Filippo et al., 2011).

Both mtDNA and Y-chromosomal phylogenies that were translated into supertree-coding dataset are of basically three forms. Apart from standard phylogenetic trees based on mtDNA or Y-chromosomal sequences and constructed by methods such as neighbor-joining or

UPGMA that show relationships of the studied populations, there are two more types of presentation.

The first type is phylogenies of haplogroup frequencies (e.g. Tishkoff et al., 2007; de Filippo et al., 2011). They comprise a phylogenetic tree of haplogroups (the branch here stands for haplogroup, not population) and a table that is listing the studied population and the frequencies (%) of concerned haplogroups within each population sample. The population was coded as present (“1”) in the given branch if frequency of a given haplogroup was over 10%. This led to the state where most of the populations in the tree were coded as present in more than one branch. The repetitive taxa were then merged and polymorphic states for them were coded (as if various taxa were both present and absent at some levels of phylogeny). The resulting input tree is frequently highly polymorphic (see Fig. 1c).

The last type is represented by phylogenetic tree of mtDNA genomes (e.g. Gonder et al., 2007; Kong, et al., 2010; Krause, et al., 2010): the tree tips stand for populations, not haplogroups, however, the same population can occur repeatedly at various positions on the tree. Again, the tree is translated as any cladogram and subsequently, the repeating populations are merged, using the polymorphism coding. The resulting input tree is frequently highly polymorphic although usually not as much as in case of haplogroup trees (see Fig 1d).

Population structure graphs (genome-wide SNP-based studies)

Apart from phylogenies based on mtDNA, Y chromosome, and segments of nuclear genome such as HLA, number of human phylogenies based on whole genomes emerged recently. These studies utilize large datasets from genomic databases like HapMap (The International HapMap Consortium, 2003) and the Human Genome Diversity Project (HGDP) (Cann et al., 2003; Cavalli-Sforza, 2005).

These genome-wide SNP (single nucleotide polymorphism)-based studies often use (except for trees) the non-tree-like outputs of STRUCTURE and FRAPPE analyses (Pritchard et al., 2000) to illustrate their results. STRUCTURE and FRAPPE analyses are Bayesian and maximum-likelihood-based algorithms developed to discover populations on the basis of recombining genetic markers. The visual output of STRUCTURE and FRAPPE analysis is a graph that represents genetic subdivision among populations based on Bayesian clustering analysis. It shows the proportions of individual multilocus genotypes attributable to clusters (denoted by K) indicated by different colors. The analysis itself assumes no grouping of

information and the individuals are arrayed by population/region/continent of origin and named only after the analysis.

These techniques have been used to infer the genetic structure and interrelationships of human populations worldwide (e.g. Rosenberg et al., 2002; Li et al., 2008; Tishkoff et al., 2009), in sub-Saharan Africa (Sikora et al., 2010; Henn et al., 2011), northern Europe (Salmela et al., 2011), South-East Asia (Wang et al., 2011), Australia (McEvoy et al., 2010), Pacific Islands (Friedlaender et al., 2008), South America (Wang et al., 2008), and elsewhere. This method has also been successfully used to explore the relations of the extinct individual (of an extinct human population) to extant human populations (Rasmussen et al., 2009). Apart from genomic patterns of variation, structure analysis has been used to infer the history of structural characteristics of languages (particularly the languages of Sahul; Reesink et al., 2009).

The population structure graphs contain some hierarchical information within that can be translated into matrix of additive binary characters and, consequently, a phylogenetic tree with hierarchical clustering and complete linkage. The information on presence/absence of a given population (culture) in each section is simply transformed into matrix component, “1” for presence and “0” for absence of the group within a given sections (see Fig. 1e). Some human populations carry clear evidence of recent genetic admixture caused by contact with “alien” population. There are numerous examples of such admixtures that are blurring the signal of the original (say “pre-Columbian”) human population structure, uncovered by whole-genome SNP-based studies. In case that a recent genetic admixture was suspected, the presence of a given “alien” genetic component in the population was coded as “?”. Numerous instances of suspected recent admixture include for example presence of modern Eurasian “Indo-European” genetic component in the genomes of the Aboriginal Australians that are indicative of genetic influence of European settlers (McEvoy, et al., 2010), presence of European genetic components in genomes of some Central and South American populations, most significantly of Mayans and Pima (Wang, et al., 2007; Tishkoff et al., 2009, McEvoy, et al., 2010) indicative of genetic influence by European conquerors, and/or European component in the genomes of Canadian and Greenland Inuit populations (Rasmussen, et al., 2010). In cases of apparently older and more elusive admixture events, such information was coded as “1” (most notable case is the significant presence of Bantu genetic component in genomes of African Pygmies and Khoe-San-speaking populations of sub-Saharan Africa (Tishkoff et al. 2009) testifying of partial assimilation of those populations by Bantu-speaking pastoralists.

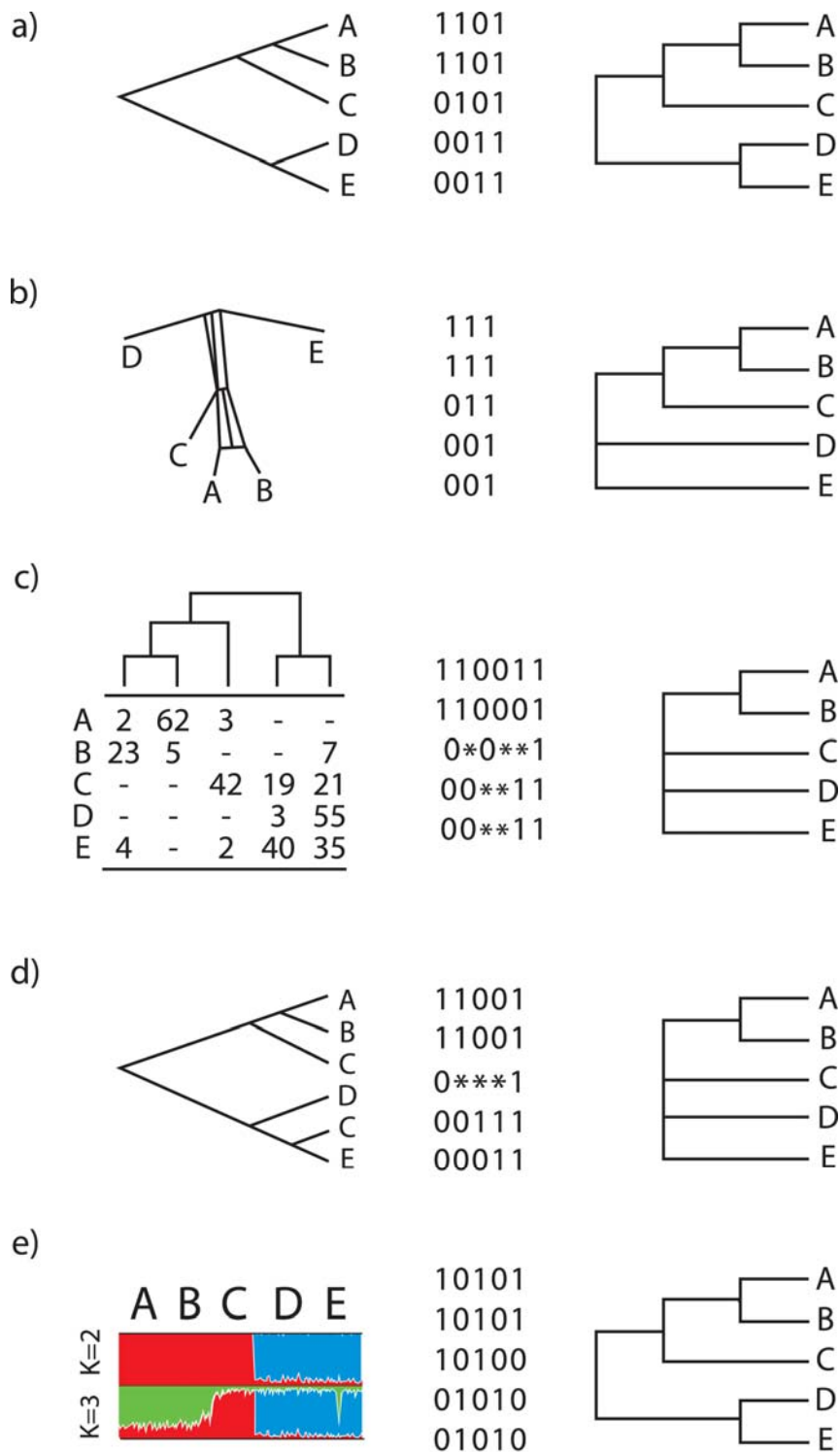


Fig. 1. The principle of translation of various types of graphic representation of shared descent, relatedness, or phylogeny, based on various types of data into partition of additive binary coding; a) standard phylogenetic tree; b) reticulated phylogenetic network; c) phylogenetic tree of haplogroup frequencies; d) phylogenetic tree of mtDNA genomes; e) population structure graph.

Supertree-coding dataset

Source studies were searched for online (Web of Science, Google Scholar etc.) using combinations of suitable keywords like human, phylogen*, evolution*, phylogeography, cladistic, tree, relationship*, population*, population structure, genetic structure, variation, diversity genom*, genom-wide, SNP, mt-DNA, Y-chromosom*, haplogroup, HLA, peopling, expansion*, language, linguistic, Bantu, Indo-European, Austronesia*, Africa, Europe*, India*, Thailand, Japan*, America*, Amerindian*, Hadza, Khoe*, San, Pygm*, Mbuti, Sandawe, Andaman*, Basque, Saami, Hungar*, Mongol*, Yukaghir, Formosa*, Ainu, Nivkh, Na-Dene, Arawak*, Mlabri, Tasmania*, Fuegian*, Yahgan etc.. The reference sections of obtained studies were searched for more potential sources of input data. Not only journal papers but also compendiums (Mace et al., 2005; Lipo et al., 2006) were used. Only studies published post 1990 were considered. The cut-off date of March 2011 was used.

The final dataset consisted of partitions based on altogether 129 input trees, reticulated networks, structure graphs etc. that came from 95 source studies. (See Supertree-coding dataset in electronic supplement for the partitions and information concerning source studies, source data, and phylogenetic methods used for creating the phylogenetic tree, network, or graph each partition was based on.)

The partitions of the combined supertree-coding dataset were then reduced dramatically in order to contain only ethno-linguistic groups that could be identified with an ethnographically documented culture (taxa whose cultural identity could be determined). The taxa denoted only by geographical location (i.e. states) were excluded from the dataset. Also taxa such as African Americans or Black Caribs of South America (Garifuna) that represent modern settlements overlaying the original patterns and amalgam of unrelated cultures were excluded from the final dataset. Extinct cultures (e.g. Akkadians, Hitties, Tocharians, or Saqqaq), with the exception of recently extinct Tasmanians, were not analyses. Most taxa were excluded from the dataset because their position on the supertre topology was not sufficiently supported (e.g. cultures present in only one partition or one source study).

The resulting dataset comprised of 574 taxa (excluding outgroups) and 5,437 “characters” (number of informative characters varies from 4,098 to 4,286, dependent on the type of rooting). There are 424 taxa (i.e. 74 %) that are present both in the supertree-building and eco-sociological datasets.

In order to compare the topologies based on combined and genomic data only, the additional supertree-coding dataset was created that consists exclusively of partitions based on genome-

wide SNP-based studies (altogether 22 studies; Rosenberg et al., 2002; Wang et al., 2007; Friedlander et al., 2008; Li et al., 2008; Abdulla et al., 2009; Kopelman et al., 2009; Li et al., 2009; Patin et al., 2009; Tishkoff et al., 2009; Verdu, 2009; Behar et al., 2010; Bryc et al., 2010; Huyghe et al., 2010; Rasmussen et al., 2010; Sikora et al., 2010; Xing et al., 2010; Xu et al., 2010; Chaubey et al., 2011; Henn et al., 2011; Listmann et al., 2011; Salmela et al., 2011; Wang et al., 2011). This dataset comprised of 246 taxa (excluding outgroups) and 1,591 characters, with 1,300 to 1,348 informative characters, and 177 taxa applicable for the optimization analyses (i.e. 72 % of taxa included).

In order to compare the results of ancestral state reconstruction in Indo-Europeans based on topology of the combined supertree with the reconstruction based on topology inferred from linguistic evidence, the additional supertree-coding dataset was created that consists exclusively of partitions based on linguistic (lexical) phylogenies concerning Indo-European-speaking populations (altogether 11 studies; Gray and Atkinson, 2003; Pagel and Meade 2005; Atkinson and Gray, 2006; Fortunato, et al. 2006; Pagel et al., 2007; Serva and Petroni, 2008; Geisler and List, 2009; Serva, 2009; Delmestri and Cristiani, 2010; Gray et al., 2010; Greenhill et al., 2010). This dataset comprised of 66 taxa (excluding the outgroup Hittite) and 586 characters (563 informative characters), with 50 taxa applicable for the optimization analyses (i.e. 76 % of taxa included).

Geographical and linguistic proximity

In order to prevent cultures (taxa) underrepresented in source trees to acquire obviously illegitimate (“wild-card”) positions on the supertree due to missing data, the topology of the combined supertree was further constrained by including information on geographic location and language affiliation of the analyzed populations (taxa). Topology of the purely genomic supertree was not constrained.

The geographic information was scored either as a single five-state non-additive character that coded World-Culture Regions as defined by Murdock (1967, i.e. Africa, Mid-Eastern, North Eurasian, East Eurasian, North American, South American; see White, 1999), or as 15 binary characters that coded presence or absence of a taxon in a given geographic cluster. These clusters were defined so they took into account not only geographic, but also genetic and linguistic boundaries and were loosely based on various, both traditional and most recent works from comparative anthropology, genomics and macrolinguistics (Burton et al., 1996,

Tishkoff et al., 2009, Lewis et al., 2009). The character definition and coding allowed some taxa to be coded as member of more than one cluster. The characters included in the geographic dataset are listed in Table 1.

South and Central Africa	including Khoe-San speakers and African Pygmies (The only discontinuous region)
West, Central, and South Africa	including mostly speakers of Nigero-Kordofan languages
West Africa	including mostly speakers of Nilo-Saharan languages
North Africa and Arabian Peninsula	including mostly speakers of Afroasiatic languages
West Eurasia	including mostly speakers of Indo-European languages (also creole languages), some Uralic languages (Hungarians, Lapps) and language isolates (Basque)
Southwest Asia and India	including both speakers of Indo-European and Dravidian languages and language isolates
Central and North Asia and Arctic region	largest geographical unit, including speakers of Altaic, Uralic, Dene-Caucasian, Chukotko-Kamchatkan, and Eskimo-Aleutian languages of Central and North Asia and North America as well as indigenous populations of Japanese Islands and Sachalin
East Asia	including mostly speakers of Sino-Tibetan Languages
South-East Asia	including mostly speakers of Austronesian languages (including Formosan and Malagasy) and other Indonesian, and oceanic cultures
Melanesia, New Guinea, and Australia	including Austronesian speakers of Melanesia and members of a Sahul lineage
New Guinea and Australia	including speakers of Indo-Pacific and Australian languages
Australia and Tasmania	including speakers of both Pama-Nyungan and non-Pama-Nyungan Australian languages and speakers of Indo-Pacific languages (Tasmanians)
America	including speakers of Amerindian languages and southern enclaves of Dene-Caucasian lineage
North America	including speakers of Amerindian languages and southern enclaves of Dene-Caucasian lineage
Central and South America	including speakers of Amerindian languages inhabiting Central and South America

Table 1. List of geographic/linguistic clusters used as binary characters to constrain the topology of the combined MRP supertree.

Moreover, the combined supertree topology has been constrained by 27 binary characters that coded presence or absence of a taxon in a linguistic macrofamily or larger linguistic group. The characters were based on the information from Ethnologue (Lewis et al., 2009), WALS database (Haspelmath et al., 2005) and the world's language families after Joseph Greenberg available at The Tower of Babel Database. The macro-linguistic characters were as follows:

Khoe-San, Nigero-Kordofanian, Bantoid (including Bantu), Nilo-Saharan, Afro-Asiatic, Afro-Asiatic (Chadic), Afro-Asiatic (Cushtic), Afro-Asiatic (Semitic), Afro-Asiatic (Berber), Indo-European, Dravidian, Uralic, Altaic, Korean-Japanese, Chukotko-Kamchatkan, Eskimo-Aleut, Na-Dene, Sino-Tibetan, Miao-Yao, Tai-Kadai, Austro-Asiatic, Papuan, Austronesian (including Formosan), Australian, Amerindian, Amerindian (Arawakan). The African Pygmies (Mbuti, Biaka, Baka, and Bakola), Adygei, Ainu, Basque, Burusho, Daghestani, Georgian, Ket, Nivkh, Anadamanese, Tasmanians, and Yukaghirs were treated as “language isolates”, not being positively scored for any of the 27 characters.

Constructing and rooting the supertree

All datasets were created in Winclada software (version 1.0000; Nixon, 1999). Supertrees were constructed by NONA software (version 2.0; Goloboff, 1999) via “Heuristic search” routine (multiple TBR + TBR search strategy).

Majority-rule consensus supertrees were constructed from supertrees based on various rooting options. Three rooting options were used that provide for nearly whole range of obtainable supertree topologies, with each output maximally dissimilar to others two. The three options were:

- 1) Rooting by one all zero outgroup followed by chimpanzee (“Chimp”)
- 2) Rooting solely by one all zero outgroup (“All-0”)
- 3) Unrooted tree (with Hadza placed on the base of the supertree) (“Unrooted”)

All zero outgroup is a made-up taxon that usually has all the characters coded as zero. In case of our dataset, only partitions where the input is a rooted phylogenetic trees or population structure graph were treated as rooted. Using the all zero outgroup is not appropriate for the unrooted trees or reticulated networks and the outgroup row was left empty in case of these inputs.

The chimpanzee-outgroup was coded only for partitions where the chimpanzee was present in the source tree. Chimpanzee outgroup occur in minority of trees of human populations, and therefore, the chimpanzee outgroup consisted of a nearly empty row (481 and 301 unambiguously coded “characters” in the combined and genomic dataset respectively).

Character dataset

The socio-cultural and ecological characters included in character dataset concerned social system, social complexity and stratification, community size, mating system, marriages, residence transfers, kinship terminology, succession and wealth acquisition rules, rituals, games, sex taboos, religious beliefs, slavery, architecture, subsistence ecology, division of labor etc. (see List of socio-cultural and ecological characters in Appendix and Character dataset in electronic supplement).

The characters were adapted from pre-coded variables in Murdock's *Ethnographic Atlas* (Murdock, 1967), its corrected version from Gray (Gray, 1999), and Standard Cross-Cultural Sample (Murdock and White, 1969). For the wealth transaction associated with marriage, additional data (28 taxa and 56 characters) were taken from Fortunato et al. (2006). Final character dataset in Winclada consisted of 1,269 taxa and 66 both binary and multistate characters, i.e. 83,809 character states. Among them, 17,907 character states were unknown and 2,189 inapplicable. Ambiguous character states account for approximately 24 % of all character states. Several characters have been coded in several alternative ways (see List of socio-cultural and ecological characters in Appendix).

Data integration

Since the anthropology lacks the standard "taxonomic" nomenclature (conversely to the Linnean binomic nomenclature or PhyloCode used in biology), various ethno-linguistic groups have more than one name. Numerous alternative names (at least different spellings) frequently exist for both (see Lewis, 2009). Moreover, there is a considerable problem of exonyms versus autonyms. The cultures (ethno-linguistic groups) present in the datasets under various names were synonymized using the descriptive information on societies in the anthropological literature and online databases (see References). Such information as geographic range of the population, geographical location (longitude and latitude) of the genotyped individual(s), language affiliation, ethnonyms, and alternate names of a culture were used in order to identify and match cultures within and between the partitions and datasets (White, 1986; Gray, 1999; Guthrie, 1967; Haspelmath, et al., 2005; Greenhill et al. 2008; Lewis, 2009). Another useful source of descriptive information and ethnonyms were the supplementary information present in studies of cultural coevolution that used the similar approach and were therefore dealing with the same issues (Holden and Mace, 2003; Fortunato et al., 2006; Jordan et al., 2009; Walker et al., 2010). Data provided in supplementary

information of Deddiu and Ladd (2007) was a useful cue to match large human populations from genomic databases with actual ethno-linguistic groups.

There was no general rule concerning naming of the taxa. The more up-to-date names from source trees were generally preferred over sometimes obsolete or slightly pejorative names in *Ethnographic Atlas* (consider “Bushmen”, “Eskimo”, “Ponapean”, “Semang” etc.). In general, I preferred autonyms over exonyms.

The taxa from the source publications were renamed and/or merged as follows:

“Central African Republic Pygmies” or “CAR Pygmies” or “Western” Pygmies were attributed to Biaka Pygmies. “Zaire Pygmies” or “ZAI Pygmies” or “Eastern Pygmies” were attributed to Mbuti Pygmies (Verdu et al., 2009).

Mbenzele pygmies were merged with Biaka pygmies in order not to exclude Mbenzele from the dataset due to their underrepresentation in source trees. This merging can be justified by the fact that Mbenzele and Biaka are closely related (e.g. Ingman et al., 2000; Coia et al., 2004).

“Maasai” from all source trees whose cultural/geographic affiliation was not closely specified were assumed to be Maasai of Kenya and Tanzania and listed as such in the dataset and combined with characters based on ethnographic data on Maasai living on the Kenyan-Tanzanian border. The exception is Maasai Ilchamus, Maasai Mumonyot and Maasai Il’gawesi, populations that were excluded from the final dataset.

All unspecified Philippine populations from source trees lacking language or geographic specification were considered Tagalog where it seemed plausible since Tagalog is the most widespread language of modern Philippines.

All Mongols (Mongolians) from the source trees were considered Khalkha Mongols (if not stated differently) and merged with ethnographic information on the latter.

Unspecified “Eskimos” from north-eastern continental part of North America present in source trees were all considered Copper Eskimos since Copper Eskimos are well documented ethnographically.

“Bedouins” is a generic name for desert-living nomads living in the area extending from the Atlantic coast of North Africa to the eastern coast of the Arabian Desert and speaking an Arabic dialect. All Bedouins in source trees were considered Rwala Bedouins (from Syria) if not stated differently

Two neighboring and linguistically closely related indigenous Taiwanese (Formosan) populations, Atayal and Taroko (Seediq) were merged due to their close relation and their uneven distribution within linguistic and molecular datasets. Taroko (Seediq) is considered merely the variety of Atayal in some sources (Lewis, 2009).

All Albanians from the source trees were merged into a single taxon and considered Gheg Albanians since the ethnographic data on Ghegs were available (Murdock, 1967; Gray, 1999). Similarly, all Tocharians, Armenians, Greeks (forming the taxon “Greek (modern)”), English, Czechs, Lusatians, Swedish, Sardinians, Bretons, Welsh, and Irish were merged into single taxa representing all forms and dialects of a given language. Afghan and Waziri were merged into a single taxon denoted “Pashtun (Afghanistan and Pakistan)” Nepali and Kashkura were merged, forming a single taxon denoted “Kashkura (Nepalese)”. Austronesian dialects that were merged into single taxa include Marshallese, Sangir, Ifugao and Manobo.

Rotokas and Aita were merged since Aita is one of the three dialects of Rotokas, the language of Bougainville Island (Lewis, 2009).

