Letters to the Editor

Am. J. Hum. Genet. 61:980–983, 1997

The Myth of Bumpy Hunter-Gatherer Mismatch Distributions

To the Editor:

Watson et al. (1996) have elaborated on figure 8 of Mountain et al. (1995) by contrasting the mtDNA mismatch distributions of African hunter-gatherers and food producers. Their claim is that food production influences mtDNA mismatch distributions so that hunter-gatherers have "bumpy" mismatch distributions, whereas food producers tend to have "bell-shaped" distributions. This is an astonishing alternative hypothesis to that of Rogers and Harpending (1992), Sherry et al. (1994), and Rogers (1995), who concluded that a Pleistocene population explosion ~60,000 years ago caused the bell-shaped mtDNA mismatch distribution found in most human populations.

Agriculture emerged after the Younger Dryas glacial interlude, that is, within the last 10,000 years, with several independent centers of origin at different times (cf. Smith 1994). Following the logic of Watson et al. (1996), all populations 10,000 years ago would have had bumpy mismatch distributions. But how many mutations can be acquired in an expanding population within 10,000 years? According to the mutation rate Watson et al. (1996) have in mind, one mutation, on average, between two sequences (underpinned by computer simulations, see Forster et al. 1996); this would just shift mismatch distributions by one step to the right with little smoothing. In particular, it cannot bridge gaps between distinct major modes of the distribution. If, for example, the Biaka (West Pygmies) of today were to start clearing their rain forest and became successful food producers, how many years would we have to wait to see the bumps in their mismatch distribution, which are $\leqslant$18 mutations apart, melt away? It is evident that a period of just a few thousand years would not suffice, even if the mutation rate were 10 times faster. (In fact, Watson et al. [1996] have already doubled the mutation rate by confusing divergence with substitution rate when citing Ward et al. [1991].)

Another problem is the definition of a bumpy distribution. Is it one with many peaks? In contrast to the authors' claim, it seems that in figure 2 of Watson et al. (1996) the food-producing Fulbe, Kanuri, and Tuaregs have a mismatch distribution that is at least as bumpy as that of the hunter-gatherer !Kung of Botswana and Sekele of Namibia (see our fig. 1), since all of these distributions have four peaks. The classification becomes even more doubtful when investigating the reasons for the peaks: a single outlier sequence in the !Kung sample causes bumpiness by disturbing an otherwise bell-shaped distribution (with mode at two mutations), while for the Kanuri sample deletion of a single sequence would not restore bell-shapedness. It thus appears that bumpiness sensu Watson et al. (1996) is highly subjective, because it does not draw on any numerical measurements, such as those suggested by Sherry et al. (1994) or Harpending et al. (1993). In fact, the mismatch distributions that they perceive as nonbumpy are, except for the Hausa and Somali samples, far from unimodal and bell shaped: the rectangular-shaped distribution of the Mandenka, for example, can arise only as a superposition of a number of distributions with quite different modes and cannot be explained by a simple expansion process, be it sudden or exponential. In a very diverse gene pool, such as that of Africa, population fusion events can easily generate bumpy mismatch distributions in expanding populations.

Even if bell-shapedness (appropriately measured) did correlate positively with food production, such a correlation would not prove a causal relationship. For example, the same geographic factors that cause a population

