Out of Africa with regional interbreeding? Modern human origins

Yoko Satta* and Naoyuki Takahata

Summary

A central issue in paleoanthropology is whether modern humans emerged in a single geographic area and subsequently replaced the preexisting people in other areas. Although the study of human mitochondrial DNAs supported this single-origin and complete-replacement model, a recent paper[1] argues that humans expanded out of Africa more than once and regionally interbred. However, both the genetic antiquity and the impact of the African contribution to modern *Homo sapiens* are so great as to view Africa as a central place of human evolution. Despite the possibility that out-of-Africa *H. sapiens* interbred with other populations, this evidence is more consistent with the uniregional hypothesis than the multiregional hypothesis of modern human origins.
***BioEssays* 24:871–875, 2002.**

Introduction

It is generally accepted that African *H. erectus* (*H. ergaster*) originated in East Africa between 2.5 and 2 million years (my) ago and shortly afterwards began to spread to the Middle East, Southeast Asia, and East Asia. It is also accepted that the descendants of *H. erectus* in Africa and/or the Old World gave rise to *H. sapiens*. However, the question of when, where and how this transformation took place continues to be highly controversial. Howell[2] expressed this situation thus: ''There is now a near consensus among students of human evolutionary biology that the origins of our own species, *Homo sapiens*, is somehow intimately linked with the first intercontinental ancient hominid, *Homo erectus*. However, neither the transformation of *erectus* to *sapiens* nor the transformation of ancient (archaic) populations of *Homo sapiens* to their anatomically modern succedents (*H s sapiens*) are matters of agreement in this scientific fraternity.''

Biosystems Science, Graduate University for Advanced Studies (Sokendai), Hayama Japan.
Funding agency: JSPS; Grant number: 12304046.
*Correspondence to: Yoko Satta, Biosystems Science, Graduate University for Advanced Studies (Sokendai), Hayama, Kanagawa 240-0193, Japan. E-mail: satta@soken.ac.jp
DOI 10.1002/bies.10166
Published online in Wiley InterScience (www.interscience.wiley.com).

There are many different theories for the origin and evolution of modern humans when interpreting hominid fossil records, archeological remains, and genetic data from the current human population. Of these, the two most popular interpretations are referred to as the uniregional and multiregional hypotheses.[3] The multiregional hypothesis[4] posits that *H. sapiens* anagenetically evolved in parallel in multiple founding populations of *H. erectus* that had been connected by gene flow. By contrast, the uniregional hypothesis[5] postulates that *H. erectus* evolved into *H. sapiens* in a single region, presumably in Africa, and then spread worldwide. The controversy became heated when Cann et al.[6] put forward their mitochondrial (mt) DNA haplotype data as evidence for a complete replacement of all the preexisting people in Africa and the rest of the then inhabited world by newly emerged *H. sapiens*.

In his recent paper, Templeton[1] describes a cladistic analysis of haplotype trees, from which he draws phylogeographic evidence for recurrent events of gene flow and at least two major range expansions out of Africa during the Pleistocene. If the most recent out-of-Africa expansion was associated with a complete replacement of the preexisting people, all the earlier phylogeographic signals must have been erased from the current human gene pool. Although it is admitted that, owing to the major expansions out of Africa, the African population has made great impacts on the genetic architecture of the entire human population, Templeton argues that the genetic data support the multiregional model with expansions followed by interbreeding.

To make such arguments, it is necessary to read off and date both historical and recurrent events. The nested clade analysis does this job. The analysis was initially developed for identifying subsets of haplotypes that are associated with significant phenotypic deviations and later modified for identifying nonrandom association of haplotypes with geographic location.[7] To extract phylogeographic information, an implemented inference key is applied to a particular clade (a group of haplotypes connected by one mutational step). To determine relative ages of smaller clades nested within a larger clade, the temporal polarity in the nested clade or the temporal frame of haplotype trees is used. However, the inference key is at best semiquantitative and the association between relative ages of clades and the geographic distribution of haplotypes is not straightforward. To substantiate the empirical finding that the nested clade analysis is not prone to false positives,[1,7] it is