Both French Basques and Spanish Basques from all source trees were listed as “Basques” since the populations are closely related. With French Basques being prominent in source studies, Basques from supertree-coding dataset were combined with the character states based on ethnographic data on Basques from the French side of Basque geographic range (Murdock, 1967; Gray, 1999).

In number of input trees and STRUCTURE graphs, specifically those from genome-wide SNP-based studies based on HGDP and Hap-Map databases, there are several recurrently occurring taxa that apparently represent composite populations or geographic clusters that are very poorly specified. These “higher-level taxa” had to be matched with actual human populations present also in other source trees and character dataset (if possible). The group of taxa replacing a single larger taxon was coded as polytomy so no additional information was added into the supertree artificially.

Taxon “South African Bantu (HGDP-CEPH)” was represented by six populations/languages it most likely consists of: Ndebele, Sotho, Swazi, Tsonga, Xhosa, and Zulu (Deddiu and Ladd, 2007)

Taxon “Bantu (HGDP-CEPH)” was represented by the following cultures/languages: Bamoun, Kikuyu (Gikuyu), Mandinka, Ndebele, Sotho, Swazi, Tsonga, Turu, Xhosa, Yoruba, and Zulu.

Unspecified Bantu speakers of Kenya (HGDP-CEPH) were replaced by Kikuyu.

Taxon “North European” that recurrently occurs in genome-wide SNP-based studies represent English-speaking populations (originally Danish people) living in the USA. They were attributed to “English” if possible. This should not present a major problem due to a low resolution of trees with this “North Europe” taxon present (e.g Ayub et al., 2003; Xing et al., 2010).

Taxon “Indigenous Taiwanese” (Ingman and Gyllensten, 2003, Tajima et al., 2004) was represented by the set of nine Formosan societies: Amis, Atayal (Taroko), Bunun, Paiwan, Puyuma, Rukai, Saisiat, Tsou, and Yami.

Taxon “Melanesia (HGDP-CEPH)” was represented by Naasioi of Bougainville and by 22 other cultures of Bougainville, New Britain and New Ireland, following Friedlaender et al. (2008).

Taxon “Micronesia” was represented by the culture of Kusaie (Kosrae).

Taxon “Polynesia” was represented by Hawaiian and Maori.

Taxon “Papua (HGDP-CEPH)”, meaning rather central Papua (Highlands) was represented by three cultures – Gimi, Goroka, and Sepik. Fore population was used to support ethnographic data on Goroka, Kwoma population was used to support ethnographic data on Sepik (Murdock, 1967; Gray, 1999).

Until recently (McEvoy et al., 2010), the Australian taxon within genome-wide SNP-based studies was based on Native Australian samples of unknown ethnic population origin (provided by European Collection of Cell Cultures in Salisbury, UK) (Tishkoff et al., 2009). The taxon “Australia” was therefore replaced by the following seven well-documented Australian cultures that it might include, i.e. Aranda, Bininj Gun-wok, Meriam Mir, Tiwi, Warlpiri, Wongaibon, and Tasmanians

Analysing datasets and optimizing characters on phylogenies

The characters were mapped onto tree topologies obtained using various rooting. These topologies were constructed using datasets with the reduced numbers of taxa. Before optimization, 73 taxa (approx. 13 %) for which the character states were unknown were excluded from the final dataset in order to avoid ambiguous reconstruction of ancestral states due to the unknown character states of terminal taxa. However, some taxa with unknown

character states were kept within the dataset in order to preserve all the higher taxa whose ancestral states were reconstructed (for example Papuan-speaking Melanesians for most of whom ethnographic data was missing). Character data were then optimized on a supertree of 502 taxa (excluding outgroups) which account for the 87% of the taxa included in the final supertrees.

Apart from the optimization on the combined supertrees constructed using various rooting, characters were optimized on genomic supertrees as a control for alternative ancestral state reconstruction since the genomic-supertree topology differs from the combined supertrees in some respects. No taxa were excluded from dataset coding genomic supertree before optimization. One character (particularly the one concerning wealth transfers at marriage) was also optimized on the purely linguistic supertree of Indo-Europeans.

The character dataset was optimized on both the combined and genomic supertrees. For maximum-parsimony (MP) reconstruction of ancestral states, NONA software (version 2.0; Goloboff, 1999) was used (option “unam”). In some instances, maximum-likelihood (ML) method of optimization was used either to control for the accuracy of MP reconstruction, or to provide the alternative reconstruction of the key ancestral society if the MP algorithm failed to reconstruct the ancestral state (due to unknown or diverse character states across terminal taxa). For the maximum-likelihood reconstruction of ancestral states, Mesquite software (Maddison and Maddison, 2009) was used. The probability of distribution of states in the internodes was calculated via trace character history routine, using the majority-rule consensus of the combined MRP supertrees. One-parameter Markov k-state model (Lewis, 2001), a generalization of the Jukes-Cantor model, was used for ML reconstructions.

RESULTS AND DISCUSSION

Phylogeny of the humankind: topology of the supertree(s)

The resulting topology of the supertrees is congruent with both traditional and modern views of human phylogeography, population movements and fundamental relationships of the major world cultures (e.g. Guthrie, 1967; Murdock, 1967; Cavali-Sforza et al., 1988; Cavalli-Sforza et al., 1994; Burton et al., 1996; Holden, 2002; Diamond and Bellwood, 2003; Li et al., 2008; Gray et al., 2009; Tishkoff et al., 2009) with some exceptions that will be discussed below.

The basal topology of inferred phylogenetic supertrees is in line with phylogenies based on large genomic datasets (e.g. Li et al., 2008; Tishkoff et al., 2009; McEvoy et al., 2010) and also on genetic phylogenies concerning the whole humankind (e.g. Cavali-Sforza et al., 1994; Ingman et al., 2000; Ingman and Gyllensten, 2003; Krause et al., 2010). Fine-scaled phylogenetic structure within the large taxonomic (geographical) units and the topology of individual demic expansions are comparable to the topology of published linguistic phylogenetic trees (e.g. Gray and Jordan, 2000; Holden, 2002; Gray and Atkinson, 2003; Rexová et al., 2006). This can be demonstrated also by comparison of the combined and purely genomic supertree (see Fig. 2 and Fig. 3).

The effect of geographic and macro-linguistic dataset (for constraining topology of the supertree) on the resulting topology of the combined supertree was negligible. The inclusion of geographic and macro-linguistic dataset prevented a number of taxa from acquiring “wild-card” positions on the tree, but since they consisted of limited numbers of the “characters” (1 + 15 + 27 characters, respectively), they did not affect the fundamental topology of the supertree in any way. Both the combined supertree topology (constrained and unconstrained) and the purely genomic (unconstrained) supertree topologies are congruent with language and geography distribution, although this congruence is not perfect and exceptions do exist. The tentative parsimonious optimization of macro-linguistic dataset (see Geographical and linguistic proximity in Materials and methods) onto the supertrees showed that languages correlated with the combined supertree topology about as good as other cultural traits. The majority of language “superphyla” form monophyletic clades in at least one supertree topology, or at least most members of a language higher taxon fall within a single, monophyletic cluster. These language groups include: South-African Khoe-San, Nigero-Kordofanian, Nilo-Saharan, Afro-Asiatic (although only Semitic, Berber and some Cushitic languages); Indo-European (truly monophyletic in only one supertree topology), Indo-Iranians, Indo-European-speaking Europeans, Eskimo-Aleut, Northern Na-Dene, Southern Na-Dene, Austronesian and its subgroups, and languages of Sahul. The exceptions include the Chadic and Cushitic lineage of Afro-Asiatic languages, Khoe-San (when including Hadza and Sandawe), Dravidian (all in one monophyletic clade but along with some Indo-European-speaking populations), Sino-Tibetan (present within one monophyletic clade with populations speaking Uralic, Altaic, Korean-Japanese, Miao-Yao, Tai-Kadai, Austro-Asiatic and isolate languages), Na-Dene (as a whole), and numerous Amerindian language groups.

The supertree(s) topology does not fit on geography well only in case of Amerindians (probably due to the lack of the underlying information rather than absence of this pattern in American population). The other exceptions concern population isolates (Hadza of Tanzania, Andaman Islanders, Finnish Saami), or once connected populations, disrupted by more recent population expansions (“Paleo-Africans” – South African Khoe-San and Central African Pygmies isolated by the Bantu populations).

The supertree topologies based on various rooting options differ substantially (see Figures A1, A2, A3, and Table 2 for comparison).

The topology of the genomic supertree(s) is much less resolved compared to the combined supertree(s). However, the major taxa in the genomic supertree(s) correspond to those in the combined supertree(s) and their fundamental topology is similar (see Fig. 2., Fig. 3, and Table 2 for comparison). There are three notable exceptions.

1. Although in the majority-rule consensus of genomic supertrees South-African Khoe-San and African Pygmies form a basal monophyletic clade, in one topology of genomic tree (“Unrooted”) they cluster with central African Nigero-Kordofanian and Afro-Asiatic speakers. This association is caused by recent genetic admixture of these groups.
2. In case of South African Bantu, genetic and linguistic phylogenetic signals strongly contradicted. Bantu of South Africa (e.g. Xhosa, Swazi, Zulu) are classified linguistically among the most derived Bantu groups (Holden and Mace, 2002; Rexová et al., 2006), while genomic studies place them near the root of human evolutionary tree (see Tishkoff et al., 2009; Xing et al., 2010; Sikora et al., 2011). This is caused by the admixture of South Bantu with indigenous South African Khoe-San populations. Therefore, in the combined supertree, South Bantu are monophyletic sister group to East Bantu and along with them are the most derived Bantu-speaking taxon, while in genomic supertree, South Bantu are polyphyletic and some of them are basal and closely related to Khoe-San.
3. In the most topologies, “Sahul” (Melanesia, Papua, and Australia) falls within the “Australasian” clade, on its unresolved base. In one genomic supertree topology (“Chimp”), Sahul is a more basal taxon, sister to the vast majority of East Eurasian (including continental East Asia, America and Austronesia). This is in congruence with the notion that Sahul was settled relatively soon after modern humans left Africa (O’Connell and Allen, 2004), which was recently supported by genomic evidence (McEvoy et al., 2010).

Taxa overview

Sub-Saharan Africa (1) – Sub-Saharan Africans is large basal taxon that includes speakers of Niger-Kordofanian languages (spoken across a broad region of Africa), Afroasiatic (more specifically, Chadic and Cushitic) languages (spoken predominantly in Sahara and eastern Africa), Nilo-Saharan languages (spoken predominantly in Sudan, Sahara, and eastern Africa), and Khoe-San languages (spoken by San in southern Africa and by Hadza and Sandawe in eastern Africa). The major division of the world population into sub-Saharan Africa and the rest of the world (Eurasia, or “Afrasia” including also populations of North and Northeast Africa) that is well supported by number genome-wide SNP-based studies (e.g. Rosenberg et al., 2002; Li et al., 2008; Tishkoff et al., 2009; Xing et al., 2010),

Hadza and Sandawe – Hadza of Tanzania, the ethnic group in constant focus of evolutionary anthropology, often used as a proxy for the ancestral human society (see Marlowe, 2005), is among basalmost human groups in both the combined and genomic supertrees. Hadza acquire position either on the base of the whole tree, or they are the first taxon to separate within sub-Saharan Africa clade, and in one genomic supertree topology (“Chimp”), Hadza acquired position on the base of Afrasia. Sandawe, other Khoe-San speakers of Tanzania, fall within the taxon consisting of sympatric Bantu-speaking societies who they were largely assimilated by (i.e. Rangi, Turu, and Burunge).

Khoe-San and African Pygmies (2) – In most supertree topologies, South African Khoe-San (3) and African Pygmies (4) form a basal monophyletic clade or a paraphyletic assemblage that is a sister taxa to all sub-Saharan Africans or to the whole mankind. African Pygmies consist of the two groups – East Pygmies (Mbuti) and West Pygmies (Biaka = Mbenzele, Baka, and Bakola). West Pygmies show a strong tendency to form a monophyletic clade in both the combined and genomic supertrees. The position of East Pygmies (Mbuti) within this taxon is unstable. Also, Hausa of Cameroon tend to cluster with African Pygmies.

Nigero-Kordofanian (Bantu) – Populations of sub-Saharan Africa (5) speaking Bantu incl. Bantoid (5) languages form a paraphyletic cluster in all combined supertree topologies. Nigero-Kordofanian speakers as a whole are strictly speaking polyphyletic in all supertree topologies since Dioula of Burkina Faso and Dogon of Mali, placed as a sister taxa to North Africa (10) also speak Nigero-Kordofanian languages. In the combined supertree, phylogeny of Bantu speakers shows the same south-east gradient as in language phylogenies (Holden and Mace, 2002; Rexová et al., 2006) with south- and eastward migration being the terminal event of Bantu expansion. South Bantu (7) forms a stable and monophyletic clade in all the

combined supertree topologies. East Bantu (**6**) are paraphyletic due to their admixture with sympatric Nilo-Saharan and Afro-Asiatic-speaking populations.

East Africa (Nilo-Saharan languages) (8) – This stable taxon consisting of inhabitants of East Africa (mostly speakers of Nilo-Saharan languages) forms a terminal section of sub-Saharan Africa (**1**). These populations are according to all supertree topologies related to East Bantu (**6**) and Afro-Asiatic (Cushitic) language speakers. Nilo-Saharan speakers form a more or less monophyletic clade within this east-African group. This clustering contradicts linguistic classification (e.g. Lewis et al., 2009) but is in line with genetic (genomic) phylogenies which does not provide evidence for a monophyletic Nilotic clade (see Gonder et al., 2007; Tishkoff, et al., 2007, 2009).

North Africa (Afro-Asiatic languages) (10) – Afro-Asiatic languages (speakers of Semitic, and Berber languages and some populations speaking Cushitic languages) form a monophyletic or tightly paraphyletic group. In one combined supertree topology (“Chimp”), they form a basal monophyletic clade within West Afrasia (**9**), sister to monophyletic West Eurasians (**11**) (see Fig. A1). In two combined and all three genomic supertree topologies Afro-Asiatic populations form a paraphyletic (ladder-like) cluster that constitutes a “bridge” between Africa and Eurasia. This should be considered an artifact caused by various degree of genetic admixture of Afro-Asiatic populations with sub-Saharan African populations. However, in combined supertree topologies (“All-0”, “Unrooted”), the topology that Afro-Asiatic speakers form is not purely ladder-like. There are monophyletic sections present that correspond to linguistic classification of Afro-Asiatic languages (Cushitic, Berber) near the base of Eurasia (see Fig. A2 and A3).

West Eurasia (11) – A large monophyletic taxon, sister group to the Afro-Asiatic clade, present in one supertree topology (“Chimp”) based on the combined dataset (see Fig. A1). It includes two monophyletic clades – South Asia (**17**) (that includes Indo-European and Dravidian language speakers and isolates like Hunza = Burusho) and Europe (**12**) (including Indo-European speakers and isolates such as Basque, Saami, and Hungarian). In most supertree topologies, it forms a paraphyletic cluster.

Europe (12) – Monophyletic taxon that consists mostly of Indo-European-speaking societies, although the basalmost Indo-Europeans (Armenian, Greek and Albanian speakers) fall outside it. Along with Indo-European speakers, it includes language isolates (Basque) and Uralic-speaking people (Saami and Hungarian). The internal topology of this taxon shows resemblance to cognate-based linguistic phylogenies (e.g. Rexová et al, 2003; Gray and

Atkinson, 2003; Pagel and Meade 2005; Pagel et al., 2007; Serva and Petroni, 2008). There are three or four monophyletic taxa within European cluster present in combined supertrees. East Europe (**13**), taxon that consists mostly of Balto-Slavonic-speaking populations but also including Finns and Lapps (Saami). Genetic (genomic) associations of Lapps and East Europeans were found recently by Huyghe et al. (2011), however, this genome-wide SNP-based study had rather inadequate taxon sample. Northwest Europe (**14**) is a taxon that includes speakers of Irish, Breton, Welsh, and also Orcadian in two combined supertrees (“All-0”, “Unrooted”). Southwest Europe (**15**) consists mostly of Italic languages, including Haiti Creoles and isolated Basque (as a sister taxon to Spanish). The rest of the Europeans (**16**) fall within taxon that includes mostly Germanic-speaking populations in Europe and former dominiums (Sranan and Afrikaans) and also Hungarians (Uralic-) and Georgians (Kartvelian-speaking populations). It is worth noting that in genomic supertree, European populations also form a monophyletic clade.

South Asia (17) – Monophyletic or paraphyletic taxon that consist of two clusters – Southwest Asia (**18**), including populations speaking Indo-Iranian languages and language isolates such as Hunza (Burusho), and India (**19**). Southwest Asia (Indo-Iranian languages) is paraphyletic (in all topologies). The situation is somehow similar to that described in the Afro-Asiatic cluster: the Indo-Iranians form a “bridge” between Europe and India. India (**19**) is a monophyletic clade that diverges into two monophyletic taxa, one including only speakers of Indo-European languages, the other including Dravidian-speaking populations along with some Indo-European-speaking populations.

East Eurasia (20) – A large monophyletic clade that includes Circum-Pacific (**21**) and Australasia (**30**), and is either a sister taxon to monophyletic West Afrasia (**9**) in one combined supertree topology (“Chimp”; see Fig. A1) or a sister taxon to paraphyletic West Eurasia (**11**) in other supertree topologies

America (incl. Beringia) (22) – Beringia and America is a monophyletic cluster placed either on the unresolved base of East Eurasia (**20**), or it is the sister taxon of Far East (**29**), or a subclade of the Far East (in genomic trees). In two out of three supertree topologies the internal topology of this taxon is largely unresolved, but in one topology (“Chimp”), it is resolved and consists of three taxa, one including the populations of Beringia and Northern Nearctic, the other two including Amerindians and Southern Na-Dene.

Beringia and North Nearctic (23) – Beringia and North Nearctic includes speakers of Chukotko-Kamchatkan, Eskimo-Aleut languages, Nivkh (Gilyak) and Northern Na-Dene

languages. Chukotko-Kamchatkan language speakers represented by Chukchi and Koryak form either monophyletic or paraphyletic cluster basal to Inuits and Northern Na-Dene. Inuit **(24)** form strictly monophyletic cluster in only one combined supertree topology (“Unrooted”) but show tendency to monophyly in all topologies. Nivkh (Gilyak) cluster with Beringia and northern Nearctic. The linguistically defined group Na-Dene is shattered into two lineages – North Na-Dene **(25)** and South Na-Dene. In combined supertrees, South Na-Dene (Apache and Navajo) are monophyletic and cluster with Amerindians. North Na-Dene (Cree, Ojibwa, Alaskan Athabaskan, and Chipewyan) cluster with Beringia and North Nearctic group (are related to Inuit) and show tendency to monophyly.

Amerindian (26) – Amerindians either form a monophyletic group (in genomic supertrees), or, in combined supertree, they cluster into two taxa (“Chimp”) that lack any clear geographic or linguistic clustering or congruence with previously published phylogenies, or they form a largely unresolved cluster (“All-0”, “Unrooted”). This is probably due to lack of underlying information. For example, there are only five populations represented in the Human Genome Diversity Project collection (Cann et al., 2002; Cavalli-Sforza, 2005), and there exists only one genome-wide study concerned predominantly the American populations (Wang et al., 2007).

Far East (27) – This taxon is either monophyletic (in combined supertrees) or paraphyletic cluster (in genomic supertrees), positioned either on the unresolved base of East Eurasia **(20)**, or as a sister taxa to America (incl. Beringia) **(22)**, together forming the monophyletic taxon called Circum-Pacific **(21)**. In combined supertrees it consists of the unresolved base (Cambodians and Burmese) followed by two monophyletic sister taxa, one including the populations of Indochinese Peninsula (e.g. Lahu, Karen or Mlabri), the other including cultures of continental East Asia speaking Sino-Tibetan, Uralic, Altaic, Miao-Yao, Tai-Kadai, Austro-Asiatic languages (and language isolates). This paraphyletic cluster also includes Korean, Ainu, Japanese and Ryukyuan (Okinawan) which form a clade also including Manchu.

Australasia (30) – Australasia is a large monophyletic taxon that includes mostly Austronesian-speaking populations of Taiwan, Indonesia, Malaysia, Madagascar, Near Oceania (West Melanesia and Papua), and Remote Oceania **(42)**, along with some apparently relict populations like Malaysian or Philippine Negritos. It also includes one Papuan-speaking culture (Tobelo of Maluku Islands). In combined and two genomic supertrees, this cluster also

includes Sahul **(31)**. In one genomic supertree (“Chimp”) Sahul is a basal taxon sister to vast majority of East Eurasian populations, not specifically related to the Australasian populations.

Sahul (31) – Sahul is a continent that existed during the last glacial maximum, consisted of present day Australia, Tasmania, New Guinea, and surrounding islands. In our supertree(s), Sahul is a monophyletic, sometimes basal clade. The base of Sahul consists of both Austronesian and Papuan-speaking cultures. This amalgam of basal taxa must be the result of the recent genetic admixture (Friedlaender et al., 2008). The Sahul crown group diverges into two monophyletic lineages, one consisting of Papuan-speaking Melanesians (and some Papuans) **(32)**, the other of Australian Aborigines (and other Papuans) **(33)**. The latter group splits into two groups – monophyletic or paraphyletic taxon including predominantly the populations of Papua mainland (e.g. Gimi and Goroka) **(34)** and a monophyletic taxon that includes aboriginal Australians and Tasmanians **(35)**. In the majority of source studies used, Aboriginal Australians were present as a single composite taxon. Therefore, no conclusions should be drawn from internal topology of Australian taxon (at least in genomic supertree). However, it is worth mentioning that Andamanese (Onge) fall within Sahul. This clustering is derived from Reesing et al. (2009) and is in concordance with the view that Andaman islanders represent remnants of the ancient expansion from Africa to Australia.

Formosan (38) – Formosans, or indigenous Taiwanese form a monophyletic clade in all combined supertree topologies. This taxon is either sister taxon to the monophyletic Philippines **(39)**, internal taxa of Philippines, or is positioned on the unresolved base of Australasia. The monophyly of Formosans is in contrast to some linguistic phylogenies of Austronesian-speaking societies where Formosan group create a paraphyletic (ladder-like) or an unresolved clade on the base of the of all Austronesian societies (e.g. Gray et al., 2009; Jordan et al., 2009; Serva, 2009). Our result is based on combination of the linguistic, genomic (Friedlaender et al., 2008), and genetic (mtDNA) studies (Jin et al., 1999; Tajima et al., 2003; Tajima et al., 2004; Tsai, 2004). While in linguistic studies Formosans play role of the basal Austronesians, genetic and genomic studies also stress their affinities to continental East Asians and other groups of the region (Cavalli-Sforza et al., 1994; Chu et al., 1998; Ingman and Gyllensten, 2003; Friedlaender et al., 2008; Li et al., 2009). The Yami people – indigenous inhabitants of the outlying island Lanyu (Orchid Island) falls within Philippines in our supertrees. In genomic supertree, populations of Philippines also show tendency to form a monophyletic cluster which includes Formosans (Atayal (Taroko)).