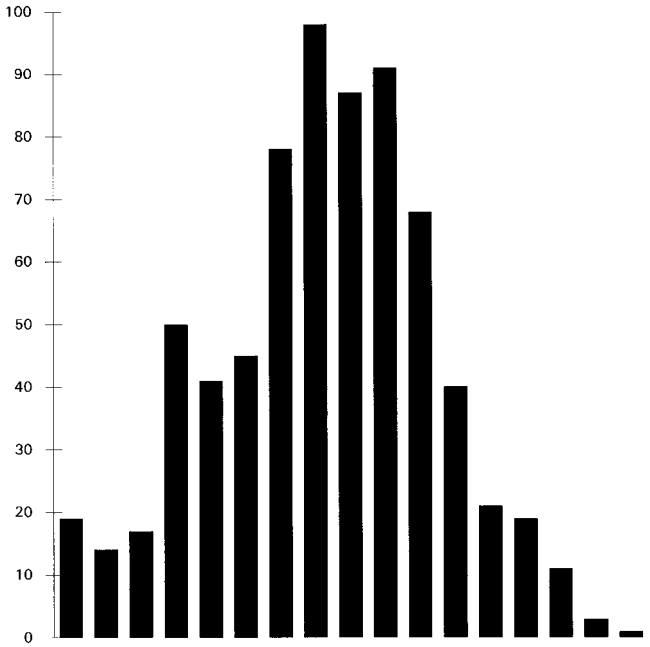

Figure 1 Distribution of pairwise differences in the Sekele of Namibia (hunter-gatherers), based on sequences for the first hypervariable segment (data from Soodyall 1993). The distribution was calculated using MacPairwise 5.0 (Macaulay and Micklem 1995).

to grow and to acquire a bell-shaped distribution, could, many tens of thousands of years later, also favor the introduction of agriculture. In order to link a population expansion event, as perceived in genetic data, to the onset of agriculture, it is therefore indispensable to date genetically the expansion and then compare it to the archaeological record for agriculture. No attempt has been made by Watson et al. (1996) to adduce such a proof, and in fact their analytical methods could not accomplish this complex task.

One of the basic assumptions of standard coalescent models is that the population in question is genetically isolated from other populations. This is strictly justified only when the human species is mismatch analyzed as a whole. While mismatch analysis of the human species may be rather insensitive to population substructure (Rogers et al. 1996), mismatch analysis of human subpopulations may be strongly affected by bottlenecks and fusion processes. The genetic exchange that populations (ethnic groups) have experienced over tens of thousands of years is impossible to model, because it would not follow any predictable regular pattern. The analysis of Watson et al. (1996) does not address this crucial point, inevitably leading to unrealistic coalescence ages, as we explain in the following. Many populations in Africa coalesce close to the human mtDNA coalescent (mtEve) for a trivial reason: there is one cluster, characterized by a transversion at nucleotide position (np) 16188 (numbering according to Anderson et al. 1981) and at least three additional transitions, that sticks out in all phylogenetic analyses of African populations where it occurs. This cluster, described by Bandelt et al. (1995), includes the major African 9-bp deletion subcluster (Soodyall et al. 1996) and is widespread in Africa. It is even found as single outliers in Sardinia, the Middle East (Di Rienzo and Wilson 1991), and Turkey (Calafell et al. 1996). According to Horai and Hayasaka (1990) as well as Tamura and Nei (1993), this cluster constitutes the deepest rooting lineage of their mtDNA trees, and in other analyses it would also branch off very deeply. Therefore, all these populations, including $\geqslant 9$ of the 13 populations used by Watson et al. (1996), such as the Senegalese Mandenka, coalesce close to mtEve. The coalescence time of 9,000–21,000 years for the Mandenka and thus for mtEve, as calculated by Watson et al. (1996) in their table 3, compares unfavorably with current estimates of 140,000–160,000 years for mtEve (Horai et al. 1995; Tamura and Nei 1993). The other populations in their table 3 fare little better. A glance at table 2 of Graven et al. (1995) suggests that the Mandenka have a pronounced population structure and can be divided into three very different population components, essentially represented by RFLP status 1-2, 2-2, and 7-2. Since these haplogroups have quite different geographic distributions and diversities in the other sub-Saharan populations, it is extremely implausible that this mix of lineages arose in the Mandenka, as the approach of Watson et al. (1996) assumes.

Genetic distances between populations as those compiled by Watson et al. (1996) contribute even less than mismatch distributions to the understanding of the relationships between populations. For instance, Kikuyu and Turkana, as well as Kanuri and Hausa, are each at distance 0 in their table 2, although even the mismatch distributions in their figure 2 can clearly distinguish these populations. Recent admixture of a group of very distant lineages (such as the 9-bp cluster) into two or more populations inflates genetic distances, and a tree analysis of these distances (their fig. 3) can misinterpret this recent admixture as an ancient population split.