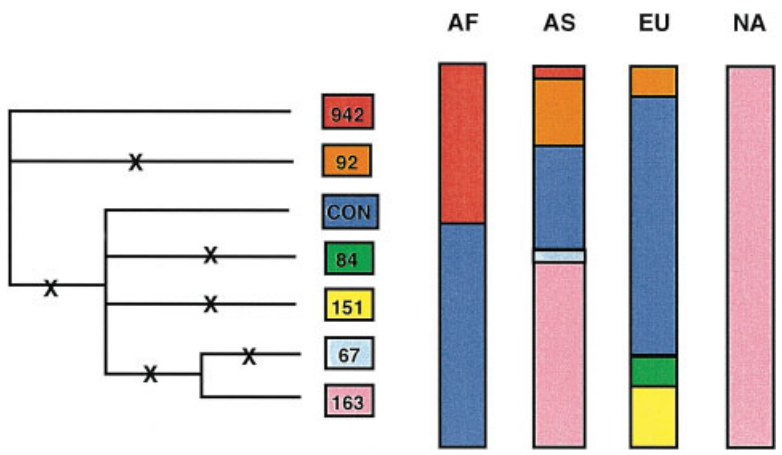

Figure 1. Haplotype tree and frequency of melanocortin 1 receptor (*MC1R*) in Africa (AF), Asia (AS), Europe (EU), and North America (NA).[13] There are seven haplotypes designated here as 942, 92, CON, 84, 151, 67, and 163. An X on the tree marks a nucleotide substitution. Range expansion out of Africa and contiguous range expansion from East and Southeast Asia to North Eurasia in the nested clade analysis[1] are respectively associated with the clade that consists of haplotypes 92 and 942 and the total clade that contains CON and 163. Note that the inference key predicts range expansion if the older haplotype in a clade is confined to a single population while the younger haplotype is widespread geographically or distantly located from the ancestral haplotype.[6]

highly desirable to demonstrate the power and reliability directly by computer simulation. Until this is done, skepticism about the analysis must remain and the result can only be accepted with some reservations.

The genetic data

Templeton[1] uses ten genetic "loci" based on four criteria concerning the number of populations, the sample size, the haplotype frequency, and the haplotype tree. These loci are abbreviated here as mtDNA, Y-DNA, *PDHA1*, *MX1*, *EDN*, *ECP*, *MC1R*, *MS205*, *Xq13.3*, and *HBB*.[8–16] Of these, it is haploid mtDNA and Y-DNA that show an out-of-Africa expansion occurred $T_{exp} = 80{,}000–150{,}000$ years ago. Equally importantly, *HBB, MS205* and *MC1R* exhibit another out-of-Africa expansion well before T_{exp}. Phylogeographic evidence for such an expansion is provided by the *MC1R* haplotype tree (Fig. 1), specifically the geographic distribution of haplotype 92 relative to that of the ancestral haplotype 942. Since this event is dated 0.64 my ago,[1] it predates the emergence of *H. sapiens*. Also, *HBB* and four other loci exhibit early recurrent gene flow. Thus, the out-of-Africa expansion indicated by mtDNA and Y-DNA has not erased signals of an ancient expansion and recurrent gene flow between African and non-African populations. It is then reasonable to conclude that the most recent out-of-Africa expansion occurred with interbreeding.

Templeton[1] refers to the result of Takahata et al.[17] who use a slightly different set of ten "loci". These include four X-linked *PLP*, *GK*, *DYS44*,[18] and ZFX[19] in addition to mtDNA, Y-DNA, *PDHA1*, *MC1R* , Xq13.3, and *HBB* that are analyzed in Templeton.[1] The data are examined with respect to the time (TMRCA) at which the most recent common ancestor (MRCA) occurred and the sampled population

Box 1: Glossary

Coalescence: Genes at a locus sampled from a population are related to each other by reproduction. When a gene is multiplied more than once and the products are transmitted to later generations, such a multiplication appears in a diagram of gene genealogy as a divergence when looked at forward in time and a coalescence when looked at backward in time. The term coalescence refers to this latter case, seeing how ancestral gene lineages merge or coalesce at times of common ancestry.