Phylogeny of Austronesian-speaking cultures is interpreted as highly pectinate (ladder-like) by linguistic phylogenetics (see e.g. Gray and Jordan, 2000; Gray et al., 2009; Jordan et al., 2009; Greenhill, et al., 2010). Phylogenetic relationships of large Austronesian-speaking groups in the combined supertrees presented here are more or less pectinate, with major clades corresponding to islands and/or archipelagos. The supertree topologies (both combined and genomic) clearly show that today's population of Melanesia consists of two different settlements, the first being the result of the ancient, "out-of-Africa" migration (O'Connell and Allen, 2004; McEvoy et al., 2010; Krause et al., 2010), the other being the result of more recent Austronesian expansion (Gray and Jordan, 2000; Diamond and Bellwood, 2003; Gray et al., 2009).

Malayo-Polynesian (Western and Eastern) – Philippines (**39**) form either monophyletic or polyphyletic taxon (when including Formosan (**38**)) with stable internal topology, situated on the unresolved base of Australasia. The basalmost taxon within the Philippines clade is Badjau ("Sea Gypsies"). Philippine Negritos (Agta and Aeta) also cluster with Philippines. Malagasy of Madagascar falls close to the Borneo populations (**40**). This Borneo-Malagasy clade is present in all combined supertree topologies that also includes an outlier – Papuan-speaking Tobelo of Moloku Islands. It should be noted that the phylogenetic position of Malagasy in our supertree is based only on linguistic phylogenies. Other Malayo-Polynesian taxa in the supertree include monophyletic clades "Indonesia" (consisting of Austronesian-speaking populations of Java, Sumatra and Borneo), "Sulawesi" (including Austronesian Muna, Wolio, Makassar and Toradja of Sulawesi), and "Wallacea" (including Austronesian populations of Sulawesi, Flores, Nusa Tenggara, Maluku, and other islands of this area).

Oceania (41) – This taxon is monophyletic in one combined supertree topology ("Unrooted") and in others largely unresolved or paraphyletic. Oceania is a sister taxon to Austronesian-speaking cultures of Papua or Melanesia and includes Near Oceania (Austronesian-speaking Melanesia) and Remote Oceania (**42**).

Remote Oceania (42) – Remote Oceania is a terminal group of Austronesian-speaking cultures, monophyletic and well resolved in two out of three supertree topologies ("All-0", "Unrooted"). In combined supertrees, Remote Oceania (**42**) consists of two monophyletic sister taxa – Micronesian (**43**) (including Kusaian, Marshallese, Kiribatense, etc.) and Polynesian (**44**) (including Samoans, Mangarevans, Hawaiians, Maori, etc.).

Character evolution

Optimization of selected socio-cultural and ecological characters (describing cultural practices, social system, and socio-political complexity) on phylogeny allowed to infer the pattern of evolutionary history (and characters' coevolution) and to reconstruct ancestral states of these traits in hypothetical common ancestors of various contemporary human populations.

For example, the society ancestral to all contemporary human populations (0) can be reconstructed as living in monogamous, occasionally polygynous nuclear families, practicing bridewealth and patrilocality. Insistence of virginity of brides was likely to be absent. This society has lived in bands or petty chiefdoms (with one level of political authority beyond the local community at the most). Local communities were likely to comprise of 100-200 individuals. There have been no class stratification and no slavery. There has been, however, the office of local headman. Leadership was likely to have been inherited patrilineally, like the movable property. Descent rules were likely to have been patrilineal or bilateral, with "Iroquois" kingship terminology. Ancestral human society lacked the religious concept of utterly transcendent supreme deity (high gods). Male genital mutilation was likely to have been absent (if present, they were likely to have been performed at adolescence). Partial segregation of adolescent males might have been present. These people lived in dwellings with circular ground plan, floor formed by the ground and hemisphere-shaped roof. Housing was likely to have been constructed predominantly by females. This society was ecologically dependent on hunting and gathering, no agriculture or animal husbandry was present. Females spent appreciably more time gathering than males while hunting have been predominantly male business. Only games based on physical skills were present (games based on chance and strategic skills were likely to have been absent). It should be noted that there is a good deal of uncertainty and ambiguity inherent to cladistic reconstructions and the results must be interpreted with caution.

I have chosen to discuss the results (based on optimization of selected socio-cultural and ecological characters on global phylogeny based on total evidence) via comparison with the results of recent studies of cultural coevolution and adaptation that utilized phylogenetic comparative approach to study evolution of similar traits using phylogenies based on linguistic data (Fortunato et al., 2006; Jordan, 2007; Jordan et al., 2009; Fortunato and Jordan, 2010; Walker and Hamilton, 2010; Fortunato 2011a; 2011b).

Additionally, I discuss the pattern of distribution and possible presence of phylogenetic signal of socio-cultural practice that have not been investigated to date using phylogenetic approach – the male genital mutilation rituals. I made an attempt to reevaluate the validity of “Sexual conflict hypothesis” that suggest that male genital mutilation represent a hard-to-fake signal of a man’s reduced ability to challenge the paternity of other man in societies with possibly high frequency of extra-pair copulations.

Case study 1: Coevolution of wealth transfers at marriage (bridewealth and dowry) and marital composition (mating system)

Marriage is a universal human institution (Murdock, 1967) and therefore is of special importance for understanding human social and sexual behavior. In most human societies marriage settlements are associated with transfers of money, property, or with services. Whatever “symbolic” aspects mark these transfers, they also have important economic aspects and often represent considerable amounts of wealth (its accumulation can cause considerable distress to a giving individuals or families; Goody, 1973). Various models were proposed to explain the function and distribution of marriage transactions, especially the two seemingly opposing marriage transactions: bridewealth (that is given to bride’s kin by husband or husband’s kin) and dowry (given to the bride by her own kin). It might be important to point out that the emphasized “opposition” of bridewealth and dowry is actually quite misleading. Bridewealth is a transaction between the kin of the groom and the kin of the bride while dowry can be seen as sort of pre-mortem inheritance to the bride.

Evolutionary psychology interprets these cultural practices as s form of sex-biased paternal investment. Since the variance of reproductive success is (in human as in most animal species) greater for males than for females (Trivers, 1972), sons are more likely to benefit from the investment of wealth in polygynous societies. In monogamous societies, however, the inclusive fitness of parents can be increased by investing wealth into daughter in order to secure a high-status partner for her. Ultimately, bridewealth and dowry represent means of resource competition for desirable spouses among potential husbands and wives (or their families; Barrett et al., 2001). Therefore, bridewealth and dowry should occur in polygynous and monogamous societies, respectively.

The association of dowry and monogamy is more than apparent in the ethnographic record. Like monogamy, dowry is rare in the ethnographic record (compared to bridewealth and

polygyny) and its geographic range is largely restricted to West Eurasia. It was inferred (from the pattern of geographic distribution of these practices) that dowry is a practice with relatively recent development, compared to evenly distributed practice of bridewealth. This, however, might not be true.

Fortunato et al. (2006) attempted to investigate the development of marriage practices systematically, utilizing Bayesian MCMC phylogenetic comparative approach to reconstruct the evolution of two forms of wealth transfers at marriage, dowry and bridewealth, for 51 Indo-European populations. The results of optimization of this binary character onto phylogeny of Indo-European languages suggest that dowry and monogamy (as mating system associated with dowry) are most likely to have been the ancestral practices which remained predominant through the course of Indo-European history (Mace and Holden, 2005; Fortunato et al. 2006). Bridewealth, in this context, represents the evolutionary novelty of one lineage of moderately polygynous Indo-Europeans that consist of populations speaking Iranian languages (Fortunato et al., 2006). Recently, the pattern of change in marriage strategies in the history of societies speaking Indo-European languages was reconstructed by Fortunato (2011a), using the same approach, providing additional evidence for practicing monogamy in proto-Indo-European society.

There are various forms of wealth transfers and services associated with marriage. They include bridewealth, bride-service, reciprocal gift exchange, exchange of sister or female relative for a bride, and dowry (Murdock, 1967; Goody, 1973). Unordered multistate character was created based on data in *Ethnographic Atlas* (Murdock, 1967; White, 1999) that resembles the character used by Fortunato et al. (2006). However, the present coding covers greater variability of cultural practices. The practices which include transfer (of money, movable property, persons, or service) from the husband's kin to wife's kin were described as "Groom's family is expected to give" (hereinafter "bridewealth") as bride-service and exchange of female relative for a bride can be considered a form of bridewealth. The second character state was denoted "Bride's family is expected to give" (hereinafter "dowry"). Our character scoring compared to that of Fortunato et al. (2006) presents the third character state that accounts for the fact that in number contemporary societies (incl. Indo-European), marriage is not associated with formalized one-way wealth transfers (Goody, 1973). This character state was denoted "absence of consideration".

The character was scored as follows: **Char. 1** – Wealth transfers at marriage sensu Fortunato et al., 2006: **State 0** – Groom's family is expected to give (including bridewealth, brideprice,

bride service, and the exchange of female relative for a bride); **State 1** – Bride’s family is expected to give (dowry); **State 2** – Absence of consideration

The parsimonious optimization of characters onto phylogeny shows that bridewealth is likely to have been the ancestral state which remained predominant through the course of human evolutionary history (see Table 3). Our results indicate that bridewealth coevolves with polygyny on a global scale (see Fig. 6). Most ancestral societies of major human subgroups are likely to have been moderately polygynous and practiced bridewealth. There is a striking uniformity especially in sub-Saharan Africa (**1**) where the ancestral bridewealth is associated with ancestral moderate polygyny or harem polygyny (in the major clade that is sister to the basalmost African populations - Hadza, Khoe-San and Pygmies). Despite greater variation in modes of marriage among populations of Afrasia, MP reconstructions of ancestral societies are rather uniform: they were reconstructed as moderately polygynous and practising bridewealth (see Fig. 6).

Concerning the proto-Indo-European society, our results contradict those of Fortunato et al. (2006) and Fortunato (2011a). This is not given by sampling or uneven distribution of ambiguous states among terminal taxa but rather by the topology of the supertree. The relevant section of the supertree differs from the tree topologies suggested by the linguistic classification. There is no support for monophyletic taxon consisting solely of Indo-European-speaking cultures within combined or genomic datasets. Our results show that whether the Indo-European-speaking populations form a monophyletic clade (“Chimp” rooting) or paraphyletic cluster, the population ancestral to all Indo-Europeans (**11**) was likely to have been moderately polygynous and practice bridewealth, and this combination was the plesiomorphic state (inherited directly from the population ancestral to West Afrasia (**9**)). Only ancestral Europeans (**12**) and their subgroups (**13, 14, and 16**) were found to have been monogamous and practised dowry. Ancestral South Asiatic people (**17**), including speakers of Indo-Iranian languages, and their subgroups (**18,19**), were polygynous and practised bridewealth.

The tentative ML optimization of our data on the topology of MRP supertree of Indo-European populations based *solely* on linguistic trees (see fig. 4) shows that with phylogeny based solely on linguistic classification, although obtained by different method, the results similar to those of Fortunato et al. (2006) are obtained (See Fig. 5).

In case of Indo-Europeans, the state “Absence of consideration” most likely represents relatively recent shift from dowry. (ML optimization suggests the absence of consideration

could have been the ancestral state for the Southwest Europe (15).) However, this state is fairly common in Afrasia (not in Africa) and its cultural (macro)evolutionary origins are more common than those of dowry. Although MP optimizations fails to account for this, ML optimization indicates that the “Absence of consideration” could have been the ancestral state for Remote Oceania subgroups (42, 43, 44), and it was likely to have been the ancestral state of all Austronesian-speaking Pacific cultures (41). It could also have been the ancestral state of Philippine (39). The absence of consideration is also fairly common among populations of America. However, the poorly resolved internal topology of this taxon does not allow to draw any conclusions regarding the ancestral state of any particular American subgroup.

There has been a long-standing conviction that dowry and monogamy results from a shift from less complex societies towards more complex societies. However, this conviction possibly stems from the lack of principled and systematic investigation of the development of these socio-cultural phenomena. From the comparison of optimization of wealth transfer at marriage and social complexity (see Fig. 6 and Fig. 8), it can be inferred that there might be the coevolution of dowry and social complexity present among Indo-European populations, however, this tendency definitely does not apply globally.

Case study 2: Evolution of postmarital residence (patri-, matri-, and ambilocality)

Post-marital residence rules specify the sex-specific dispersal and kin associations (they describe the kin group with whom a couple lives after marriage). These rules are often, but not always, correlated with the descent rules (and other social norms and cultural practices). It was long hypothesised that changes in the residence system would cause changes in descent, not the other way around (e.g. Marlowe, 2004; Jordan, 2007).

The evolution of postmarital residence was recently investigated via phylogenetic comparative methods in Austronesian populations (Jordan, 2007; Jordan et al., 2009; Fortunato and Jordan, 2010), in Indo-European populations (Fortunato and Jordan, 2010; Fortunato, 2011b), and in populations of lowland South America (Walker et al., 2010). They came to conclusion that, in Austronesians, postmarital residence coevolve with descent system in a way that changes in postmarital residence precede changes in descent system (Jordan, 2007) and that ancestral Austronesian society practiced matrilocality (uxorilocality) (Jordan et al., 2009; Fortunato and Jordan, 2010). Ancestral Indo-European society practiced patrilocality (virilocality) (Fortunato and Jordan, 2010; Fortunato, 2011b). Matrilocality was

found to be relatively unstable compared to patrilocality that, once gained, is rarely lost (Jordan, 2007; Jordan et al., 2009; Fortunato and Jordan, 2010; Walker et al., 2010).

The multistate and unordered character was created based on data in *Ethnographic Atlas* (Murdock, 1967; White, 1999) that resembles the character used for optimization by Jordan et al. (2009). The post-marital residence rules that imply living with husbands kin were described as “patrilocal (including virilocal)” (hereinafter “patrilocal”). The second character included states when postmarital residence rule imply living with wives kin were described as “matrilocal (including uxorilocal and avunculocal)”, hereinafter “matrilocal”. The difference of our character scoring compared to that of Jordan et al. (2009) present the third character state that accounts for the situation when there is no strict rule for postmarital residence and the couple may live after marriage either with husband’s kin or with wife’s kin or elsewhere. This character state was denoted “ambilocal (including neolocal)”, hereinafter “ambilocal”.

The character was scored as follows: **Char. 2** – Postmarital residence sensu Jordan et al., 2009: **State 0** – Patrilocal (including virilocal); **State 1** – Matrilocal (including uxorilocal and avunculocal); **State 2** – Ambilocal (including neolocal).

The results of MP optimization of characters onto phylogeny show that patrilocality is likely to have been the ancestral state which remained predominant through the course of human evolutionary history (see Table 3). Most ancestral societies of all major human subgroups are likely to have been patrilocal (see Fig. 7). The apparent stability of patrilocality and instability of matrilocality is in line with findings of previous studies (Jordan, 2007; Jordan et al., 2009; Fortunato and Jordan, 2010, Fortunato 2011).

Our results indicate that society ancestral to all Indo-Europeans (**12**) was likely to have been patrilocal and patrilineal. The reconstruction of early Indo-European patrilineality is in line with the prevalent scenarios derived from the linguistic evidence and with findings of other studies utilizing phylogenetic comparative approach (Fortunato and Jordan, 2010, Fortunato 2011). Also, societies ancestral to European subgroups (**13, 14, 16**) were reconstructed as patricentric and bilateral, except for society ancestral to Southwest Europeans (**15**) that was reconstructed as ambilocal and bilateral.

Our results indicate that the society ancestral to all Austronesian-speaking populations (**30**) was likely to have been patrilocal. Societies ancestral to Oceania (**41**), Remote Oceania (**42**), Micronesia (**43**), and Polynesia (**44**), were also reconstructed as patrilocal. Matrilocality is relatively common in Austronesian-speaking Oceanic societies but since matricentric societies are, in the topology of our supertree, rather terminal taxa within Austronesian subgroups (that

applies especially to Micronesia), matrilocality would originate later in Micronesian lineage of Austronesian-speaking populations and other instances of matrilocality within this group represent a few unique transitions from patrilocality.

Matrilocality occurs in Formosans (**38**) whose ancestral population was, however, reconstructed as patrilocal, and in Philippine (39) whose ancestral population was reconstructed as ambilocal. This is rather important since in linguistic phylogenetic trees, the inferences of Jordan (2007), Jordan et al. (2009), and Fortunato and Jordan (2010), and Fortunato (2011) are based on, Formosans and Austronesian-speaking Philippine represent the basal taxa of rather pectinate topology, while in our supertree(s), they represent monophyletic clades within more or less polytomic section of the tree. Apart from different topology and character scoring that accounts for the existence of ambilocality, our results could also be caused by limited sampling compared to that of Jordan et al. (2009).

Matrilineal descent is fairly rare and was not likely to have been ancestral for any particular geographic or linguistic group with the exception of Equatorial West Africa (West-Central Bantu) where ancestral matrilineality is associated with ancestral patrilocality, although matrilocality is also common among populations belonging to this group (see Fig. 7). Ancestral matrilineality was possibly present also in Polynesia (**44**) according to ML optimization. The society ancestral to Oceania (41) was likely patrilineal while the societies ancestral to Remote Oceania (42), Micronesia (43), and Polynesia (44) were reconstructed with ambiguous character state, so they might have been ambilineal or matrilineal. In Micronesia, the prevailing ambilocality is associated with prevailing matrilineality, in Polynesia, the prevailing patrilocality is associated with prevailing ambilineality (see Fig 7).

Ambilineal descent is, on the other hand, common in all regions except sub-Saharan Africa and could have been the ancestral state for number of large linguistically or geographically defined groups, e.g. Europeans (**12**), Ancestral America (incl. Beringia) (**22**) and some of its internal subgroups (**23, 24, 24**), and also Australasia (**30**) which includes all populations speaking Austronesian languages. Our results concerning proto-Austronesians are therefore in contradiction with results of Jordan (2007), Jordan et al. (2009), and Fortunato and Jordan, 2010 but in congruence with Lane (1961) who proposed, relatively long ago, selection for malleable social structures in early history of Austronesian societies.

The pattern of evolution of postmarital residence and descend rules closely resembles that of wealth transfers at marriage. The loss of predominant state, patrilocality, to matrilocality is uncommon, but loss of predominant strict rule, here patrilocality, to absence of rule (here

ambilocality, or in case of wealth transfers at marriage, the “absence of consideration”) is much more common. This might imply that socio-cultural characters for optimizations on phylogenies should be defined as “rule present” – “rule absent” rather than as optimization of two strict rules that are often only seemingly opposing, as in the case of bridewealth and dowry (Goody, 1973). Perhaps there is potential for generalization: concerning social organization, the transition from one strict rule (e.g. bridewealth, patrilocality, patrilineality, or harem polygyny) to another (e.g. dowry, matrilocality, matrilineality, or strict monogamy) is far less common than the transition to less strict, more fluid state (e.g. absence of consideration or reciprocal gift exchange, ambilocality, abilineality, or moderate polygyny).

Case study 3: Evolution of socio-political complexity (jurisdictional hierarchy beyond local community)

How and why small-scale societies evolve into large-scale societies, and how and why large-scale societies inevitably fail is one of the fundamental questions of anthropology (e.g. Johnson and Earle, 2000; Diamond, 2005). The transition from foraging groups to agrarian states is a gradual and complex process that is linked to number of demographic and socio-ecological factors. Among factors that are thought to determine the level of socio-political complexity are community size, population density, mode of subsistence, and the intensity of agriculture (and also the degree of economic dependence on agriculture). Walker and Hamilton (2010) utilized Bayesian MCMC phylogenetic comparative approach to evaluate pattern evolution of socio-political complexity in Austronesian and Bantu language expansions using language phylogenies.

The *Ethnographic Atlas* (Murdock, 1967; White, 1999) provides data on socio-political complexity described as number of levels of “political authority beyond local community”. The character used for optimizations was identical to the one used by Walker and Hamilton (2010).

It was scored as follows: **Char. 3** – Socio-political complexity sensu Walker and Hamilton, 2010: **State 0** – Absent (no political authority beyond community); **State 1** – Petty chiefdom (one level); **State 2** – Large chiefdom (two levels); **State 3** – State (three or more levels). Conversely to Walker and Hamilton (2010), the positive state changes (rises) in social complexity are referred to as ascensions, and the negative state changes (declines) in social complexity are referred to as descensions.

Our results show, on the global scale, the similar pattern of evolution of socio-political complexity that Walker and Hamillton (2010) found using linguistic phylogenies of Bantu and Austronesian-speaking populations. The evolution of socio-political complexity is not simple, linear process, but a more complex, wave like process, when social complexity tends to build up and decline in an incremental fashion. Both MP and ML optimization suggest that the predominant level of social complexity is petty chiefdom (one level of political authority beyond community), which seems logical given the inherent demographic instability of populations during expansion phases.

Compared to social norms and cultural practices, social complexity is less stable on phylogeny and its evolution is much less linear. Social complexity is also the character which evolution and ancestral states are more difficult to reconstruct using MP algorithm (see Fig 8). This is due to multistate nature of the character and the uneven distribution of character states among terminal taxa. This is particularly true for East Eurasia (**20**) and its internal groups.

ML character optimization (see Fig. A4 in electronic supplement) suggest that ancestral society of West Eurasia (**11**) was likely to have been complex (either states or chiefdoms), and ancestral society of Europe (**12**) was likely to have been highly complex (either states or large chiefdoms). Ancestral societies of South Asia (**17**), Southwest Asia (**18**), and India (**19**) were likely to have been highly complex (states). Ancestral society of Far East (**27**) was reconstructed as ambiguous or as highly complex (states) if including Cambodians and Burmese, which would imply that numerous descensions occurred during consecutive population history of this region. Other cases of high ancestral social complexity includes East Africa (East Bantu) (**6**) (large chiefdoms) where societies developed into states in some cases and ancestral Australasia (**36**) (large chiefdoms) where societies experienced multiple descensions later and these descensions were followed by later multiple ascensions in societies of Indonesia, Madagascar and Remote Oceania.

Our results indicate that several human groups experienced radical decline of social complexity (descension), usually associated with decline of local community size and change in mode of subsistence. These descensions of social complexity include:

Ancestral America (**22**) and ancestral Beringia and North Nearctic (**23**) and also the internal groups of North Nearctic taxon, Inuit (**24**) and North Na-Dene (**25**). This descensions might have been associated with entering Americas via Beringia and reaching the marginal habitat of northern Nearctic. It might also have been associated with the “Beringean standstill” (Tamm et al., 2007). Another descension concerns ancestral society of Sahul (**31**) and its

internal groups, particularly ancestral Australian Aborigines (36). This descension might have been associated with migration to different environment and the abandonment of horticulturalism to hunting and gathering. It is unclear whether the low level of socio-political complexity in populations of Sahul (31), especially of Australia (35), represents the secondary loss of complexity or retention of the ancestral state. Our results suggest that some secondary loss of socio-political complexity occurred in populations of Sahul. Although low level of social complexity in Australian Aborigines is generally thought to represent the ancestral state since Aboriginal Australians are descendants of early out-of-Africa migration (e.g. O'Connell and Allen, 2004; Krause et al., 2010; McEvoy et al., 2010), to keep this assumption valid would require to assume a high number of multiple ascensions to have occurred independently in Afrasia.

Another apparent descension concerns ancestral Philippine (39) and Austronesian-speaking Melanesians. This descension might have been associated with decrease of local community size in the initial stages of Austronesian expansion. It is worth noting that ascensions occurred later in the history of Amerindians and in Austronesian expansion, but not in Sahul, perhaps since Aboriginal Australians have never reached more favorable environment and their continuing expansion to the south and into more marginal habitats led to one of the most radical reductions of social complexity in recent human history in case of Aboriginal Tasmanians (see e.g. Diamond, 1993).