Finally, we briefly present an analysis of the !Kung of Botswana to demonstrate that most of their mtDNA diversity predates their ethnogenesis, and therefore the total diversity should not be used to infer the private history of the !Kung. Although the !Kung sample consists of only 20 sequences, it coalesces very deeply (cf. Tamura and Nei 1993). The major !Kung cluster (comprising all !Kung except one outlier) is also shared by other Khoisan populations: We collect all sequences from Soodyall (1993) (cf. Sherry et al. 1994), Vigilant (1990), and Watson et al. (1996) sharing the motif np 16187, 16230, and 16243, which are the most conservative positions (according to Wakeley 1993) in the !Kung consensus sequence. To take reversals in the motif into

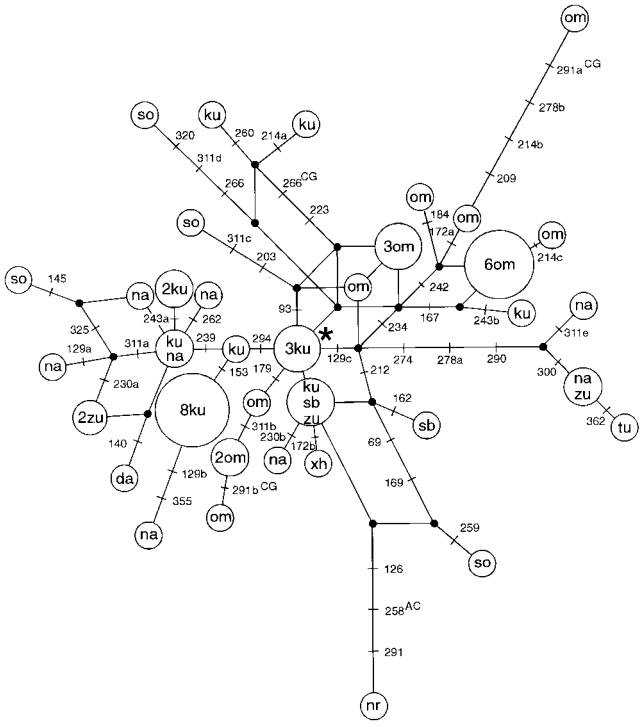

Figure 2 Reduced median network (Bandelt et al. 1995) of !Kung-motif lineages. Network comprises all sequences (taken from Soodyall 1993; Vigilant 1990; and Watson et al. 1996; np 16069–16362) with the !Kung motif np 16187, 16230, and 16243, plus the sequences differing by one motif mutation from these sequences or their immediate neighbors. The two-letter code designates the following populations: ku = !Kung and nr = Naron (Vigilant et al. 1991); da = Dama, na = Nama, om = Sekele, so = Sotho, xh = Xhosa, and zu = Zulu (Soodyall 1993); sb = southeast Bantu (Soodyall et al. 1996); and tu = Turkana (Watson et al. 1996). The node marked with an asterisk (*) represents the !Kung consensus np 16129, 16187, 16189, 16223, 16230, 16243, and 16311 and the putative root of the network. Numbers preceding the code indicate sequence frequency; numbers along links indicate np's (less 16,000) of mutations, and transversions are specified in superscript; resolved recurrent mutations are labeled alphabetically. Note that it is necessary to correct the reading frame shift by one position at np 16176 in Soodyall (1993) and that ambiguous sequences are likely to cause additional reticulations and minor topological changes in the network.

account, we include closely related sequences. It is interesting to note that a few of these sequences were found in non-Khoisan populations, and these sequences presumably represent recent admixture. The reduced median network (fig. 2) for this data reveals that the !Kung lineages are inseparable by all but six private mutations from other southern African populations. In particular, 8 of 18 Nama (Khoi) lineages are interspersed in the !Kung (San) cluster. Even the outlier sequence in the !Kung (outside the network) is close to other Khoisan sequences: it differs from a Sekele sequence at only one position. The !Kung hence seem to represent only a splinter of a former widespread Khoisan population, and their differentiation from other Khoisan populations may have occurred quite recently relative to the !Kung coalescence time.