MRCA: If the ancestral lineages of genes at a locus sampled from a population are traced back, they gradually merge to common ancestors and are eventually derived from single common ancestors. Among these, there is an ancestor that is closest to all sampled genes and referred to as the MRCA or the most recent common ancestor.

TMRCA: This term is defined as the past time at which MRCA for a sample of genes occurs. Since coalescence is a random process, TMRCA is also determined by a probability distribution with a large variance. The larger the population size, the larger the mean of TMRCA.

PMRCA: The coalescence theory was initially developed for a randomly mating population. However, a population is often subdivided and mating occurs only within each local population. In such a structured population, it is important to examine time and space of coalescence, or when and in which local population MRCA occurs. The term PMRCA refers to the local population in which the inferred ancestral haplotype is found most frequently. By this definition, PMRCA is not always identical to the local population in which MRCA exists.

(PMRCA) in which the ancestral haplotype is found most frequently (see Glossary, Box 1). The estimated TMRCA of mtDNA and Y-DNA is within the range of 0.1–0.2 my ago and that of the nuclear loci ranges from 0.41 my ago at *GK* to 1.59 my ago at *PDHA1*. All of these TMRCAs are restricted within the Middle and Lower Pleistocene. During this period several major founding populations of *H. sapiens* would have existed worldwide if the multiregional hypothesis is valid. The maximum parsimony analysis reveals nine African PMRCAs out of the ten loci. The sole exception is *GK*, although there is only one nucleotide site that supports the Asian PMRCA.

Templeton[1] takes this 90% African PMRCA as evidence against a total replacement hypothesis. It is based on the prediction that, under this hypothesis, all haplotype trees with coalescent times greater than T_{exp} years must be rooted in Africa.

The uniregional and multiregional coalescence

In reality, the above prediction does not always hold true even under the uniregional hypothesis. This is due to our inability to directly observe an ancestral population in which MRCA at a locus actually occurs. It can happen that whereas MRCA must occur in Africa, PMRCA is a non-African population. This discordance results from randomness in both the coalescence process and the mutational process, although the extent depends on the values of various parameters. These are the migration rate (m), the sample size (n), and the effective size (N) of descendant African and non-African populations that have existed over the past T_{exp} years. The per-generation mutation rate (u) per locus is also important to obtain a reliable haplotype tree.

Intuitively, it is clear that the larger the values of n and N in the African population relative to those in non-African populations, the higher the proportion of African PMRCA. A high migration rate also favors an increase in genes of African ancestry, because it allows non-African haplotypes to coalesce with African haplotypes during the past T_{exp} years. It turns out, however, that in order to obtain the observed 90% African PMRCA, it is necessary to assume very unrealistic values of m, n and N for African and non-African populations. Essentially, it would be required that all non-African haplotypes coalesced rapidly with African haplotypes but that some other African haplotypes remained that did not coalesce directly with non-African haplotypes. These African-specific haplotypes, if they exist, are likely to be direct descendants of MRCA and contribute to a high proportion of African PMRCA. However, the condition for this African ancestry is stringent and often violated. The simulated coalescence process can substantiate this conclusion (Fig. 2a).

For the multiregional hypothesis, the condition becomes even more stringent. Under this hypothesis, non-African haplotypes tend to segregate from African haplotypes for a relatively long time and these distinct ancestral haplotypes may coalesce relatively slowly (Fig. 2b). Because of this long