Our results suggest that the process of expansion is also associated with loss of class stratification (wealth distinctions or hereditary aristocracy). Our results also suggest, although it should be interpreted with caution, that human societies tend to lose "faith" (the religious concept of high gods interfering with human affairs) during phases of demographic and/or geographic expansions. Some ancestral societies appear to have lacked high gods entirely. Class stratification and belief in utterly transcendent supreme deities are also linked to the social complexity (see Table 3).

Case study 4: Male genital mutilation as an adaptation to sexual conflict

Male genital mutilation (hereinafter MGM) is any permanent modification of the external genitalia that involves the ablation of tissue and is normative for all males within a society (Murdock, 1967). It takes several forms and occurs in about 25% of societies (Wilson, 2008). The societies performing either form of circumcision usually refer to it as to way to make the boy "strong, fearless, valorous and respectful". Number of theories regarding the underlying

function of MGM does exist, including the improved hygiene hypothesis, modern “prophylactic” hypothesis, Oedipal interpretation of MGM – the psychodynamic hypotheses, and others (see Wilson 2008 for a critical review). These theories, even if some surely are testable, are unlikely to represent the adaptive function of MGM since hygiene risks of MGM surgery clearly outweighs its benefits, and MGM rituals obviously predates the occurrence of HIV. The ultimate function of MGM should be explained by the means of human evolutionary psychology.

Wilson (2008) tested for the hypothesis concerning function of MGM that integrates the signaling theory of ritual with principles of sexual selection. This “sexual conflict” hypothesis suggests that MGM is likely to reduce insemination efficiency, consequently reducing a man’s capacity for successful extra-pair copulations (EPC) by impairing sperm competition. MGM may therefore represent a hard-to-fake signal of a man’s reduced ability to challenge the paternity of older, already married men. MGM as a signal of sexual obedience may gain social benefits if married men are selected to offer social trust and investment preferentially to peers who are less threatening to their paternity.

Wilson (2008) found support for his theory in cross-cultural data. MGM rituals are highly public, watched by mainly male audience, MGM facilitates access to social benefits (and its absence lead to social segregation or outcasting of such males). Presence of MGM rituals correlated positively with polygyny and co-wife residence (MGM is widespread where risk of EPC is highest).

Since cases of inter-cultural transmission of practice of circumcision have been documented and cross-cultural correlation of occurrence of various forms of MGM and domestic organization (mating system) was found, it is tempting to investigate the evolution of MGM via phylogenetic comparative methods. Based on data in *Ethnographic Atlas* (Murdock, 1967; White, 1999) binary character was created that described presence of absence of MGM in the given society. It was scored as follows: **Char. 15** – Male genital mutilations: **State 0** – Absent; **State 1** – Present.

The results of MP character optimization suggest that MGM rituals represent highly conservative cultural practice that shows a strong association with phylogeny and macroscale coevolution with mating system (see Fig. 9). The results of MP and ML optimization suggest that presence of MGM is an ancestral state predominant to sub-Saharan Africa (**1**) and West Afrasia (**9**) and the absence of MGM is the ancestral predominantly to East Eurasia. According to results of MP optimization, that macro-evolution of MGM rituals occurred at

least twice independently in recent human history: in sub-Saharan Africa **(1)**, and in Polynesia **(44)**, and possibly earlier in history of Austronesian expansion, since there are multiple unique occurrences of MGM rituals in Austronesian-speaking populations. MGM rituals also occurred independently in some Australian Aboriginal populations. MP optimization suggests that in sub-Saharan Africa, MGM rituals could have evolved multiple times independently. The presence of MGM rituals might have been the ancestral state of all sub-Saharan Africans **(1)** and even of all humans **(0)**. This would either suggest the ancestral nature of human polygyny or perhaps that MGM works mainly as hard-to-fake demonstration of group membership, rather than hard-to-fake signal of sexual obedience in polygynous societies (it could certainly be both).

The results of ML optimization suggest that in ancestral Khoe-San and African Pygmies **(2)**, MGM rituals were likely to have been absent (the presence of MGM in Mbuti Pygmies can be the result of cultural transmission in this group acultured by groups practicing MGM). Even if predominant and likely ancestral in populations of sub-Saharan Africa and West Afrasia, these “African” MGM rituals may not present a single homologous phenomenon. The tentative optimization of character 16 – “Male genital mutilations (age at performing)” (see List of socio-cultural and ecological characters in Appendix) suggests that MGM rituals ancestral to and predominant in populations of sub-Saharan Africa **(1)** are those performed in adolescence or early adulthood, while MGM rituals ancestral to and predominant in populations of West Afrasia **(9)** are those performed in infancy or childhood. Moreover, various forms of MGM were documented across cultures (Wilson, 2008).

MGM rituals have likely been absent in society ancestral to European populations and this loss appears to have been linked to loss of polygyny and to ancestral monogamy in this population.

Our results show that MGM rituals are associated with polygynous societies. However, they occur not only in societies practicing harem polygyny but also in vast number of moderately polygynous societies. This might be caused by phylogenetic inertia (conservativeness) of this cultural practice that endures even if marital composition changes to moderate polygyny and to monogamy (e.g. in Mozabite of Algeria). MGM could be associated with societies at high risk for extramarital sex regardless the marital composition (social mating system practiced). This association of MGM and high frequency of EPC can be in line with the (macro)evolution of MGM in moderately polygynous populations of Polynesia **(44)**.

CONCLUSIONS

This study presents, to our knowledge, the first use of the supertree approach in cultural phylogenetics. There has only been a few studies worth mentioning in this respect that examine genetic and linguistic coevolution in populations of South America, North America, and northern Melanesia that were based on consensus of genetic phylogenies and language phylogenies (Hunley and Long, 2005; Hunley et al., 2007; Hunley et al., 2008), although those studies used consensus trees, not composite trees. The results of this study demonstrate that not only phylogenetic trees but also reticulated networks and results of STRUCTURE and FRAPPE analysis (Pritchard et al., 2000) can be utilized as inputs to obtain well-resolved and comprehensive phylogenies using standard matrix representation with parsimony (Baum, 1992; Ragan, 1992). This conclusion applies more generally, not only to human phylogeography.

The constructed supertree presents the largest, most comprehensive human population-level phylogeny available to-date. Also, it provides for the largest phylogenetic-comparative cross-cultural analysis to-date, compared to the previous studies both on global (Holden and Mace, 1997; Sellen and Mace, 1997; Holden and Mace, 1999) and the local scale (e.g. Fortunato et al., 2006; Jordan et al., 2009; Fortunato and Jordan, 2010; Walker et al., 2010). Of course, there is only partial overlap of phylogenetic and cross-cultural data available. Many cultures not crucial for understanding human population history were thoroughly documented (e.g. Native North Americans) and ethnographic data on some cultures that might be important are lacking (e.g. African click-speaking tribes, west Melanesians, west Papuans etc.). Also, some cultures that might be crucial for our understanding of human evolution and ecological adaptation are still underrepresented in recent phylogeographic studies (e.g. West Pygmies, Paleo-Asiatic populations, Andamanese, Australian Aboriginals, Tasmanians, populations of southern tip of South America and many others).

The resulting topologies of the supertrees based on linguistic and genetic evidence are in general congruent both with traditional and modern views of human phylogeography, population movements and fundamental relationships of the major world cultures (e.g. Guthrie, 1967; Murdock, 1967; Cavalli-Sforza et al., 1988; Cavalli-Sforza et al., 1994; Burton et al., 1996; Holden, 2002; Diamond and Bellwood, 2003; Li et al., 2008; Gray et al., 2009; Tishkoff et al., 2009). The most important implications of obtained supertree topology includes the basal position of Hadza and South-African Khoe-San-Pygmy clade in line with assumption of these populations early divergence (e.g. Semino et al., 2002; Tishkoff et al.,

2007, Tishkoff et al., 2009; Henn et al., 2011), the major division of the human world population into sub-Saharan Africa and the rest of the world (Eurasia, or Afrasia if also including populations of North and Northeast Africa) that is apparent in number of genome-wide SNP-based studies (e.g. Rosenberg et al., 2002; Li et al., 2008; Tishkoff et al., 2009; Xing et al., 2010), the likely sister relationship of West Eurasiatic and East Eurasiatic population, the relatively basal position of Sahul (Australia, Tasmania, New Guinea, and surrounding islands) due to early colonization (O'Connell and Allen, 2004; Krause et al., 2010; McEvoy et al., 2010), and the apparently disparate nature of today's population of Melanesia consisting of two different settlements (Diamond and Bellwood, 2003; Friedlaender, J.S., et al., 2008; Gray et al., 2009).

Our results suggest that phylogenetic comparative methods can be successfully applied to large composite phylogenies based on various sources of data and that the analysis of present-day cross-cultural variation allows us to directly address the questions regarding evolutionary processes concerning social organization, cultural practices and ecological adaptations in prehistory. Some cultural traits appear to be very conservative phylogenetically and well-comparable to biological characters in this respect.

In the present state, however, some of our results contradict those of recent cultural phylogenetic studies based on linguistic phylogenies (Fortunato et al., 2006; Jordan, 2007; Jordan et al., 2009; Fortunato and Jordan, 2010, Fortunato, 2011a). The cause of this incongruence could be threefold: the difference in topology, the difference in taxon sampling, and different character scoring.

The topology of the relevant sections of the combined MRP supertree and linguistic phylogenetic trees of languages expansions differ to various degree. Such differences (e.g. switching basal and terminal taxa within clade) can lead to different interpretation of evolutionary polarity of the concerned characters. Language phylogenies are in fact cultural phylogenies as well, and the results of optimization of cultural characters on linguistic trees demonstrate the “culture-culture” coevolution rather than declared “gene-culture” coevolution, with cases of imperfect coevolution (homoplasies) nearly impossible to interpret.

The different taxon sample, either missing taxa or additional taxa (most notably those absent in linguistic phylogenies since they speak different language like Bantu-speaking Pygmies closely related to the click-language-speaking African populations), might also affect the results significantly, especially if the linguistically distinct taxon is positioned on the base of the concerned clade.

Studies using the language phylogenies are unable to account for the effect of the cultural evolution preceding the emergence of the concerned clade (the inherited state, the true plesiomorphy) that might be vital for the accurate reconstruction of the proto-society. In other words, the studies limited to a single clade (often even linguistics-based) lack suitable outgroups and are, consequently, often unable to polarize character states evolutionarily.

Other cause of incongruence of the results lies within the different (more detailed) character scoring that was used in some characters. The most of the recent cultural phylogenetic studies (e.g. Fortunato et al. 2006; Jordan et al., 2009; Fortunato, 2011a) use, for the sake of simplification, binary characters, even if the binary nature of the character is apparently not supported by the observed cultural/behavioral variation, therefore falling for the false dichotomy (bridewealth vs. dowry, matri- vs. patrilocality, etc.). Attempts to reconstruct ancestral states of social organization and dispersal in proto-societies as strict rules could be misleading, since ancestral societies (societies in the state of expansion or foraging) might have frequently lacked these rules and possessed rather malleable cultural norms (Lane, 1961; Vayda and Rappaport, 1963; Marlowe, 2004), with individual decisions resting for example on considerations of childcare and care of elderly parents (Marlowe, 2004; 2005). Some of the discussed cultural phylogenetic studies might suffer greatly from the simplification and distortion of the observed cultural/behavioral variation. Most certainly, the criterion used to code the phenomena characterized by qualitative and quantitative variation, both within and across populations, in discrete categories is the inevitable source of potential bias in all phylogenetic cross-cultural analyses, including ours. In the reality, human behavior is almost never of “either/or” kind. There are possible implications of behavioral variation that the phylogenetic comparative methods based on distinct character states might not be able to perceive.

Reconstruction of human population history and cultural evolution presents a fascinating challenge. Our study demonstrate the potential of supertree approach for creating detailed and comprehensive human phylogenies based on various sources of data. It also demonstrates the potential of such phylogenies for making valuable inferences concerning human cultural macroevolution and adaptation on a global scale, utilizing phylogenetic comparative approach. We hope that growing body of research of human phylogeography and incorporation of additional ethnographic data will ensure the future improvement of our datasets that will provide a useful tool for testing number of explicit hypothesis concerning evolution of modern humans.

REFERENCES

- Abdulla, M.A., Ahmed, I., Assawamakin, A., et al. (92 co-authors), 2009. Mapping human genetic diversity in Asia. *Science* 326, 1541–1545.
- Arnaiz-Villena, A., Moscoso, J., Granados, J., Serrano-Vela, J.I., de L Peña, A., Reguera, R., Ferri, A., Seclen, E., Izaguirre, R., Perez-Hernandez, N., Vargas-Alarcon, G., 2007. HLA genes in Mayos population from Northeast Mexico. *Curr. Genomics* 8, 466–475.
- Arnaiz-Villena, A., Parga-Lozano, C., Moreno, E., Reces, C., Rey, D., Gomez-Prieto, P., 2010. The origin of Amerindians and the peopling of the Americas according to HLA genes: admixture with Asian and Pacific People. *Curr. Genomics* 11, 103–114.
- Atkinson, Q.D., Gray, R.D., 2005. Curious parallels and curious connections: phylogenetic thinking in biology and historical linguistics? *Syst. Biol.* 54, 513–526.
- Atkinson, Q.D., Gray, R.D., 2006. How old is the Indo-European language family? Progress or more moths to the flame? In: Forster, P., Renfrew, C. (Eds.), *Phylogenetic Methods and the Prehistory of Languages*. Cambridge: McDonald Institute for Archaeological Research, pp. 91–109.
- Ayub, Q., Mansoor, A., Ismail, M., Khaliq, S., Mohyuddin, A., Hameed, A., Mazhar, K., Rehman, S., Siddiqi, S., Papaioannou, M., Piazza, A., Cavalli-Sforza, L.L., 2003. Reconstruction of human evolutionary tree using polymorphic autosomal microsatellites. *Am. J. Phys. Anthropol.* 122, 259–268.
- Baillie, J.E.M., Hilton-Taylor, C., Stuart, S.N., 2004. 2004 IUCN Red List of Threatened Species: A Global Species Assessment. IUCN, Gland, Switzerland and Cambridge.
- Bannai, M., Ohashi, J., Harihara, S., Takahashi, Y., Juji, T., Omoto, K., Tokunaga, K., 2000. 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.
- Barrett, L., Dunbar, R.I.M., Lycett, J., 2001. *Human Evolutionary Psychology*. Palgrave Macmillan, London.
- 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, 3–10.
- Behar, D.M., Yunusbayev, B., Metspalu, M., Metspalu, E., Rosset, S., Parik, J., Rootsi, S., Chaubey, G., Kutuev, I., Yudkovsky, G., Khusnutdinova, E.K., Balanovsky, O., Semino, O., Pereira, L., Comas, D., Gurwitz, D., Bonne-Tamir, B., Parfitt, T., Hammer, M.F., Skorecki, K. and Villems, R., 2010. The genome-wide structure of the Jewish people. *Nature* 466, 238–242.
- Bininda-Emonds, O.R.P., Bryant, H.N., 1998. Properties of matrix representation with parsimony analyses. *Syst. Biol.* 47, 497–508.
- Bininda-Emonds, O.R.P., 2004. The evolution of supertrees. *Trends Ecol. Evol.* 19, 315–322.
- Bininda-Emonds, O.R.P., Cardillo, M., Jones, K.E., MacPhee, R.D., Beck, R.M., Grenyer, R., Price, S.A., Vos, R.A., Gittleman, J.L., Purvis, A., 2007. The delayed rise of present-day mammals. *Nature* 446, 507–512.
- Bininda-Emonds, O.R.P., Gittleman, J.L., Steel, M., 2002. The (super)tree of life: procedures, problems, and prospects. *Annu. Rev. Ecol. Syst.* 33, 265–289.

- Bíró, A.Z., Zalán, A., Völgyi, A., Pamjav, H., 2009. A Y-chromosomal comparison of the Madjars (Kazakhstan) and the Magyars (Hungary). *Am. J. Phys. Anthropol.* 139, 305–310.
- Blackmore, S., 1999. *The Meme Machine*. Oxford University Press.
- Borgerhoff Mulder, M., 2001. Using phylogenetically based comparative methods in anthropology: more questions than answers. *Evol. Anthropol.* 10, 99–111.
- Borgerhoff Mulder, M., George-Cramer, M., Eshleman, J., Ortolani, A., 2001. A study of East African kinship and marriage using phylogenetically-based comparative methods. *Am. Anthropol.* 103, 1059–1082.
- Borgerhoff Mulder, M., Nunn, C.L., Towner, M.C., 2006. Cultural macroevolution and the transmission of traits. *Evol. Anthropol.*, 15, 52–64.
- Brown, G.R., Dickins, T.E., Sear, R., Laland, K.N., 2011. Evolutionary accounts of human behavioural diversity. *Phil. Trans. R. Soc. B* 366, 313–324
- Bryant, D., Filimon, F., Gray, R.D., 2005. Untangling our past: languages, trees, splits and networks. In: Mace, R., Holden, C., Shennan, S.J. (Eds.), *The Evolution of Cultural Diversity: A Phylogenetic Approach*. University College London Press, London, UK, pp. 67–85.
- Bryant, D., Moulton, V., 2004. NeighborNet: an agglomerative method for the construction of planar phylogenetic networks. *Mol. Biol. Evol.* 21, 255–265.
- Bryc, K., Auton, A., Nelson, M.R., Oksenberg, J.R., Hauser, S.L., Williams, S., Froment, A., Bodo, J.M., Wambebe, C., Tishkoff, S.A., Bustamante, C.D., 2010. Genome-wide patterns of population structure and admixture in West Africans and African Americans. *Proc. Natl. Acad. Sci. USA* 107, 786–791.
- Buchanan, B., Collard, M., 2007. Investigating the peopling of North America through cladistic analyses of early paleoindian projectile points. *Anthropol. Archaeol.* 26, 59–76.
- Burton, M.L., Moore, C.C., Whiting, J.W.M., Romney, A.K., 1996. Regions based on social structure. *Curr. Anthropol.* 37, 87–123.
- Cann, H.M., de Toma, C., Cazes, L., et al. (38 co-authors), 2002. A human genome diversity cell line panel. *Science* 296, 261.
- Cavalli-Sforza, L.L., 2005. The human genome diversity project: past, present and future. *Nature Rev. Genet.* 6, 333–340.
- Cavalli-Sforza, L.L., Minch, E., Mountain, J.L., 1992. Coevolution of genes and languages revisited. *Proc. Natl. Acad. Sci. USA* 89, 5620–5624.
- Cavalli-Sforza, L.L., Feldman, M.W., 1981. *Cultural Transmission and Evolution*. Stanford, CA: Stanford University Press.
- Cavalli-Sforza, L.L., Menozzi, P., Piazza, A., 1994. *The History and Geography of Human Genes*. Princeton, Princeton University Press.
- Cavalli-Sforza, L.L., Piazza, A., Menozzi, P., and Mountain, J., 1988. Reconstruction of human evolution: bringing together genetic, archaeological, and linguistic data. *Proc. Natl. Acad. Sci. USA* 85, 6002–6006.
- Chaubey, G., Metspalu, M., Choi, Y., Magi, R., Romero, I.G., Soares, P., van Oven, M., Behar, D.M., Rootsi, S., Hudjashov, G., Mallick, C.B., Karmin, M., Nelis, M., Parik, J., Reddy, A.G., Metspalu, E., van Driem, G., Xue, Y., Tyler-Smith, C., Thangaraj, K., Singh, L., Remm, M., Richards, M.B., Lahr, M.M., Kayser, M., Villems, R., Kivisild, T., 2011.

- Population genetic structure in Indian Austroasiatic speakers: the role of landscape barriers and sex-specific admixture. *Mol. Biol. Evol.* 28, 1013–1024.
- Chu, J.Y., Huang, W., Kuang, S.Q., Wang, J.M., Xu, J.J., Chu, Z.T., Yang, Z.Q., Lin, K.Q., Li, P., Wu, M., Geng, Z.C., Tan, C.C., Du, R.F., Jin, L., 1998. Genetic relationship of populations in China. *Proc. Natl. Acad. Sci. USA* 95, 11763–11768.
- Coia, V., Caglia, A., Arredi, B., Donati, F., Santos, F.R., Pandya, A., Taglioli, L., Paoli, G., Pascali, V., Destro-Bisol, G., Tyler-Smith, C., 2004. Binary and microsatellite polymorphisms of the Y-chromosome in the Mbenzele pygmies from the Central African Republic. *Am. J. Human Biol.* 16, 57–67.
- Collard, M., Shennan, S.J., Tehrani, J.J., 2006. Branching versus Blending in Macroscale Cultural Evolution: A Comparative Study. In: Lipo, C.P., O'Brien, M.J., Shennan, S.J., Collard, M., (Eds.), *Mapping Our Ancestors: Phylogenetic Methods in Anthropology and Prehistory*, Aldine, Hawthorne, NY, pp. 53–63.
- Collard, M., Tehrani, J., 2005. Phylogenesis versus ethnogenesis in Turkmen cultural evolution. In: Mace, R., Holden, C., Shennan, S.J. (Eds.), *The Evolution of Cultural Diversity: A Phylogenetic Approach*. University College London Press, London, UK, pp. 109–131.
- Coward, F., Shennan, S.J., Colledge, S., Conolly, J., Collard, M., 2008. The spread of Neolithic plant economies from the Near East to Northwest Europe: a phylogenetic analysis. *J. Archaeol. Sci.* 35, 42–56.
- 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. *Phil. Trans. R. Soc. B* 365, 3903–3912.
- Darwin, C., 1859. *On the Origin of Species*. London, John Murray.
- Darwin, C.R., 1871. *The Descent of Man, and Selection in Relation to Sex*. 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. *Proc. R. Soc. B* 277, 1597–1606.
- Dawkins, R., 1976. *The Selfish Gene*. Oxford University Press, New York.
- Dawkins, R., 1982. *The Extended Phenotype*. Oxford University Press, New York.
- Deddiu, D., Ladd, D.R., 2007. Linguistic tone is related to the population frequency of the adaptive haplogroups of two brain size genes, ASPM and Microcephalin. *Proc. Natl. Acad. Sci. USA* 104, 10944–10949.
- de Filippo, C., Barbieri, C., Whitten M., Mpoloka, S.W., Gunnarsdóttir, E.D., Bostoen, K., Nyambe, Beyer, K., Schreiber, H., de Knijff, P., Luiselli, D., Stoneking, M., Pakendorf, B., 2011. Y-chromosomal variation in sub-Saharan Africa: insights into the history of Niger-Congo groups. *Mol. Biol. Evol.* 28, 1255–1269.
- Delmestri, A., Cristianini, N., 2010. *Linguistic Phylogenetic Inference by PAM-like Matrices*. Technical Report DISI-10-058, Ingegneria e Scienza dell'Informazione, University of Trento.
- Diamond, J., 1993. Ten thousand years of solitude: What really happens when a society is forced to go it alone? *Discover* 14, 48–57.
- Diamond, J., 2005. *Collapse: How Societies Choose to Fail or Succeed*. Penguin, New York.