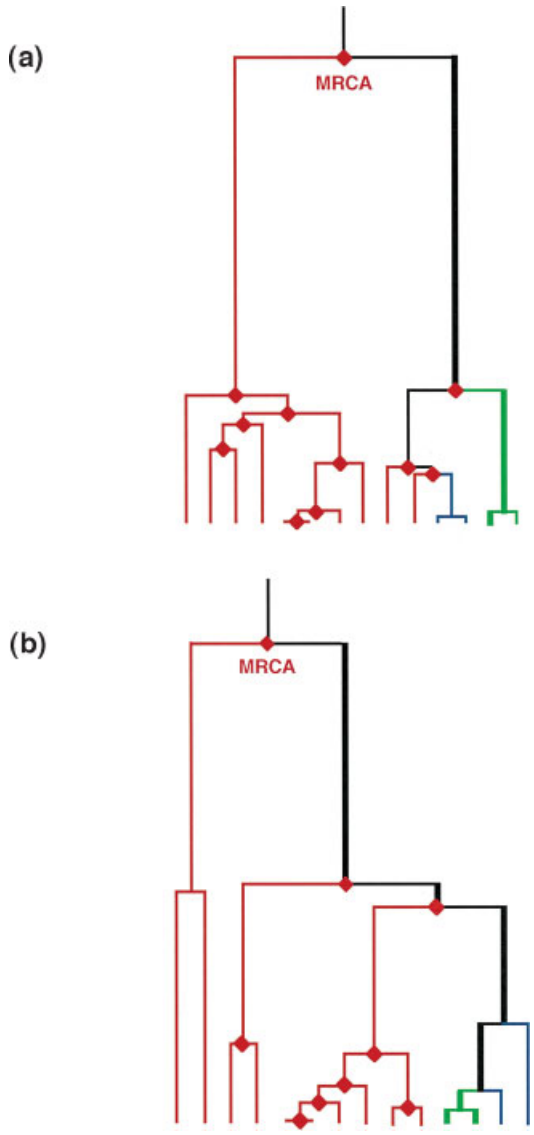

Figure 2. Simulated examples for discordance between PMRCA determined by the maximum parsimony method and the ancestral population in which MRCA actually occurs in the coalescence process.[17] There are three populations (Africa, Asia and Europe) from which $n = 10, 2$ and 2 genes at a locus are respectively sampled. These populations persist for 0.2 my under the uniregional model (a) and for 2 my under the multiregional model (b). In both models, it is assumed that the size of the African population ($N = 10{,}000$) is ten times larger than that of the Asian or European population, $Nm = 2$, $Nu = 5$, and the generation time is 20 years. Each red diamond stands for coalescence in Africa. Red, green and blue lines stand for the ancestral lineages of African, Asian and European haplotypes, respectively. A connected thick line down from MRCA to a haplotype at the tip shows the ancestral lineage for PMRCA. In simulations each with 1000 repeats, the proportion of African PMRCA is 76% in (a) and 72% in (b), most of which can be accounted for by the biased sample size for Africans.

persistence of non-African haplotype lineages, it becomes difficult to obtain more than 80% of African PMRCA.

What then is the effect of multiple out-of-Africa expansions when they are incorporated into the multiregional hypothesis? On the one hand, with a total replacement followed by each expansion, all non-African haplotype lineages can be traced back to Africa so that the situation becomes the same as that of the uniregional hypothesis. On the other hand, when replacement is incomplete and interbreeding is permitted, a different situation may be expected. Nevertheless, even substantial degrees of interbreeding would not help the multiregional hypothesis to account for the high proportion of African PMRCA. Whereas a certain fraction of non-African haplotype lineages can be traced to Africa at the time of expansion, the remaining haplotype lineages persist within non-African populations and it is these lineages that contribute to non-African PMRCA. Thus, a multiregional model with out-of Africa expansions followed by interbreeding creates an intermediate situation but the proportion of African PMRCA cannot be higher than expected under the uniregional hypothesis.

The interpretation and implication

It is the rule rather than the exception that ancestral haplotypes can be found outside Africa with an appreciable frequency. In this regard, the observed human haplotype trees are unusual. At present, excluding *MS205* and *MX1*, there are in total 12 loci available for the analysis of TMRCA and PMRCA. The analysis yields the result that all the TMRCAs occur between 0.1 and 2 my ago and that there are 11 African PMRCAs. The proportion of African PMRCA has increased by additional *EDN* and *ECP* loci,[12] and this is further evidence for an incompatibility with either the uniregional or multiregional hypothesis. An important feature is that, at the eleven loci, there are distinct haplotype lineages in Africa that do not coalesce with non-African haplotype lineages until MRCA. One obvious possibility is that mating within Africa has not been at random. If there have been some African populations that did not directly pump emigrants into non-African populations (Fig. 3), these African populations may preserve ancient haplotype lineages. In fact, the possibility of a complex African population structure is suggested in light of the high level of haplotype diversity at the *CD4* locus: Sub-Saharan populations of diverse geographic origins are more variable than Northeast African or non-African populations.[20] However, it should be kept in mind that for MRCA at most loci to be found during the Pleistocene, such population structure would have to have been rather slight and/or it did not last over 2 my.

Regarding possible interbreeding, it is interesting to ask the conditions for detecting signals of interbreeding or admixture in haplotype trees. It is shown that the statistical

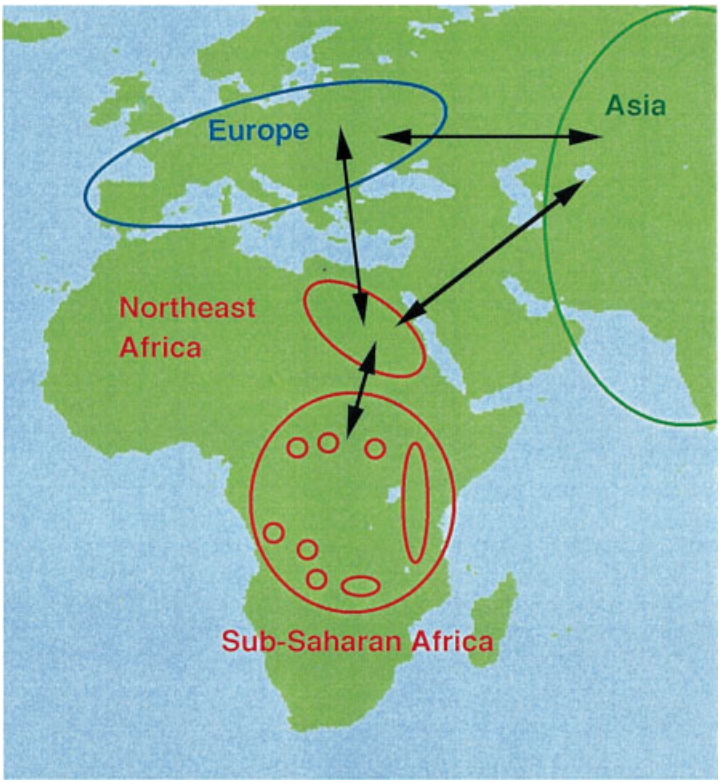

Figure 3. Population structure within Africa that may account for the high haplotype diversity[20] and the high proportion of African PMRCA.[17] It is assumed that the Sub-Saharan African population as a whole is somehow genetically isolated and exchanges migrants with Asian and European populations only through the Northeast African population. Gene flow is indicated by arrows.

power to detect ancient admixture by a single locus haplotype tree is very low unless the admixture involving highly diverged populations was recent.[21] The power depends on the extent of admixture as well. It is argued that, if extensive admixture between Neanderthals and modern human ancestors occurred 30,000 years ago, close to the extinction time of the Neanderthals, and if these two populations diverged 0.4–0.5 my ago, we may (but actually did not) detect admixture in a sample of current human mtDNAs.[17] Use of X-linked or autosomal loci may detect early admixture because of relatively long coalescence times, but it is required that the two populations under study diverged from each other correspondingly early. Single locus information is thus limited and the use of many independent loci would increase the power for detecting multiple out-of-Africa expansions.[1]

Templeton[1] suggests a model that emphasizes the role of out-of-Africa expansion coupled with gene flow and not replacement. Such a model may be realistic and will be strongly supported if the nested clade analysis is shown to be quantitatively reliable. However, it is not immediately clear whether the model can be equated to the multiregional model with expansions followed by admixture. What is shown is that Pleistocene expansions always occurred from Africa. Moreover, in spite of suggested interbreeding in our ancestors, the ancestral haplotype lineage is traced back to Africa at all but one locus. The genetic impact of the African population upon the human gene pool is thus great and that of Eurasian *H. erectus* is correspondingly minor. It therefore seems that these observations actually support the notion of African origins of modern humans and the view that the central role of the African population in human evolution must be associated with its internal population structure.