- Diamond, J., Bellwood, P., 2003. Farmers and their languages: the first expansions. *Science* 300, 597–603.
- Downey, S.S., Hallmark, B., Cox, M.P., Norquest, P., Lansing, S.J., 2008. Computational feature sensitive reconstruction of language relationships: Developing the ALINE distance for comparative historical linguistic reconstruction. *J. Quant. Ling.* 15, 340–369.
- Dunn, A., Terrill, G., Reesink, R.A., Foley, Levinson, S.C., 2005. Structural phylogenetics and the reconstruction of ancient language history, *Science* 309, 2072–2075.
- Eff, E.A., 2004. Does Mr. Galton still have a problem? Autocorrelation in the standard cross-cultural sample. *World Cultures* 15, 153–170.
- Eshleman, J.A., Malhi, R.S., Johnson, J.R., Kastle, F.A., Lorenz, J., Smith, D.G., 2004. Mitochondrial DNA and prehistoric settlements: native migrations on the western edge of North America. *Hum. Biol.* 76, 55–75.
- Farjadian, S., Ota, M., Inoko, H., Ghaderi, A., 2009. The genetic relationship among Iranian ethnic groups: an anthropological view based on HLA class II gene polymorphism. *Mol. Biol. Rep.* 36, 1943–1950.
- Ferri, A., Seclen, E., Izaguirre, R., Perez-Hernandez, N., Vargas-Alarcon, G., 2007. HLA genes in Mayos population from Northeast Mexico. *Curr. Genomics* 8, 466–475.
- Fortunato, L., 2011a. Reconstructing the history of marriage strategies in Indo-European—speaking societies: monogamy and polygyny. *Hum. Biol.* 83, 87–105.
- Fortunato, L., 2011b. Reconstructing the history of residence strategies in Indo-European—speaking societies: Neo-, uxori-, and virilocality. *Hum. Biol.* 83, 107–128.
- Fortunato, L., Holden, C. & Mace, R., 2006 From bridewealth to dowry? A Bayesian estimation of ancestral states of marriage transfers in Indo-European groups. *Hum. Nat.* 17, 355–376.
- Fortunato, L., Jordan, F., 2010. Your place or mine? A phylogenetic comparative analysis of marital residence in Indo-European and Austronesian societies. *Phil. Trans. R. Soc. B* 365, 3913–3922.
- Friedlaender, J.S., Friedlaender, F.R., Reed, F.A., Kidd, K.K., Kidd, J.R., Chambers, G.K., Lea, R.A., Loo, J.H., Koki, G., Hodgson, J.A., Merriwether, D.A., Weber, J.L., 2008. The genetic structure of Pacific Islanders. *PLoS Genet.* 4, e19.
- Garcia-Ortiz, J.E., Sandoval-Ramirez, L., Rangel-Villalobos, H., Maldonado-Torres, H., Cox, S., Garcia-Sepulveda, C.A., Figuera, L.E., Marsh, S.G., Little, A.M., Madrigal, J.A., Moscoso, J., Arnaiz-Villena, A., Arguello, J.R., 2006. High-resolution molecular characterization of the HLA class I and class II in the Tarahumara Amerindian population. *Tissue Antigens* 68, 135–146.
- Geisler, H., List, J., 2009. Beautiful trees on unstable ground. Ulm University, Ulm. (http://www.uniulm.de/fileadmin/website_uni_ulm/med.evo/List_Geisler_trees.pdf)
- Goloboff, P. 1999. NONA (NO NAME) ver. 2.0. Published by the author, Tucumán, Argentina.
- Goody, J., 1973. Bridewealth and Dowry in Africa and Eurasia. In: Goody, J., Tambiah, S.J., (Eds.), *Bridewealth and Dowry*. Cambridge University Press, Cambridge, pp. 1–58.

- 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. *Mol. Biol. Evol.* 24, 757–768.
- Gray, J.P., 1999. A Corrected Ethnographic Atlas. *World Cultures* 10, 24–136.
- Gray, R. D., Greenhill, S. J. & Ross, R. M. 2007. The pleasures and perils of Darwinizing culture (with phylogenies). *Biol. Theor.* 2, 360–375.
- Gray, R.D, Atkinson, Q.D., 2003. Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature* 426, 435–439.
- Gray, R.D., Bryant, D., Greenhill, S.J., 2010. On the shape and fabric of human history. *Phil. Trans. R. Soc. B* 365, 3923–3933.
- Gray, R.D., Drummond, A.J., Greenhill, S.J., 2009. Language phylogenies reveal expansion pulses and pauses in Pacific settlement. *Science* 323, 479483.
- Gray, R.D., Jordan, F.M., 2000. Language trees support the express-train sequence of Austronesian expansion. *Nature* 405, 1052–1055.
- Greenhill, S.J., Atkinson, Q.D., Meade, A., Gray, R.D., 2010. The shape and tempo of language evolution. *Proc. R. Soc. B* 277, 2443–2450.
- Greenhill, S.J., Blust, R., Gray, R.D., 2008. The austronesian basic vocabulary database: From bioinformatics to lexomics. *Evol. Bioinform.* 4, 271–283. Austronesian Basic Vocabulary database available online at <http://language.psy.auckland.ac.nz/austronesian/>
- Greenhill, S.J., Currie, T.E., Gray, R.D., 2009. Does horizontal transmission invalidate cultural phylogenies? *Proc. R. Soc. B* 276, 2299–2306.
- Greenhill, S.J., Gray, R.D., 2005. Testing population dispersal hypotheses: pacific settlement, phylogenetic trees and austronesian languages. In: Mace, R., Holden, C., Shennan, S.J. (Eds.), *The Evolution of Cultural Diversity: A Phylogenetic Approach*. London, UK: University College London Press, pp. 31–52.
- Guglielmino, C.R., Viganotti, C., Hewlett, B., Cavalli-Sforza, L.L., 1995. Cultural variation in Africa: role of mechanisms of transmission and adaptation. *Proc. Natl. Acad. Sci. USA* 92, 7585–7589.
- Guthrie, M. 1967. *Comparative Bantu*. Farnborough: Gregg International Publishers Ltd. Vols. 1–4. List of Bantu Language Names (in order by Guthrie number) available online at <http://linguistics.berkeley.edu/CBOLD/Lgs/LgsbyGN.html>
- Haspelmath, M., Dryer, M., Gil, D., Comrie, B. (Eds.), 2005. *The World Atlas of Language Structures*. Oxford, England, Oxford University Press. Available online at <http://wals.info/index>
- Henn, B.M., Gignoux, C.R., Jobin, M., Granka, J.M., Macpherson, M.M., Kidd, J.M., Rodríguez-Botigué, L., Ramachandran, S., Hon, L., Brisbin, A., Lin, A.A., Underhill, P.A., Comas, D., Kidd, K.K., Norman, P.J., Parham, P., Bustamante, C.D., Mountain, J.L., Feldman, M.W., 2011. Hunter-gatherer genomic diversity suggests a southern African origin for modern humans. *PNAS*, Published online before print March 7, 2011.
- Hewlett, B.S., de Silvestri, A., Guglielmino, C.R., 2002. Semes and Genes in Africa. *Curr. Anthropol.* 43, 313–321.
- Holden, C. J., 2002. Bantu language trees reflect the spread of farming across sub-Saharan Africa: a maximum-parsimony analysis. *Proc. R. Soc. Lond. B* 269, 793–799.

- Holden, C., Mace, R., 1997. Phylogenetic analysis of the evolution of lactose digestion in adults. *Hum. Biol.* 69, 605–628.
- Holden, C., Mace, R., 1999. Sexual dimorphism in stature and women's work: a phylogenetic cross-cultural analysis. *Am. J. Phys. Anthropol.* 110, 27–45.
- Holden, C., Shennan, S.J., 2005. Introduction to part 1: how treelike is cultural evolution?
In: Mace, R., Holden, C., Shennan, S.J. (Eds.), *The Evolution of Cultural Diversity: A Phylogenetic Approach*. London, UK: University College London Press, pp.13–30.
- Holden, C.J., Gray, R.D., 2006. Rapid radiation, borrowing and dialect continua in the Bantu languages. In: Foerster, P., Renfrew, C. (Eds.), *Phylogenetic Methods and the Prehistory of Languages*. McDonald Institute for Archaeological Research, Cambridge, pp. 19–31.
- Holden, C.J., Mace, R., 2003. Spread of cattle led to the loss of matrilineal descent in Africa: a coevolutionary analysis. *Proc. R. Soc. Lond. B* 270, 2425–2433.
- Horai, S., Murayama, K., Hayasaka, K., Matsubayashi, S., Hattori, Y., Fucharoen, G., Harihara, S., Park, K.S., Omoto, K., Pan, I.H., 1996. mtDNA polymorphism in east Asian populations, with special reference to the peopling of Japan. *Am. J. Hum. Genet.* 59, 579–590.
- Hunley, K., Cabana, G.S., Merriwether, D.A., Long, J.C., 2007. A formal test of linguistic and genetic coevolution in native Central and South America. *Am. J. Phys. Anthropol.* 132, 622–631.
- Hunley, K., Dunn, M., Lindstrom, E., Reesink, G., Terrill, A., Healy, M.E., Koki, G., Friedlaender, F.R., Friedlaender, J.S., 2008. Genetic and linguistic coevolution in northern island Melanesia. *PLoS Genet.* 4, e1000239.
- Hunley, K., Long, J.C., 2005. Gene flow across linguistic boundaries in Native North American populations. *Proc. Natl. Acad. Sci. USA* 102, 1312–1317.
- Huson, D.H., Bryant, D., 2006. Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* 23, 254–267.
- Huxley, J.S., 1942. *Evolution: The Modern Synthesis*. Allen & Unwin.
- Huyghe, J.R., Fransen, E., Hannula, S., Van Laer, L., Van Eyken, E., Mäki-Torkko, E., Aikio, P., Sorri, M., Huentelman, M.J., Camp, G.V., 2011. A genome-wide analysis of population structure in the Finnish Saami with implications for genetic association studies. *Eur. J. Hum. Genet.* 19, 347–352.
- Ingman, M., Gyllensten, U., 2003. Mitochondrial genome variation and evolutionary history of Australian and New Guinean Aborigines. *Genome Res.* 13, 1600–1606.
- Ingman, M., Kaessmann, H, Paabo, S., Gyllensten, U., 2000. Mitochondrial genome variation and the origin of modern humans. *Nature* 408, 708–713.
- Jin, F., Saitou, N., Ishida, T., Sun, C.S., Pan, I.H., Omoto, K., Horai, S., 1999. Population genetic studies on nine aboriginal ethnic groups of Taiwan. I. Red cell enzyme systems. *Anthropol Sci* 107, 229–246.
- Jobling, M.A., Tyler-Smith, C., 2003. The human Y chromosome: an evolutionary marker comes of age. *Nat. Rev. Genet.* 4, 598–612.
- Johnson, A., Earle, T., 2000. *The Evolution of Human Societies: From Foraging Group to Agrarian State*. Stanford University Press, Stanford, CA.

- Jordan, F., Mace, R., 2007. Changes in post-marital residence precede changes in descent systems in Austronesian societies. The European Human Behaviour and Evolution Conference 2007, London School of Economics, London.
- Jordan, F.M., Gray, R.D., Greenhill, S.J., Mace, R., 2009. Matrilocal residence is ancestral in Austronesian societies. *Proc. R. Soc. B* 276, 1957–1964.
- Jordan, P., O’Neill, S., 2010. Untangling cultural inheritance: language diversity and long-house architecture on the Pacific northwest coast. *Phil. Trans. R. Soc. B* 365, 3875–3888.
- Kang, L.L., Li, S.L., Gupta, S., Zhang, Y.G., Liu, K., Zhao, J.M., Jin, L., Li, H., 2010. Genetic structures of the Tibetans and the Deng people in the Himalayas viewed from autosomal STRs. *J. Hum. Genet.* 55, 270–277.
- 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. *Proc. R. Soc. B* 276, 2703–2710.
- Kohlrausch, F.B., Callegari-Jacques, S.M., Tsuneto, L.T., Petzl-Erler, M.L., Hill, K., Hurtado, A.M., Salzano, F.M., Hutz, M.H., 2005. Geography influences microsatellite polymorphism diversity in Amerindians. *Am. J. Phys. Anthropol.* 126, 463–470.
- Kong, Q.P., Sun, C., Wang, H.W., Zhao, M., Wang, W.Z., Zhong, L., Hao, X.D., Pan, H., Wang, S.Y., Cheng, Y.T., Zhu, C.L., Wu, S.F., Liu, L.N., Jin, J.Q., Yao, Y.G., Zhang, Y.P., 2011. Large-scale mtDNA screening reveals a surprising matrilineal complexity in East Asia and its implications to the peopling of the region. *Mol. Biol. Evol.* 28: 513–522.
- Kopelman, N.M., Stone, L., Wang, C., Gefel, D., Feldman, M.W., Hillel, J., Rosenberg, N.A., 2009. Genomic microsatellites identify shared Jewish ancestry intermediate between Middle Eastern and European populations. *BMC Genet.* 10, 80.
- Krause, J., Fu, Q., Good, J.M., Viola, B., Shunkov, M.V., Derevianko, A.P., Pääbo, S., 2010. The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. *Nature* 464, 894–897.
- Lane, R. B., 1961. A reconsideration of Malayo-Polynesian social organization. *Am. Anthropol.* 63, 711–720.
- Lewis, M. P., (Ed.), 2009. *Ethnologue: Languages of the World*, 16th ed. Dallas, Texas: SIL. Available online at <http://www.ethnologue.com/>.
- Laland, K.N., Galef, B. (Eds.), 2009. *The Question of Animal Culture*. Harvard University Press, Cambridge, MA.
- Lewis, P.O., 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. *Syst. Biol.* 50, 913–925.
- Li, H., Cho, K., Kidd, J.R., Kidd, K.K., 2009. Genetic landscape of Eurasia and “admixture” in Uyghurs. *Am. J. Hum. Genet.* 85: 934–937.
- Li, J.Z., Absher, D.M., Tang, H., Southwick, A.M., Casto, A.M., Ramachandran, S., Cann, H.M., Barsh, G.S., Feldman, M., Cavalli-Sforza, L.L., 2008. Worldwide human relationships inferred from genome-wide patterns of variation. *Science* 319, 1100–1104.
- Lipo, C.P., O’Brien, M.J., Collard, M., Shennan, S., (Eds.) 2006. *Mapping Our Ancestors: Phylogenetic Approaches in Anthropology and Prehistory*. Aldine, New York, NY.
- Listman, J.B., Malison, R.T., Sanichwankul, K., Ittiwut, C., Mutirangura, A., Gelernter, J., 2011. Southeast Asian origins of five Hill Tribe populations and correlation of genetic to

- linguistic relationships inferred with genome-wide SNP data. *Am. J. Phys. Anthropol.* 144, 300–308.
- Lycett, S.J., 2007. Why is there a lack of Mode 3 Levallois technologies in East Asia? A phylogenetic test of the Movius-Schick hypothesis. *J. Anthropol. Archaeol.* 26, 541–575.
- Lycett, S.J., Collard, M., McGrew, W.C., 2009. Cladistic analyses of behavioral variation in wild Pan troglodytes: exploring the chimpanzee culture hypothesis. *J. Hum. Evol.* 57, 337–349.
- Mace, R., 2005. On the use of phylogenetic comparative methods to test co-evolutionary hypotheses across cultures. In: Mace, R., Holden, C., Shennan, S.J. (Eds.), *The Evolution of Cultural Diversity: A Phylogenetic Approach*. London, UK: University College London Press, pp. 235–256.
- Mace, R., Holden, C.J., Shennan, S., (Eds.) 2005. *The Evolution of Cultural Diversity: A Phylogenetic Approach*. University College London Press, London, UK.
- Mace, R., Pagel, M., 1994. The comparative method in anthropology. *Curr. Anthropol.* 35, 549–564.
- Mace, R., Jordan, F.M., 2011. Macro-evolutionary studies of cultural diversity: a review of adaptation empirical studies of cultural transmission and cultural adaptation. *Phil. Trans. R. Soc. B* 366, 402–411.
- Maddison, W. P., Maddison, D.R., 2009. Mesquite: a modular system for evolutionary analysis. Version 2.71. Available online at <http://mesquiteproject.org>.
- Marlowe, F.W., 2005. Hunter-gatherers and human evolution. *Evol. Anthropol.* 14, 54–67.
- Marlowe, F.W. 2004. Marital residence among foragers. *Curr. Anthropol.* 45, 277–284.
- McEvoy, B.P., Lind, J.M., Wang, E.T., Moyzis, R.K., Visscher, P.M., Pellekaan, S.M.V., Wilton, A.N., 2010. Supplemental Data Whole-Genome Genetic Diversity in a Sample of Australians with Deep Aboriginal Ancestry. *Am. J. Hum. Genet.* 87, 297–305.
- Mesoudi, A., Whiten, A., Laland, K.N., 2004. Is human cultural evolution Darwinian? Evidence reviewed from the perspective of *The Origin of Species*. *Evolution* 58, 1–11.
- Mesoudi, A., Whiten, A., Laland, K., 2006. Toward a unified science of cultural evolution. *Behav. Brain Sci.* 29, 329–383.
- Murdock, G.P., 1967. *Ethnographic Atlas: A summary*. *Ethnology* 6, 109–236.
- Murdock, G.P., White, D.R., 1969. Standard cross-cultural sample. *Ethnology* 8:329–69.
- Nettle, D., 2007. Language and genes: A new perspective on the origins of human cultural diversity. *Proc. Nat. Acad. Sci.* 104, 10755–10756.
- Nixon, K.C., 1999. WinClada ver. 1.0000. Published by the author, Ithaca, NY, USA.
- Novembre, J., Johnson, T., Bryc, K., Boyko, A., Auton, A., Indap, A., King, K., Bergmann, S., Nelson, M., Stephens, M., Bustamante, C.D., 2008. Genes mirror geography within Europe. *Nature* 456, 98–101.
- Nunn, C.L., Arnold, C., Matthews, L., Borgerhoff Mulder, M., 2010. Simulating trait evolution for cross-cultural comparison. *Phil. Trans. R. Soc. B* 365, 3807–3819.
- O’Connell, J.F., Allen, J., 2004. Dating the colonization of Sahul (Pleistocene Australia–New Guinea): a review of recent research. *J. Archaeol. Sci.* 31, 835–853.

- Omoto, K., Saitou, N., 1997. Genetic origins of the Japanese: a partial support for the dual structure hypothesis. *Am. J. Phys. Anthropol.* 102, 437–446.
- Oven, M.V., Kayser, M., 2009. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum. Mut.* 30, E386–E394.
- Pagel, M., Atkinson, Q.D., Meade, A., 2007. Frequency of word-use predicts rates of lexical evolution throughout Indo-European history. *Nature* 449, 717–720.
- Pagel, M., 2009. Human language as a culturally transmitted replicator. *Nat. Rev. Genet.* 10, 405–415.
- Pagel, M., Meade, A., 2005. Bayesian estimation of correlated evolution across cultures: a case study of marriage systems and wealth transfer at marriage. In: Mace, R., Holden, C., Shennan, S.J. (Eds.), *The Evolution of Cultural Diversity: A Phylogenetic Approach*. London, UK: University College London Press, pp. 235–256.
- Patin, E., Laval, G., Barreiro, L.B., Salas, A., Semino, O., Santachiara-Benerecetti, S., Kidd, K.K., Kidd, J.R., Van der Veen, L., Hombert, J.M., Gessain, A., Froment, A., Bahuchet, S., Heyer, E., Quintana-Murci, L., 2009. Inferring the demographic history of African farmers and Pygmy hunter-gatherers using a multilocus resequencing data set. *PLoS Genet.* 5, e1000448.
- Pisani D, Yates AM, Langer MC, Benton MJ. 2002. A genus-level supertree of the Dinosauria. *Proc. R. Soc. B.* 269, 915–21.
- Pritchard, J.K., Stephens, M., Donnelly, P.J., 2000. Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959.
- Ragan, M.A., 1992. Phylogenetic inference based on matrix representation of trees. *Mol. Phylogenet. Evol.*, 1, 53–58.
- Rama, T., Singh, A.K., 2009. From bag of languages to family trees from noisy corpus. In: *Proceedings of the Conference on Recent Advances in Natural Language Processing*, Borovets, Bulgaria, pp. 355–359.
- Rasmussen, M., Li, Y., Lindgreen, S., Pedersen, J.S., Albrechtsen, A., Moltke, I., Metspalu, M., Metspalu, E., Kivisild, T., Gupta, R. et al. (52 co-authors), 2010. Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature*, 463, 757–762.
- Reesink, G., Singer, R., Dunn, M., 2009. Explaining the linguistic diversity of Sahul using population models. *PLoS Biol.* 7, e1000241.
- Rexová, K., Frynta, D., Zrzavý, D., 2003. Cladistic analysis of languages: Indo-European classification based on lexicostatistical data. *Cladistics* 19, 120–127.
- Rexová, K., Bastin, Y., Frynta, D., 2006. Cladistic analysis of Bantu languages: a new tree based on combined lexical and grammatical data, *Naturwissenschaften* 93, 189–194.
- Richerson, P.J., Boyd, R., 2005. *Not by Genes Alone: How Culture Transformed Human Evolution*. University of Chicago Press, Chicago.
- 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, 2381–2385.
- Ruhlen, M., 1991. *A Guide to the World's Languages*. Edward Arnold, London.
- Salmela, E., Lappalainen, T., Liu, J., Sistonen, P., Andersen, P.M., Schreiber, S., Savontaus, M.L., Vžene, K., Lahermo, P., Hall, P., Kere, J., 2011. Swedish population substructure revealed by genome-wide single nucleotide polymorphism data. *PLoS ONE* 6, e16747.

- Sellen, D.W., Mace, R., 1997. Fertility and mode of subsistence: a phylogenetic analysis. *Curr. Anthropol.* 38, 878–889.
- Semino, O., Santachiara-Benerecetti, A.S., Falaschi, F., Cavalli-Sforza, L.L., Underhill, P.A., 2002. Ethiopians and Khoisan share the deepest clades of the human Y-chromosome phylogeny. *Am. J. Hum. Genet.* 70, 265–268.
- Serva, M., 2009. Automated languages phylogeny from Levenshtein distance. Conference proceedings, Instituto de Estudos Avancados Transdisciplinares - UFMG (Belo Horizonte).
- Serva, M., Petroni, F., 2008. Indo-European languages tree by Levenshtein distance. *EuroPhys. Lett.* 81, 68005.
- Shennan, S., 2008. Evolution in archaeology. *Annu. Rev. Anthropol.* 37, 75–91.
- Sikora, M., Laayouni, H., Calafell, F., Comas, D., Bertranpetit, J., 2011. A genomic analysis identifies a novel component in the genetic structure of sub-Saharan African populations. *Eur. J. Hum. Genet.* 19, 84–88.
- Steele, J., Kandler, A., 2010. Language trees \neq gene trees. *Theor. Biosci.* 129, 223–233.
- Sulcebe, G., Sanchez-Mazas, A., Tiercy, J.M., Shyti, E., Mone, I., Ylli, Z., Kardhashi, V., 2009. HLA allele and haplotype frequencies in the Albanian population and their relationship with the other European populations. *Int. J. Immunogenet.* 36, 337–343.
- Tajima, A., Hayami, M., Tokunaga, K., Juji, T., Matsuo, M., Marzuki, S., Omoto, K., Horai, S., 2004. Genetic origins of the Ainu inferred from combined DNA analyses of maternal and paternal lineages. *J. Hum. Genet.* 49, 187–193.
- Tajima, A., Sun, C.S., Pan, H.I., Ishida, T., Saitou, N., Horai, S., 2003. Mitochondrial DNA polymorphisms in nine aboriginal groups of Taiwan: implications for the population history of aboriginal Taiwanese. *Hum. Genet.* 113, 24–33.
- Tamm, E., Kivisild, T., Reidla, M., Metspalu, M., Smith, D.G., Mulligan, C.J., Bravi, C.M., Rickards, O., Martinez-Labarga, C., Khusnutdinova, E.K., Fedorova, S.A., Golubenko, M.V., Stepanov, V.A., Gubina, M.A., Zhadanov, S.I., Ossipova, L.P., Damba, L., Voevoda, M.I., Dipierri, J.E., Villems, R., Malhi, R.S., 2007. Beringian standstill and spread of native American founders. *PLoS ONE* 2, e829.
- The International HapMap Consortium, 2003. The international HapMap project. *Nature* 426, 789–795.
- Thompson, L.C., Kincade, M.D., 1990. Languages. In: Suttles (Ed.), *Handbook of North American Indians*, vol. 7. Northwest Coast. Washington, DC: Smithsonian Institution. pp. 30–51.
- Tishkoff, S.A., Gonder, M.K., Henn, B.M., Mortensen, H., Knight, A., Gignoux, C., Fernandez-Pulle, N., Lema, G., Nyambo, T.B., Ramakrishnan, U., Reed, F.A., Mountain, J.L., 2007. History of click-speaking populations of Africa inferred from mtDNA and Y chromosome genetic variation. *Mol. Biol. Evol.* 24, 2180–2195.
- Tishkoff, S.A., Reed, F.A., Friedlaender, F.R., et al. (25 co-authors), 2009. The genetic structure and history of Africans and African Americans. *Science* 324, 1035–1044.
- Tower of Babel: Evolution of Human Language Project. Available online at <http://starling.rinet.ru/index2.php?lan=en>
- Trivers, R.L., 1972. Parental investment and sexual selection. In: Campbell, B., (Ed.) *Sexual Selection and the Descent of Man, 1871–1971*. Heinemann, London, pp. 136–179.