References

1. Templeton AR. Out of Africa again and again. Nature 2002;416:45–51.
2. Howell FC. In Smith FH, Spencer F, eds; The origins of modern humans. Introduction. New York: Alan R. Liss. 1984. pp. xiii–xxii.
3. Klein J, Takahata N. Where do we come from? The molecular evidence for human descent. Berlin: Springer; 2002. pp. 250–370.
4. Wolpoff MH, Wu X, Thorne AG. In Smith FH, Spencer F, eds; The origins of modern humans. Modern *Homo sapiens* origins: a general theory of hominid evolution involving the fossil evidence from East Asia. New York: Alan R. Liss. 1984. pp. 411–483.
5. Stringer CB, Andrew P. Genetic and fossil evidence for the origin of modern humans. Science 1988;239:1263–1268.
6. Cann RL, Stoneking M, Wilson AC. Mitochondrial DNA and human evolution. Nature 1987;325:31–36.
7. Templeton AR, Routman E, Phillips CA. Separating population structure from population history: A cladistic analysis of the geographical distribution of mitochondrial DNA haplotypes in the tiger salamander, *Ambystoma tigrinum*. Genetics 1995;140:767–782.
8. Ingman M, Kaessmann H, Pääbo S, Gyllensten U. Mitochondrial genome variation and the origin of modern humans. Nature 2000;408:708–713.
9. Hammer MF, Karafet T, Rasanayagam A, Wood ET, Altheide TK, Jenkins T, Griffiths RC, Templeton AR, Zegura SL. Out of Africa and back again: Nested cladistic analysis of human Y chromosome variation. Mol Biol Evol 1998;15:427–441.
10. Harris EE, Hey J. X chromosome evidence for ancient human histories. Proc Natl Acad Sci USA 1999;96:3320–3324.
11. Jin L, Underhill PA, Doctor V, Davis RW, Shen P, Cavalli-Sforza LL, Oefner PJ. Distribution of haplotypes from a chromosome 21 region distinguishes multiple prehistoric human migrations. Proc Natl Acad Sci USA 1999;96:3796–3800.
12. Zhang J, Rosenberg HF. Sequence variation at two eosinophil-associated ribonuclease loci in humans. Genetics 2000;156:1949–1958.
13. Rana BK, Hewett-Emmett D, Jin L, Chang BH-J, Sambuughin N, Lin M, Watkins S, Bamshad M, Jorde JB, Ramsay M, Jenkins T, Li W-H. High polymorphism at the human melanocortin 1 receptor locus. Genetics 1999;151:1547–1557.
14. Rogers EJ, Shone AC, Alonso S, May CA, Armour JAL. Integrated analysis of sequence evolution and population history using hypervariable compound haplotypes. Hum Mol Genet 2000;9:2675–2681.
15. Kaessmann H, Heiβig F, Haeseler AV, Pääbo S. DNA sequence variation in a non-coding region of low recombination on the human X chromosome. Nature Genet 1999;22:78–81.
16. Harding RM, Fullerton MS, Griffths RC, Bond J, Cox MJ, Schneider JA, Moulin DS, Clegg JB. Archaic African and Asian lineages in the genetic ancestry of modern humans. Am J Hum Genet 1997;60:772–789.
17. Takahata N, Lee S-H, Satta Y. Testing multiregionality of modern human origins. Mol Biol Evol 2001;18:172–183.
18. Nachman MW, Bauer VL, Crowell SL, Aquadro CF. DNA variability and recombination rates at X-linked loci in humans. Genetics 1998;150:1133–1141.
19. Jaruzelska J, Zietkiewicz E, Batzer M, Cole DEC, Moisan J-P, Scozzari R, Tavaré S, Labuda D. Spatial and temporal distribution of the neutral polymorphisms in the last ZFX intron: analysis of the haplotype structure and genealogy. Genetics 1999;152:1091–1101.
20. Tishkoff SA, Dietzsch E, Speed W, Pakstis AJ, Kidd JR, Cheung K, Bonné-Tamir B, Santachiara-Benerecetti AS, Moral P, Krings M, Pääbo S, Watson E, Risch N, Jenkins T, Kidd KK. Global patterns of linkage disequilibrium at the CD4 locus and modern human origins. Science 1996;271:1380–1387.
21. Nordborg M. In Donnelly P, Foley RA, eds; Genes, fossils and behaviour. On detecting ancient admixture. Amsterdam: IOS Press. 2001. pp. 123–136.