- Tsai, L., 2004. A comparative study of Sakizaya and Amis in Hualien by mitochondrial DNA sequences analysis. Master Thesis, Graduate Institute of Anthropology, Tzu Chi University, Hualien.
- Tsuneto, L., Probst, C., Hutz, M., Salzano, F., Rodriguez-Delfin, L., Zago, M., Hill, K., Hurtado, A., Ribeiro-dos-Santos, A., Petzl-Erler, M., 2003. HLA class II diversity in seven Amerindian populations: clues about the origins of the Aché. *Tissue Antigens* 62, 512–526.
- Tylor, E.B., 1871. *Primitive Culture: Researches into the Development of Mythology, Philosophy, Religion, Art, and Custom*. John Murray, London, UK.
- Vayda, A.P., Rappaport, R.A., 1963. Island cultures. In: Fosberg, F.R., (Ed.) *Man's Place in the Island Ecosystem*. Bishop Museum Press, Honolulu, HI, pp. 133–142.
- Verdu, P., Austerlitz, F., Estoup, A., Vitalis, R., Georges, M., Théry, S., Froment, A., Le Bomin, S., Gessain, A., Hombert, J.M., Van der Veen, L., Quintana-Murci, L., Bahuchet, S., Heyer, E., et al., 2009. Origins and genetic diversity of Pygmy hunter-gatherers from western Central Africa. *Curr. Biol.* 19, 312–318.
- Walker, R.S., Hamilton, M.J., 2010. Social complexity and linguistic diversity in the Austronesian and Bantu population expansions. *Proc. R. Soc. B*, Published online before print October 20, 2010.
- Walker, R.S., Flinn, M.V., Hill, K.R., 2010. Evolutionary history of partible paternity in lowland South America. *Proc. Natl. Acad. Sci. USA* 107, 19195–19200.
- Walker, R.S., Ribeiro, L.A., 2011. Bayesian phylogeography of the Arawak expansion in lowland South America. *Proc. R. Soc. B*, Published online before print January 19, 2011.
- Wang, B., Zhang, Y.B., Zhang, F., Lin, H., Wang, X., Wan, N., Ye, Z., Weng, H., Zhang, L., Li, X., Yan, J., Wang, P., Wu, T., Cheby, L., Wang, J., Wang, D.M., Ma, X., Yu, J., 2011. On the origin of Tibetans and their genetic basis in adapting high-altitude environments. *PLoS ONE* 6, e17002.
- Wang, S., Lewis, C.M. Jr., Jakobsson, M., et al. (27 co-authors), 2007. Genetic variation and population structure in Native Americans. *PLoS Genet.* 3, e185.
- Wells, R.S., Yuldasheva, N., Ruzibakiev, R., et al. (27 co-authors), 2001. The Eurasian heartland: a continental perspective on Y-chromosome diversity. *Proc. Natl. Acad. Sci. USA* 98, 10244–10249.
- White, D.R., 1986. Ethnographic bibliography for the Standard Cross-Cultural Sample. *World Cultures* 2, 16–125.
- Whiten, A., Goodall, J., McGrew, W.C., Nishida, T., Reynolds, V., Sugiyama, Y., Tutin, C.E.G., Wrangham, R.W., Boesch, C., 1999. Cultures in chimpanzees. *Nature* 399, 682–85.
- Whitfeld, J., 2008. Across the curious parallel of language and species evolution. *PLOS Biol.* 6, e186.
- Wilson, C.G., 2008. Male genital mutilation: an adaptation to sexual conflict *Evol. Hum. Behav.* 29, 149–164.
- Xing, J., Watkins, W.S., Witherspoon, D.J., Zhang, Y., Guthery, S.L., Thara, R., Mowry, B.J., Bulayeva, K., Weiss, R.B., Jorde, L.B., 2009. Fine-scaled human genetic structure revealed by SNP microarrays, *Genome Res.* 19, 815–825.
- Xing, J., Watkins, W.S., Shlien, A., Walker, E., Huff, C.D., Witherspoon, D.J., Zhang, Y., Simonson, T.S., Weiss, R.B., Schiffman, J.D., Malkin, D., Woodward, S.R., Jorde,

L.B., 2010. Toward a more uniform sampling of human genetic diversity: A survey of worldwide populations by high-density genotyping. *Genomics* 96, 199–210.

Xu, S., Kangwanpong, D., Seielstad, M., Srikummool, M., Kampuansai, J., Jin, L., 2010. Genetic evidence supports linguistic affinity of Mlabri - a hunter-gatherer group in Thailand. *BMC Genet.* 11, 18.

Zhu, B., Yang, G., Shen, C., Qin, H., Liu, S., Deng, Y., Fan, S., Deng, L., Chen, F., Zhang, P., Fang, J., Chen, L., Wang, H., Wang, Z., Lucas, R., 2010. Distributions of HLA-A and -B alleles and haplotypes in the Yi ethnic minority of Yunnan, China: relationship to other populations. *J. Zhejiang Univ. Sci. B* 11, 127–135.

Table and figure legends

Table 2. List of 44 selected taxa present/absent in topologies of six MRP supertrees (combined and genomic datasets, using three rooting options for each). “M” denotes a perfect monophyly, “M+” an imperfect monophyly (with some internal taxa falling outside the monophyletic clade), “P” a compact paraphyly, “P+” a disrupted paraphyly (with some external taxa disrupting the compact paraphyletic cluster), “A” the absence of taxon due to polyphyly, “X” the absence of the taxon due to the absence of elementary taxa that constitutes it, “U” unresolved (a polytomy with potential hidden monophyly, either perfect or imperfect). Two abbreviations separated with slash denote the combination of two of the above (e.g. “P / M+” denotes paraphyletic cluster with some of the internal taxa forming a monophyletic clade within the paraphyletic cluster).

Table 3. Reconstructions of ancestral states for 45 higher taxa (ancestral nodes) and 11 selected characters obtained by MP optimizations of character data onto three combined MRP supertrees based on various rooting (“Chimp” / “All-0” / “Unrooted”), and onto majority-rule consensus of three genomic supertrees. Consensus reconstructions of ancestral states based on four output trees are given. “?” denotes ambiguous reconstruction of ancestral state, “X” the absence of taxon in a given topology.

Fig. 2. Majority rule consensus of MRP supertrees resulting from combined dataset using various rooting options.

Fig. 3. Majority rule consensus of MRP supertrees based on genomic datasets.

Fig. 4. MRP supertree of Indo-European language trees based on lexical data.

Fig. 5. Result of ML optimization of character “Wealth transfers at marriage sensu Fortunato et al., 2006” on topology of MRP supertree of Indo-European language trees.

Fig. 6. Optimization of two characters on the combined supertree. Left, char. 7b “General mating system”; right, char. 1 “Wealth transfers at marriage sensu Fortunato et al., 2006”.

Fig. 7. Optimization of two characters on the combined supertree. Left, char. 2 “Postmarital residence sensu Jordan et al., 2009”; right, char. 12 “Descent: major type”.

Fig. 8. Optimization of char. 3 “Socio-political complexity sensu Walker and Hamilton, 2010” on the combined supertree.

Fig. 9. Optimization of two characters on the combined supertree. Left, char. 7b “General mating system”; right, char. 15 “Male genital mutilations”.

Table 2

	Combined Supertree			Genomic Supertree		
	Chimp	All-0	Unrooted	Chimp	All-0	Unrooted
1 sub-Saharan Africa	M	M	M	M+	U (P)	U (P)
2 Khoe-San + African Pygmy	M+	U (M)	M	M	P	M
3 Khoe-San	M+	M	M	M	M	M
4 African Pygmy	M+	M+	M+	M+	M	M+
5 Bantu incl. Bantoid	P+	P+	P+	A	A	A
6 East Africa (East Bantu)	A	P	P	A	A	X
7 Southeast Africa (South Bantu)	M	M	M	A	A	A
8 East Africa (Nilo-Saharan languages)	P+ / M+	P+ / M+	P+ / M+	P+ / M+	U (M)	U (M)
9 West Afrasia (incl. Afro-Asiatic languages)	M	P	P	P	P	P
10 North Africa (Afro-Asiatic languages)	M	P / M+	P / M+	P	P	P
11 West Eurasia (excl. Afro-Asiatic languages)	M	P	P	P	P	P
12 Europe	M+	M+	M	M+	M+	M+
13 East Europe (Balto-Slavic, and Uralic languages)	M	M	M	M	U (M)	M
14 Northwest Europe (Celtic languages)	M	U (M+)	U (M+)	X	X	X
15 Southwest Europe (Italic languages and Basque)	M	M	M	M	P	A
16 Europe (Germanic and Creole languages)	M+	M+	M+	X	X	X
17 South Asia	M	P+	P	P	P+	P+
18 Southwest Asia (Indo-Iranian languages)	P / M+	A / M+	A / M+	P+	P+	P+
19 India (Indo-European + Dravidian languages)	M	M+	M+	P+ / M+	P+ / M+	P+ / M+
20 Ancestral East Eurasia (incl. Sahul, Oceania, and America)	M	M+	M	M	M+	M+
21 Circum-Pacific (incl. Far East and America)	P	M+	P	P+	P	P+
22 America (incl. Beringia)	M	M	M	M+	M	M+
23 Beringia and North Nearctic	M	U (M)	U (M)	A	P	A
24 Inuit	M+	U (M)	M	X	X	X
25 North Na-Dene	M	U (M)	U (M)	P	U (M+)	M
26 Amerindian	P	U (M)	U (M)	M	M	M
27 Far East (incl. Indochina)	M	M+	M+	P+	P+	P+
28 Indochina	M+	M+	M+	U (M)	M+	M+
29 Far East (excl. Indochina)	M	M+	M	P+	P+	P+
30 Australasia (incl. Sahul)	M	M	M	A	M	M
31 Sahul (incl. West Melanesia, Papua, Australia and Tasmania, and Andaman Islands)	M	M	M	M	M	M
32 West Melanesia (Papuan languages)	P+ / M+	P+ / M+	P+ / M+	P+	P+	P+
33 Papua + Australia	M	M	M	P+	P+	P+
34 Papua	M	P	P	P	P	P
35 Australia (Incl. Tasmania) (monophyletic)	M	M+	M	M	M	M
36 Australasia (excl. Sahul)	A	A	A	M	A	A
37 Philippines (incl. Formosa)	M	U (M)	A	X	X	X
38 Formosa	M	M	M	X	X	X
39 Philippines (excl. Formosa)	A	M	M	U (M)	M+	M+
40 Borneo + Madagascar	M	M	U	X	X	X
41 Oceania	U (M)	U (M)	M+	U (M)	P+	P+
42 Remote Oceania	U (M)	M	M	U (M)	P+	P+
43 Micronesia	U (M)	M	M	U (M)	U (M)	P+
44 Polynesia	U (M)	M	M	U (M)	P+	U (M)

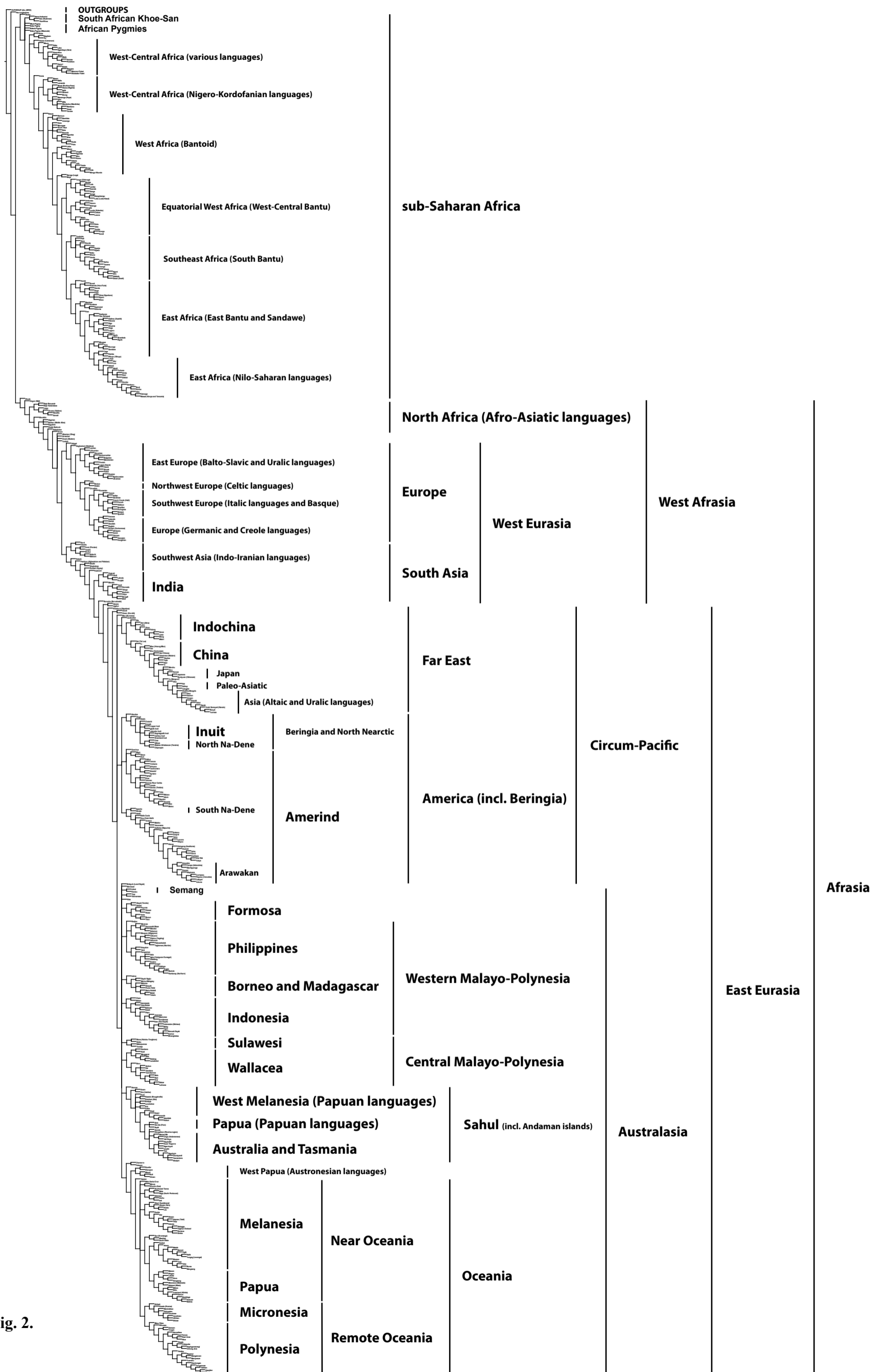


Fig. 2.

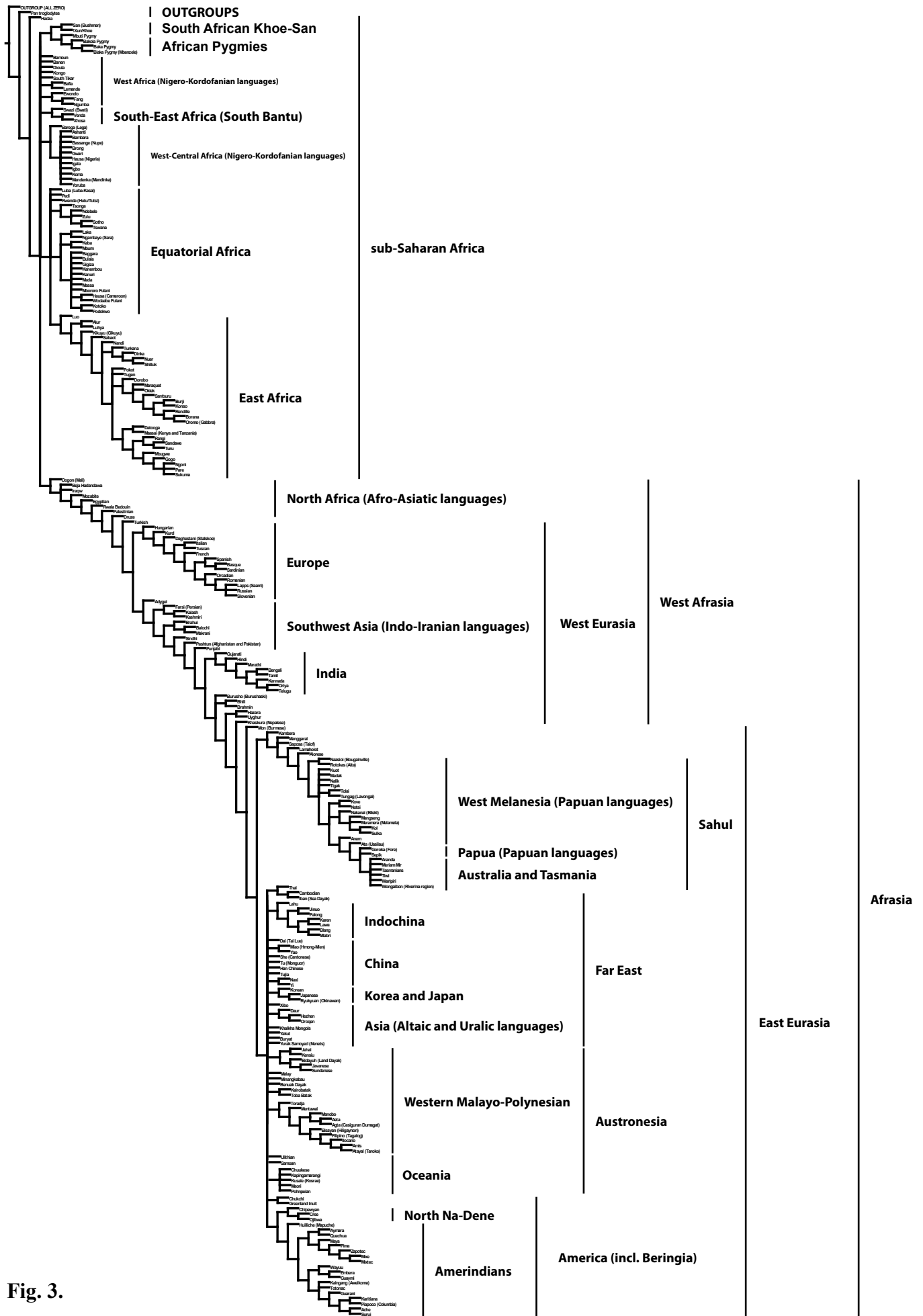


Fig. 3.

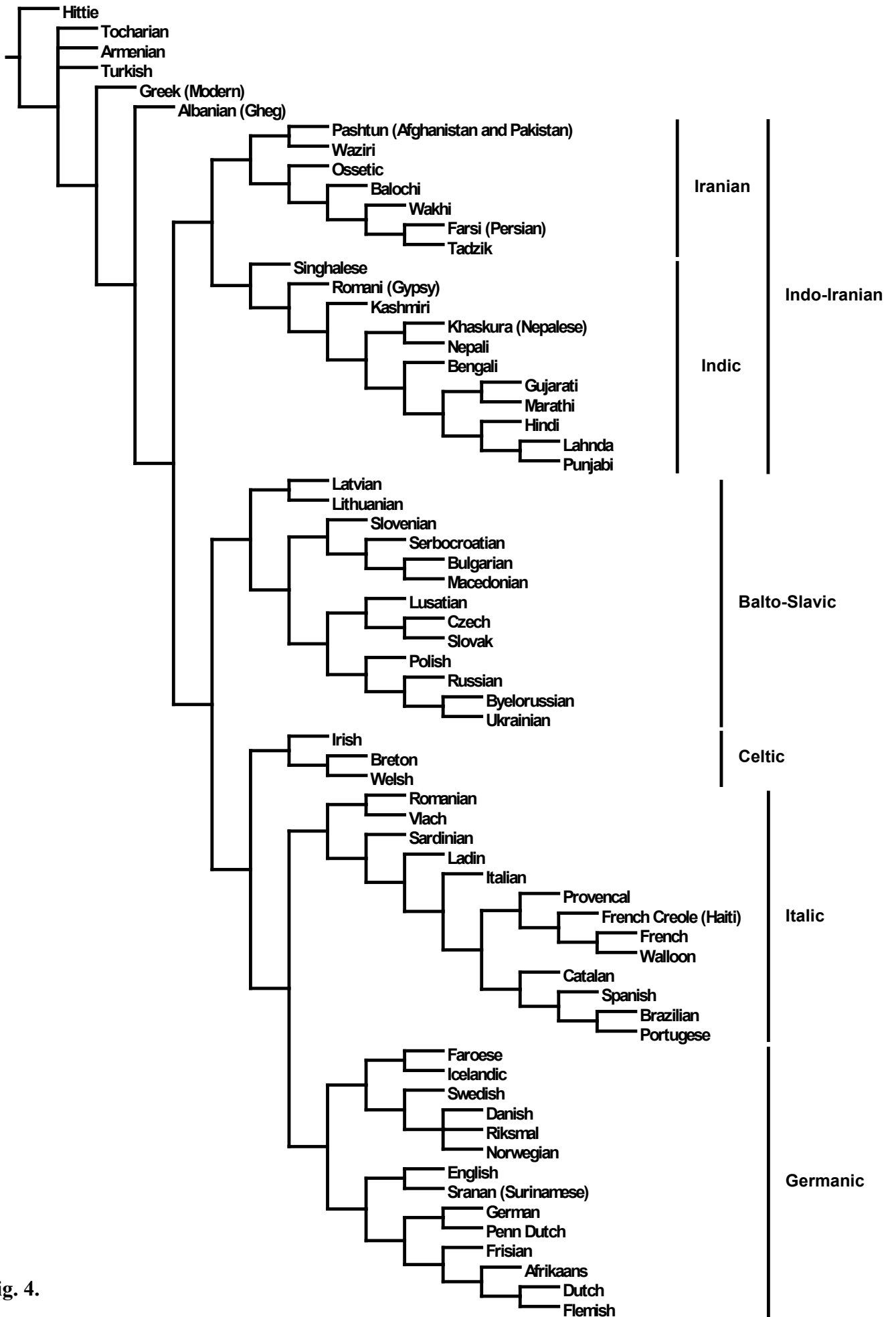


Fig. 4.

Char. 1 - Wealth transfers at marriage
sensu Fortunato et al., 2006

bridewealth
 dowry
 absence of consideration
 ambiguity

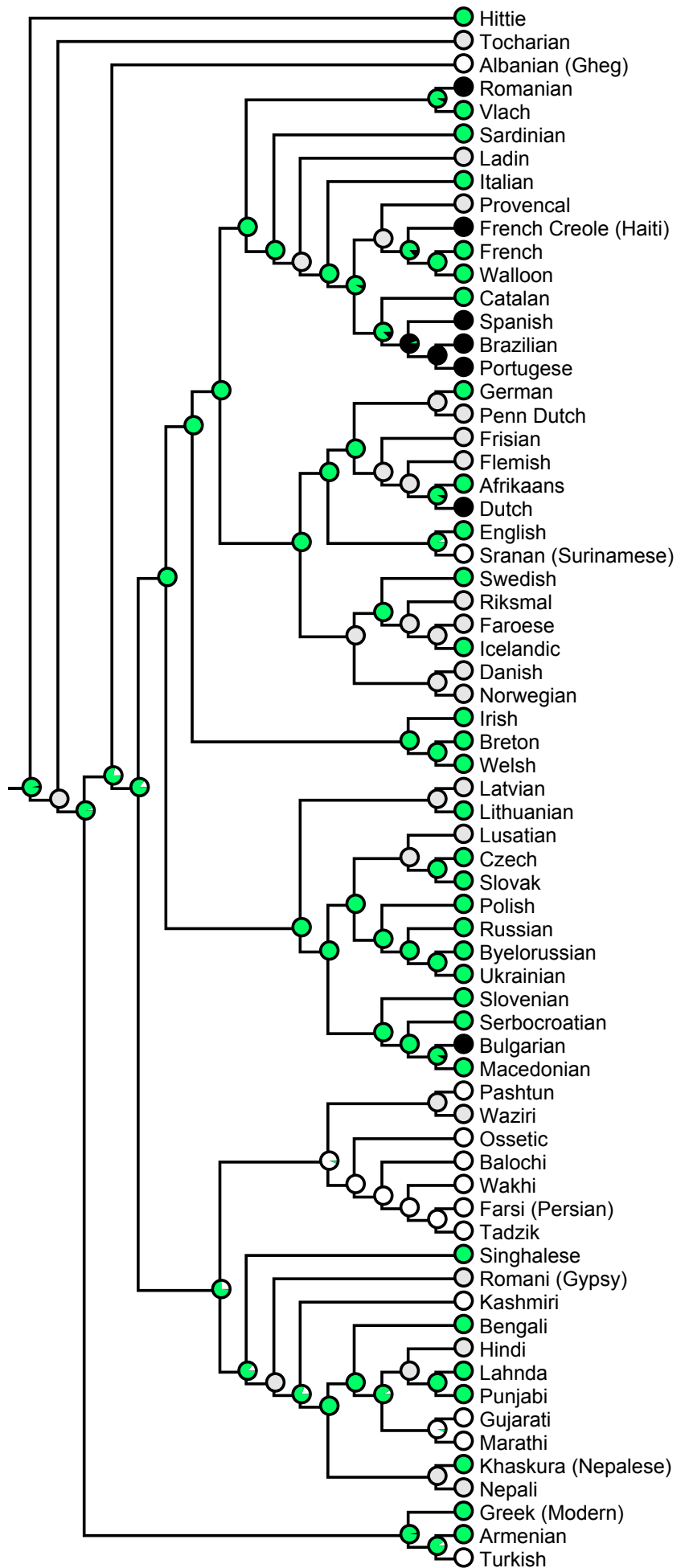


Fig. 5.

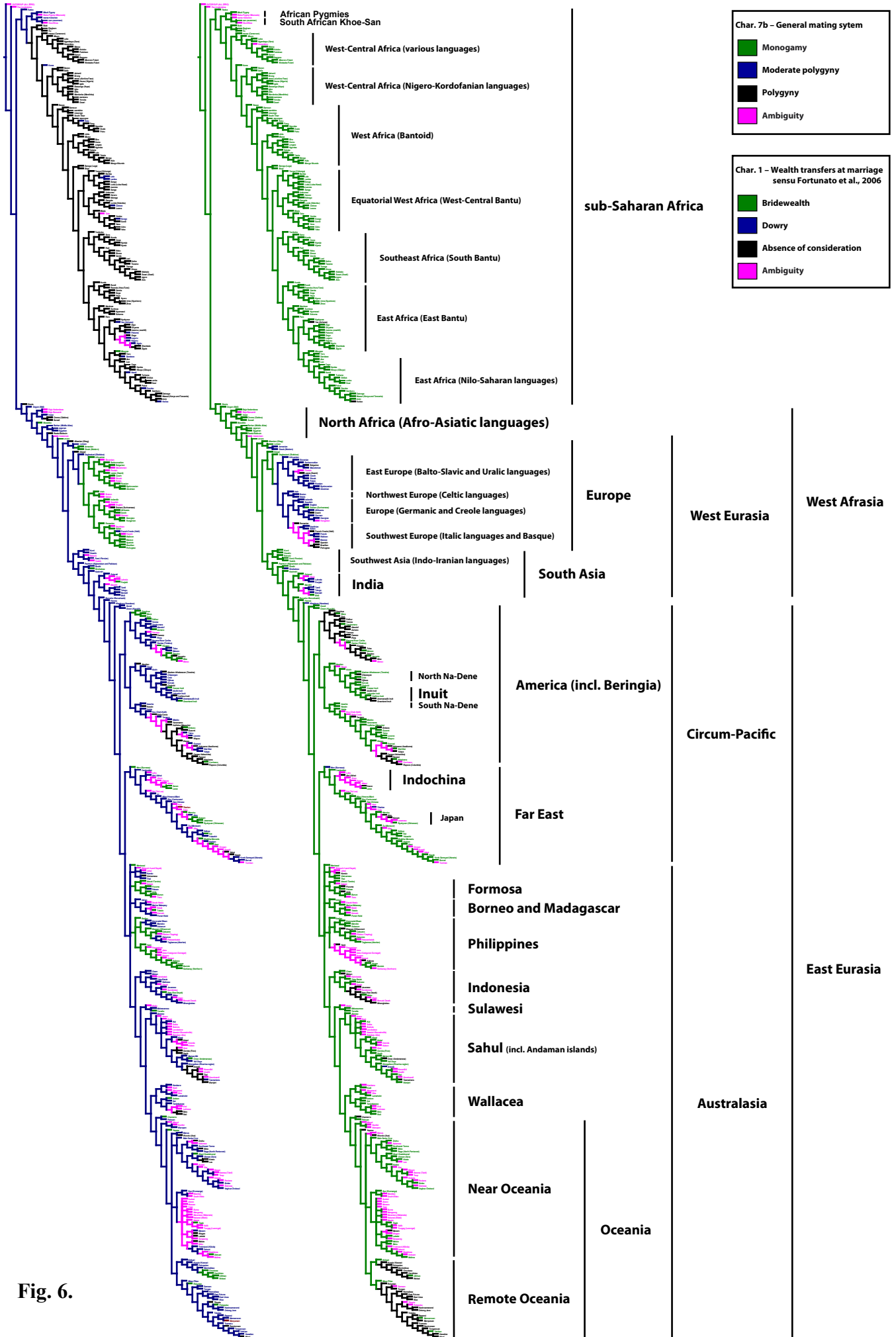


Fig. 6.

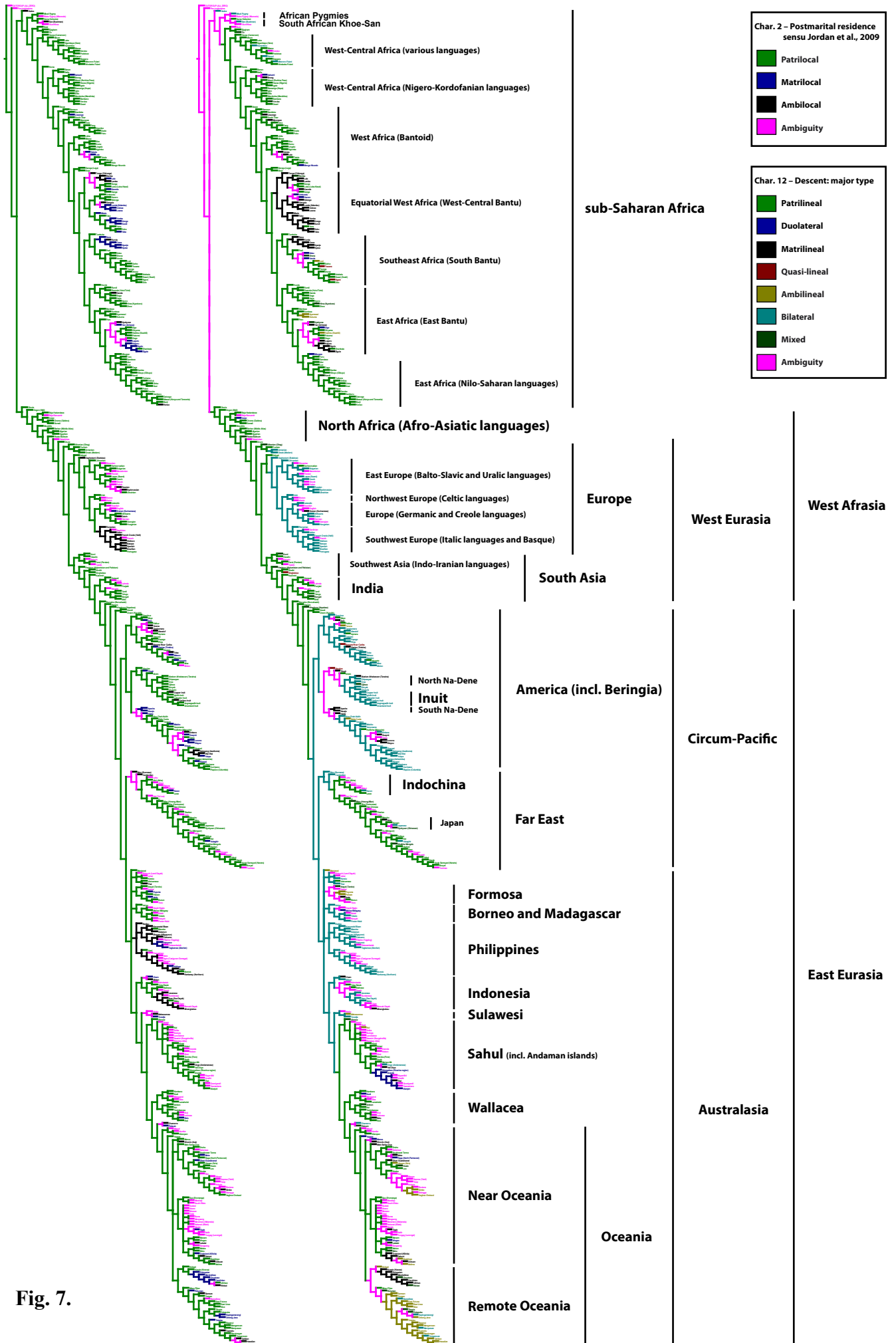


Fig. 7.

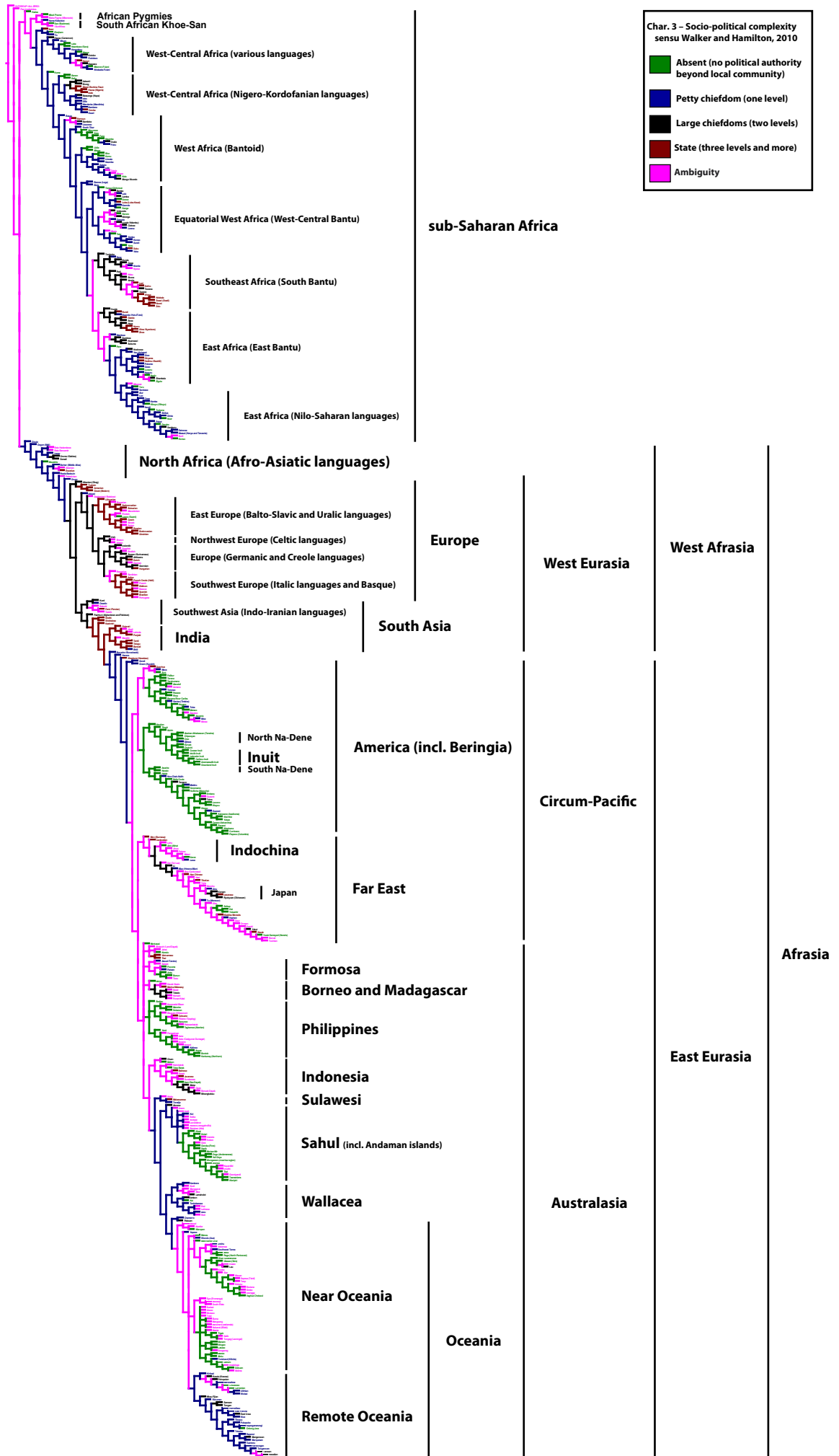


Fig. 8.

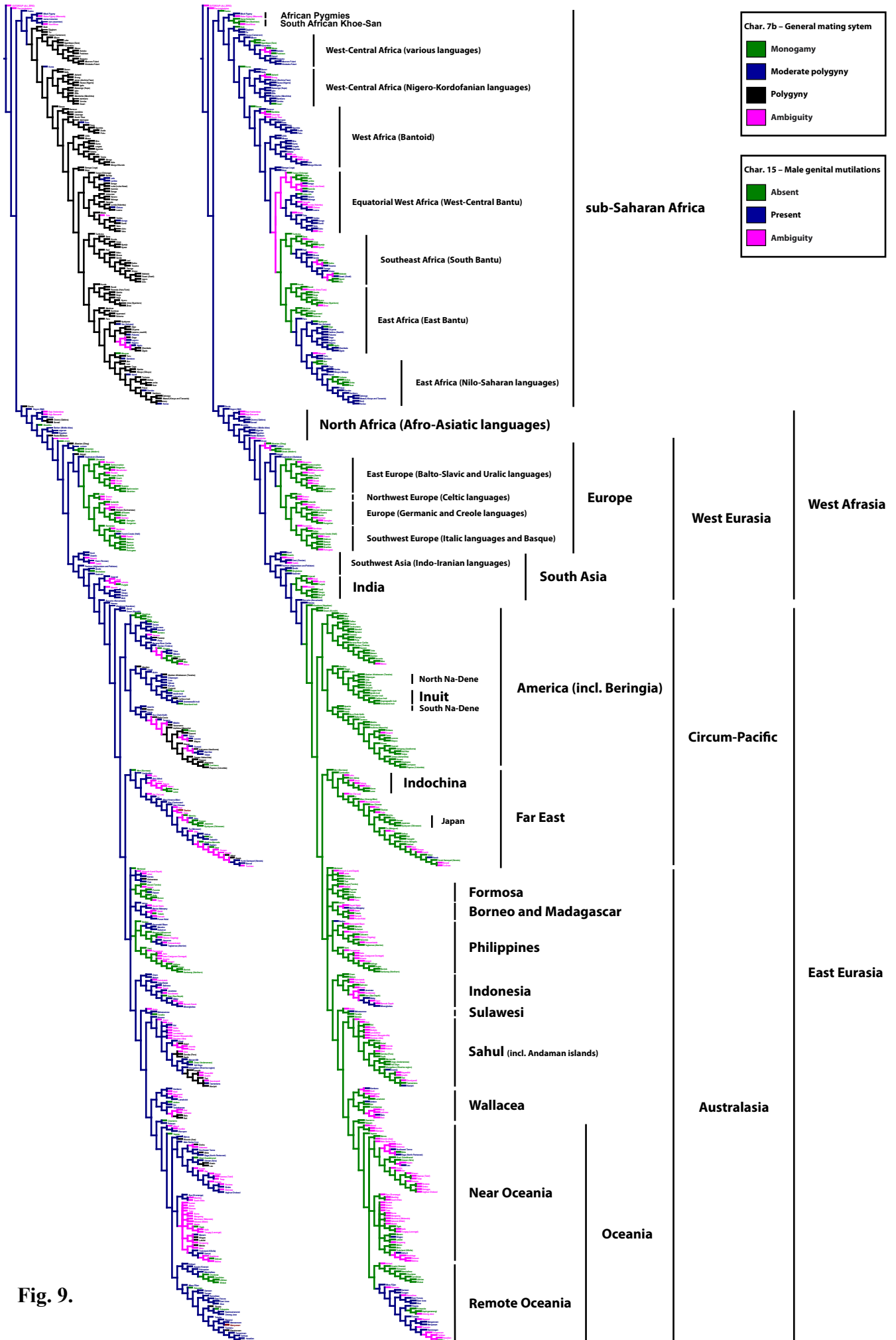


Fig. 9.

APPENDIX

List of socio-cultural and ecological characters

Char. 1 – Wealth transfers at marriage sensu Fortunato et al., 2006

State 0 – Groom's family is expected to give (bridewealth, brideprice, bride service, and the exchange of female relative for a bride included)

State 1 – Bride's family is expected to give (dowry)

State 2 – Absence of consideration

Unordered

Char. 2 – Postmarital residence sensu Jordan et al., 2009

State 0 – Patrilocal (including virilocal)

State 1 – Matrilocal (including uxorilocal and avunculocal)

State 2 – Ambilocal (including neolocal)

Unordered

Char. 3 – Socio-political complexity sensu Walker and Hamilton, 2010

State 0 – Absent (no political authority beyond community)

State 1 – Petty chiefdom (one level)

State 2 – Large chiefdom (two levels)

State 3 – State (three or more levels)

Ordered

Char. 4 – Primary mode of marriage

State 0 – Bridewealth (given to bride's family)

State 1 – Bride service to bride's family

State 2 – Token bride price

State 3 – Reciprocal gift exchange

State 4 – Sister or female relative exchanged for bride

State 5 – Absence of consideration

State 6 – Dowry (given to bride from bride's family)

Unordered

Char. 5 – Alternate mode of marriage

State 0 – Absent

State 1 – Present

Char. 6 – Domestic organization

State 0 – Independent monogamous, nuclear families

State 1 – Independent occasionally polygynous, nuclear families

State 2 – Independent polyandrous families

State 3 – Polygynous families

State 4 – Small extended families

State 5 – Large extended families

Unordered

Char. 7a – Marital composition

State 0 – Independent monogamous, nuclear families

State 1 – Independent occasionally polygynous, nuclear families

State 2 – Sororal polygyny (cowives in same dwelling)

State 3 – Sororal polygyny (cowives in separate dwellings)

State 4 – Non-sororal polygyny (cowives in same dwelling)

State 5 – Non-sororal polygyny (cowives in separate dwellings)

State 6 – Independent polyandrous families

Unordered

Char. 7b – General mating system

State 0 – Monogamy

State 1 – Moderate polygyny

State 2 – Polygyny (sororal or non-sororal)

State 3 – Polyandry (fraternal or non-fraternal)

Unordered

Char. 8 – Establishment of common household in the first years of marriage

State 0 – Absent

State 1 – Present

Char. 9 – Transfer of marital residence (after first years)

- State 0 – Wife to husband’s group
- State 1 – Neolocal
- State 2 – Husband to wife’s group

Unordered

Char. 10 – Marital residence with kin (after first years)

- State 0 – Avunculocal (with husband’s mother’s brother’s kin)
- State 1 – Ambilocal (with either wife’s or husband’s kin)
- State 2 – Matrilocal (with wife’s kin)
- State 3 – Neolocal (separate from kin)
- State 4 – Patrilocal (with husband’s kin)
- State 5 – Uxorilocal (with wife’s kin)
- State 6 – Virilocal (with husband’s kin)

Unordered

Char. 11 – Kinship terminology (Morgans classification)

- State 0 – “Crow”
- State 1 – Descriptive
- State 2 – “Eskimo”
- State 3 – “Hawaiian”
- State 4 – “Iroquois”
- State 5 – “Omaha”
- State 6 – “Sudanese”

Unordered

Char. 12 – Descent: major type

- State 0 – Patrilineal
- State 1 – Duolateral
- State 2 – Matrilineal
- State 3 – Quasi-lineal
- State 4 – Ambilineal
- State 5 – Bilateral

State 6 – Mixed

Unordered

Char. 13 – High gods (utterly transcendent supreme deity)

State 0 – Absent or not reported

State 1 – Present, but not active in human affairs

State 2 – Present, active in human affairs, not supportive of human morality

State 3 – Present, supportive of human morality

Ordered

Char. 14 – Post-partum sex taboo

State 0 – None or very short (1 month or less)

State 1 – 1 to 6 months

State 2 – 6 months to 1 year

State 3 – More than one to two years

State 4 – More than one to two years

State 5 – Over two years

Ordered

Char. 15 – Male genital mutilations

State 0 – Absent

State 1 – Present

Char. 16 – Male genital mutilations (age at performing)

State 0 – Within two months after birth

State 1 – Two months to two years

State 2 – 2 to 5 years

State 3 – 6 to 10 years

State 4 – 11 to 15 years

State 5 – 16 to 25 years

State 6 – 26 to 50 years

State 7 – After 50 years

Ordered

Char. 17 – Segregation of adolescent boys

State 0 – Absent

State 1 – Partial segregation

State 2 – Complete segregation (with relatives outside nuclear family)

State 3 – Complete segregation (with non-relatives)

State 4 – Complete segregation (with peers)

Ordered

Char. 18 – Milking of domestic animals

State 0 – Absent (little or no milking)

State 1 – Present (milked more often than sporadically)

Char. 19 – Prevailing type of dwelling: ground plan

State 0 – Semicircular

State 1 – Circular

State 2 – Elliptical or elongated with rounded ends

State 3 – Polygonal

State 4 – Rectangular or square

State 5 – Quadrangular around inner court

Unordered

Char. 20 – Prevailing type of dwelling: floor level

State 0 – Subterranean or semi-subterranean (ignoring cellars)

State 1 – Floor formed by ground

State 2 – Elevated slightly or on raised platform

State 3 – Raised substantially on piles, posts, or piers

Unordered

Char. 21 – Prevailing type of dwelling: shape of roof

State 0 – Rounded or semi-cylindrical

- State 1 – Dome or hemisphere
- State 2 – Beehive with pointed peak
- State 3 – Conical
- State 4 – Semi-hemisphere
- State 5 – Shed (one slope)
- State 6 – Flat or horizontal
- State 7 – Gabled (two slopes)
- State 8 – Hipped or pyramidal (four slopes)

Unordered

Char. 22a – Norms of premarital sexual behavior of women

- State 0 – Insistence on virginity at marriage
- State 1 – Prohibited but weakly censured and not infrequent
- State 2 – Allowed, censured only if pregnancy results
- State 3 – Trial marriage, promiscuous relations prohibited
- State 4 – Freely permitted, even if pregnancy results

Ordered

Char. 22b – Insistence on virginity of brides

- State 0 – Insistence on virginity absent
 - State 1 – Insistence on virginity present (or marriage at or before puberty)
-

Char. 23 – Office of local headman

- State 0 – Absent
 - State 1 – Present
-

Char. 24 – Succession to the office of local headman

- State 0 – Patrilineal heir
- State 1 – Matrilineal heir
- State 2 – Appointment by higher authority, nonhereditary
- State 3 – Seniority or age, nonhereditary
- State 4 – Influence, wealth or social status, nonhereditary
- State 5 – Election or other formal consensus, nonhereditary
- State 6 – Informal consensus, nonhereditary

Unordered

Char. 25 – Type of hereditary succession (succession to the office of local headman)

State 0 – Hereditary by son (patrilineal)

State 1 – Hereditary by other patrilineal heir (e.g., younger brother)

State 2 – Hereditary by a sister's son (matrilineal)

State 3 – Hereditary by other matrilineal heir (e.g., younger brother)

State 4 – Nonhereditary

Unordered

Char. 26 – Inheritance rules for land

State 0 – Matrilineal (sister's sons)

State 1 – Other matrilineal heirs (e.g., younger brothers)

State 2 – Children, with daughters receiving less

State 3 – Children, equally for both sexes

State 4 – Other patrilineal heirs (e.g., younger brothers)

State 5 – Patrilineal (sons)

Unordered

Char. 27 – Inheritance distribution for land

State 0 – Equal or relatively equal

State 1 – Exclusively or predominantly to the one adjudged best qualified

State 2 – Ultimogeniture (to the junior individual)

State 3 – Primogeniture (to the senior individual)

Unordered

Char. 28 – Inheritance rules for movable property

State 0 – Matrilineal (sister's sons)

State 1 – Other matrilineal heirs (e.g., younger brothers)

State 2 – Children, with daughters receiving less

State 3 – Children, equally for both sexes

State 4 – Other patrilineal heirs (e.g., younger brothers)

State 5 – Patrilineal (sons)

Unordered

Char. 29 – Inheritance distribution for movable property

State 0 – Equal or relatively equal

State 1 – Exclusively or predominantly to the one adjudged best qualified

State 2 – Ultimogeniture (to the junior individual)

State 3 – Primogeniture (to the senior individual)

Unordered

Char. 30 – Class stratification

State 0 – Absent among freemen

State 1 – Wealth distinctions

State 2 – Elite (based on control of land or other resources)

State 3 – Dual (hereditary aristocracy)

State 4 – Complex (social classes)

Unordered

Char. 31 – Slavery

State 0 – Absent

State 1 – Present

Char. 32 – Type of slavery

State 0 – Incipient or nonhereditary

State 1 – Hereditary and socially significant

Char. 33 – Class stratification (endogamy)

State 0 – Absent or insignificant

State 1 – Despised occupational group(s)

State 2 – Ethnic stratification

State 3 – Complex

Unordered

Char. 34a – Dependence on gathering

Char. 35a – Dependence on hunting

Char. 36a – Dependence on fishing

Char. 37a – Dependence on animal husbandry

Char. 38a – Dependence on agriculture

State 0 – 0 – 5% Dependence

State 1 – 6 – 15% Dependence

State 2 – 16 – 25% Dependence

State 3 – 26 – 35% Dependence

State 4 – 36 – 45% Dependence

State 5 – 46 – 55% Dependence

State 6 – 56 – 65% Dependence

State 7 – 66 – 75% Dependence

State 8 – 76 – 85% Dependence

State 9 – 86 – 100% Dependence

Ordered

Char. 34b – Dependence on gathering II

Char. 35b – Dependence on hunting II

Char. 36b – Dependence on fishing II

Char. 37b – Dependence on animal husbandry II

Char. 38b – Dependence on agriculture II

State 0 – 0 – 15% Dependence

State 1 – 16 – 35% Dependence

State 2 – 36 – 55% Dependence

State 3 – 56 – 75% Dependence

State 4 – 76 – 100% Dependence

Ordered

Char. 39 – Subsistence economy

- State 0 – Gathering contributes most
- State 1 – Fishing contributes most
- State 2 – Hunting contributes most
- State 3 – Pastoralism contributes most
- State 4 – Agriculture contributes most
- State 5 – Complex subsistence economy (more sources contribute equally)

Unordered

Char. 40 – Intensity of agriculture

- State 0 – No agriculture
- State 1 – Casual agriculture, incidental to other subsistence modes
- State 2 – Extensive or shifting agriculture
- State 3 – Horticulture, vegetal gardens or groves of fruit trees
- State 4 – Intensive agriculture
- State 5 – Intensive irrigated agriculture

Unordered

Char. 41 – Major crop type

- State 0 – Non-food crops only, such as cotton or tobacco
- State 1 – Vegetables
- State 2 – Tree fruits
- State 3 – Roots or tubers
- State 4 – Cereal grains

Unordered

Char. 42 – Settlement patterns

- State 0 – Nomadic or fully migratory
- State 1 – Seminomadic
- State 2 – Semisettled
- State 3 – Compact but impermanent settlements
- State 4 – Neighborhoods of dispersed family homesteads
- State 5 – Separated villages, forming a single community
- State 6 – Compact and relatively permanent settlements
- State 7 – Complex settlements

Unordered

Char. 43a – Mean size of local communities

- State 0 – Fewer than 50
- State 1 – 50-99
- State 2 – 100-199
- State 3 – 200-399
- State 4 – 400-1000
- State 5 – 1,000 without any town of more than 5,000
- State 6 – Towns of 5,000-50,000 (one or more)
- State 6 – Cities of more than 50,000 (one or more)

Ordered

Char. 43b – Mean size of local communities II

- State 0 – Fewer than 99
- State 1 – 100-299
- State 2 – 1,000 without any town of more than 5,000
- State 3 – Towns of 5,000-50,000 (one or more)
- State 4 – Cities of more than 50,000 (one or more)

Ordered

Char. 44 – Jurisdictional hierarchy of local community

- State 0 – Two levels
- State 1 – Three levels
- State 2 – Four levels (e.g., nuclear family, extended family, clan barrio, village levels)

Ordered

Char. 45 – Jurisdictional hierarchy beyond local community

- State 0 – No levels (no political authority beyond community)
- State 1 – One level (e.g., petty chiefdoms)
- State 2 – Two levels (e.g., larger chiefdoms)
- State 3 – Three levels (e.g., states)
- State 4 – Four levels (e.g., large states)

Ordered

Char. 46 – Types of games: physical skill

Char. 47 – Types of games: chance

Char. 48 – Types of games: strategy

State 0 – Absent

State 1 – Present

Char. 49a – Animals and plow cultivation

State 0 – Absent (no plow animals)

State 1 – Not aboriginal but well established at period of observation

State 2 – Present (aboriginal prior to contact)

Ordered

Char. 49b – Animals and plow cultivation II

State 0 – Absent (no plow animals)

State 1 – Present (Aboriginal prior to contact)

Char. 50 – Predominant type of animal husbandry

State 0 – Absence or near absence of large domestic animals

State 1 – Pigs are the only large domestic animals

State 2 – Sheep and/or goats without larger domestic animals

State 3 – Equine animals (horses, donkeys)

State 4 – Deer (reindeer)

State 5 – Camels

State 6 – Llamas or alpacas

State 7 – Bovine animals (cattle, gayal, yak, water buffalo)

Unordered

Char. 51 – Sex differences: house construction

Char. 52 – Sex differences: gathering

Char. 53 – Sex differences: hunting

Char. 54 – Sex differences: fishing

Char. 55 – Sex differences: animal husbandry

Char. 56 – Sex differences: agriculture

State 0 – Males only or almost alone

State 1 – Males appreciably more

State 2 – Differentiated but equal participation

State 3 – Equal participation, no marked differentiation

State 4 – Females appreciably more

State 5 – Females only or almost alone

State 6 – Absent or unimportant activity

Unordered

References

Gray, J.P., 1999. A Corrected Ethnographic Atlas. *World Cultures* 10, 24-136.

Jordan, F.M., Gray, R.D., Greenhill, S.J., Mace, R., 2009. Matrilocal residence is ancestral in Austronesian societies. *Proc. R. Soc. B* 276, 1957–1964.

Murdock, G.P., 1967. Ethnographic atlas: A summary. *Ethnology* 6, 109-236.

Murdock, G.P., White, D.R., 1969. Standard cross-cultural sample. *Ethnology* 8:329–69.

Figure legends

Fig. A1. Strict consensus of MRP supertrees of human cultures resulting from combined dataset using the “Chimp” rooting.

Fig. A2. Strict consensus of MRP supertrees of human cultures resulting from combined dataset using the “All-0” rooting.

Fig. A3. Strict consensus of MRP supertrees of human cultures resulting from unrooted combined dataset.

Fig. A4. Result of ML optimization of character “Socio-political complexity sensu Walker and Hamilton, 2010” on topology of majority-rule consensus of combined MRP supertrees (see Electronic supplement).

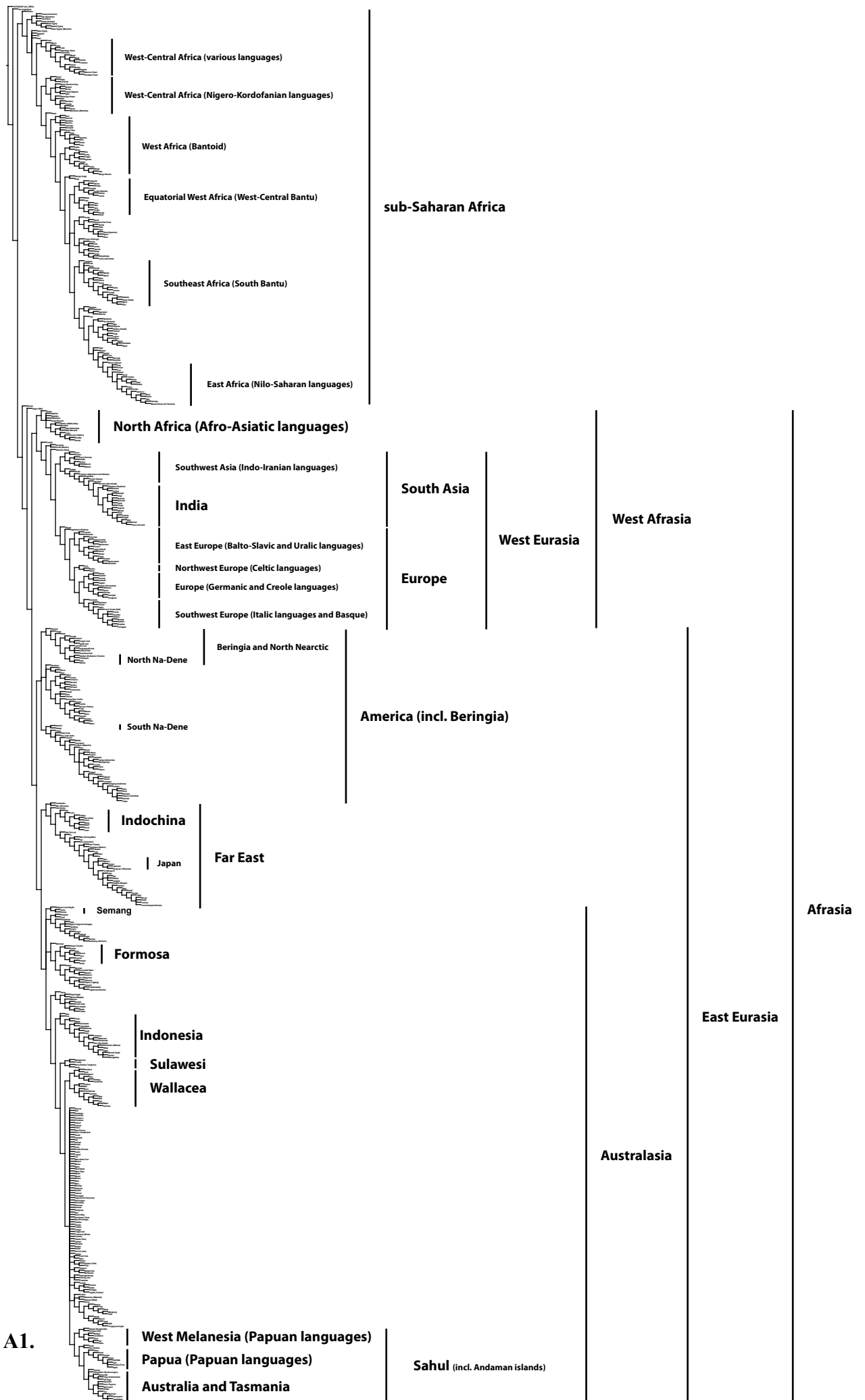


Fig. A1.

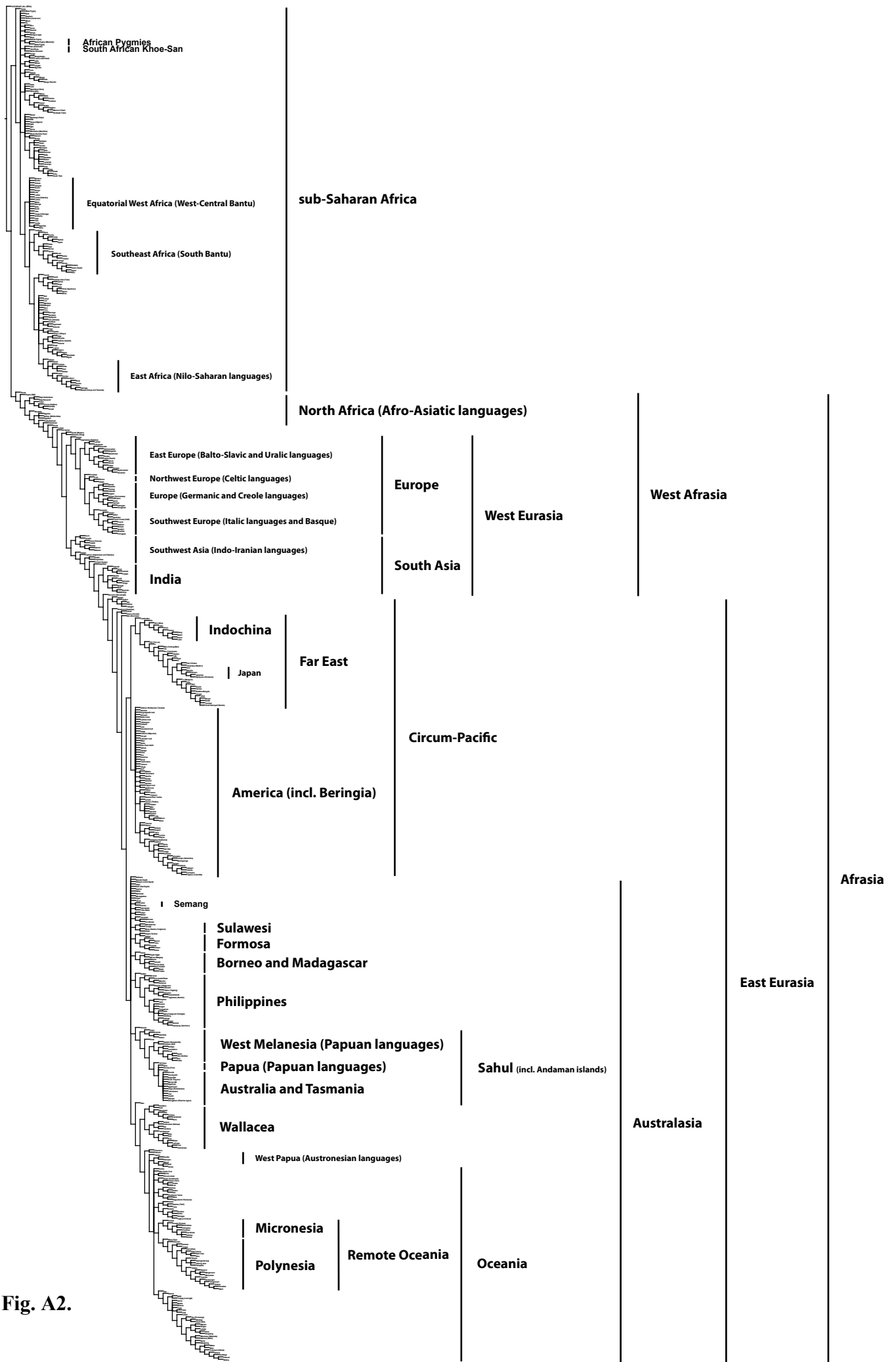


Fig. A2.

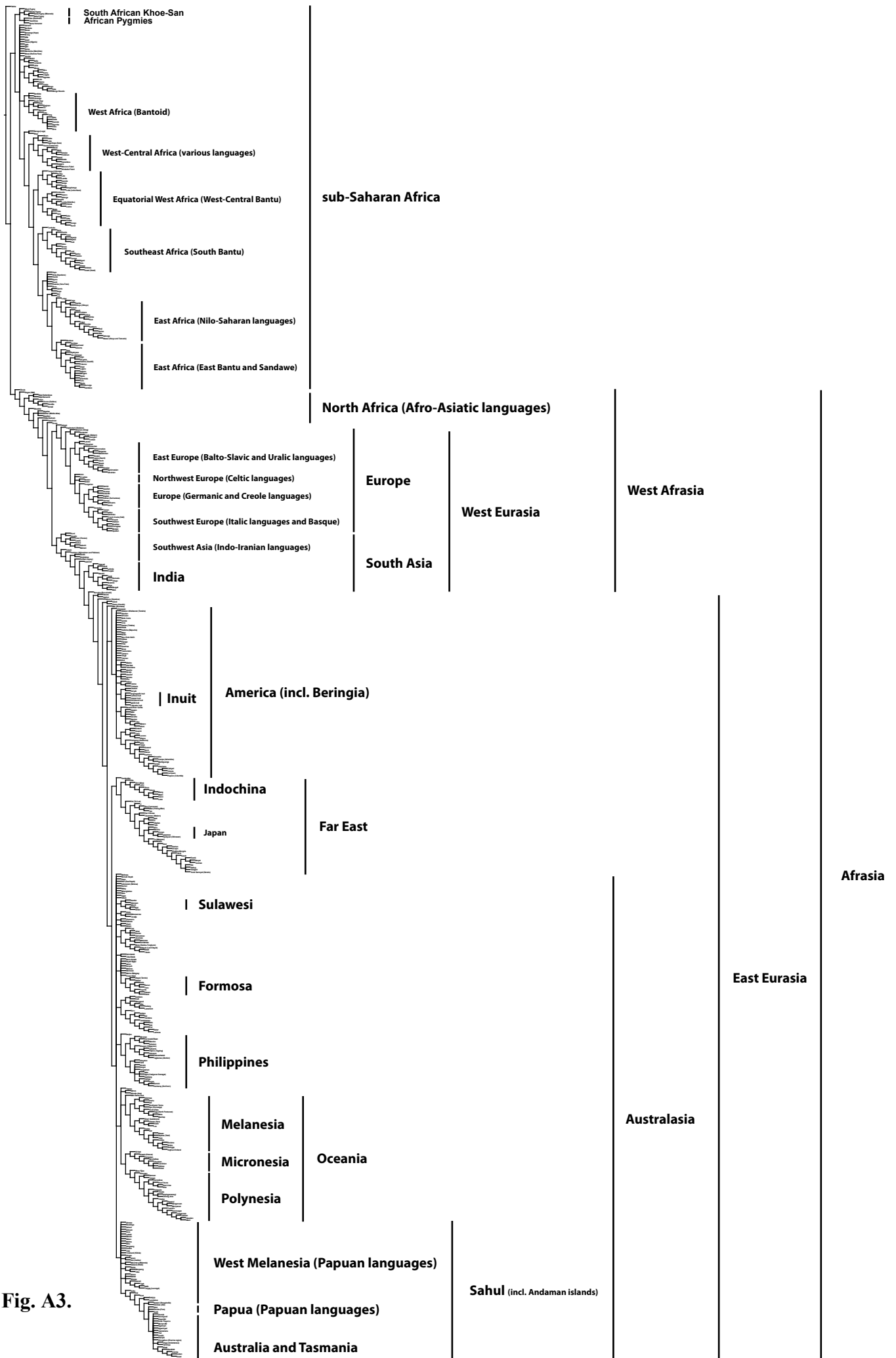


Fig. A3.

Char. 3 – Socio-political complexity sensu Walker and Hamilton, 2010

- Absent (no political authority beyond community)
- Petty chiefdom (one level)
- Large chiefdoms (two levels)
- State (three levels and more)
- Ambiguity

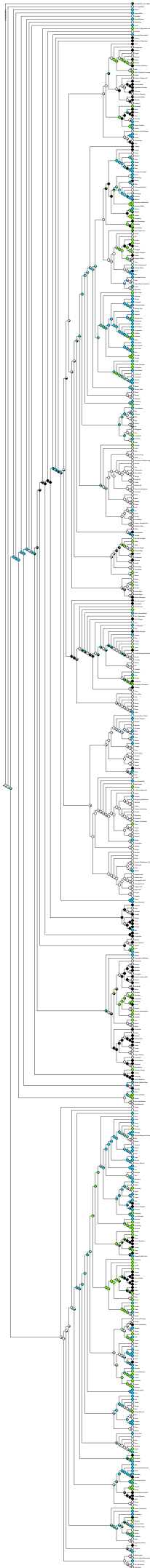


Fig. A4.