In summary, mismatch and distance analyses alone are insufficient tools to uncover the history of ethnic groups or mankind as a whole, since "information is lost in studying pairwise difference data rather than the sequences themselves" (Marjoram and Donnelly 1994, pp. 680–681). Phylogenetic analysis of human mtDNA data is a daunting task, but it nevertheless must be tackled if we are to reconstruct our genetic prehistory.

HANS-JÜRGEN BANDELT[1] AND PETER FORSTER[2]
[1]*Mathematisches Seminar and* [2]*Heinrich-Pette-Institut für Experimentelle Virologie und Immunologie, Universität Hamburg, Hamburg*

References

Anderson S, Bankier AT, Barrell BG, de Bruijn MHL, Coulson AR, Drouin J, Eperon IC, et al (1981) Sequence and organisation of the human mitochondrial genome. Nature 290: 457–465

Bandelt H-J, Forster P, Sykes BC, Richards MB (1995) Mitochondrial portraits of human populations using median networks. Genetics 141:743–753

Calafell F, Underhill P, Tolun A, Angelicheva D, Kalaydjieva L (1996) From Asia to Europe: mitochondrial DNA sequence variability in Bulgarians and Turks. Ann Hum Genet 60: 35–49

Di Rienzo A, Wilson AC (1991) Branching pattern in the evolutionary tree for human mitochondrial DNA. Proc Natl Acad Sci USA 88:1597–1601

Forster P, Harding RH, Torroni A, Bandelt H-J (1996) Origin and evolution of Native American mtDNA variation: a reappraisal. Am J Hum Genet 59:935–945

Graven L, Passarino G, Semino O, Boursot P, Santachiara-Benerecetti S, Langaney A, Excoffier L (1995) Evolutionary correlation between control region sequence and restriction polymorphisms in the mitochondrial genome of a large Senegalese Mandenka sample. Mol Biol Evol 12:334–345

Harpending HC, Sherry ST, Rogers AR, Stoneking M (1993) The genetic structure of ancient human populations. Curr Anthropol 34:483–496

Horai S, Hayasaka K (1990) Intraspecific nucleotide sequence differences in the major noncoding region of human mitochondrial DNA. Am J Hum Genet 46:828–842

Horai S, Hayasaka K, Kondo R, Tsugane K, Takahata N (1995) Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs. Proc Natl Acad Sci USA 92:532–536

Macaulay VA, Micklem KJ (1995) MacPairwise 5.0. Computer program at the Institute of Molecular Medicine, University of Oxford, Oxford

Marjoram P, Donnelly P (1994) Pairwise comparisons of mitochondrial DNA sequences in subdivided populations and implications for early human evolution. Genetics 136:673–683

Mountain JL, Hebert JM, Bhattacharyya PA, Underhill PA, Ottolenghi C, Gadgil M, Cavalli-Sforza LL (1995) Demographic history of India and mtDNA-sequence diversity. Am J Hum Genet 56:979–992
Rogers AR (1995) Genetic evidence for a Pleistocene population explosion. Evolution 49:608–615
Rogers AR, Fraley AE, Bamshad MJ, Watkins WS, Jorde LB (1996) Mitochondrial mismatch analysis is insensitive to the mutational process. Mol Biol Evol 13:895–902
Rogers AR, Harpending H (1992) Population growth makes waves in the distribution of pairwise differences. Mol Biol Evol 9:552–569
Sherry ST, Rogers AR, Harpending H, Soodyall H, Jenkins T, Stoneking M (1994) Mismatch distributions of mtDNA reveal recent human population expansions. Hum Biol 66: 761–775
Smith BD (1994) The emergence of agriculture. Scientific American Library, New York
Soodyall H (1993) Mitochondrial DNA polymorphisms in Southern African populations. PhD thesis, University of the Witwatersrand, Johannesburg
Soodyall H, Vigilant L, Hill AV, Stoneking M, Jenkins T (1996) mtDNA control-region sequence variation suggests multiple independent origins of an "Asian-specific" 9-bp deletion in sub-Saharan Africans. Am J Hum Genet 58:595–608
Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol 10:512–526
Vigilant L (1990) Control region sequences from African populations and the evolution of human mitochondrial DNA. PhD thesis, University of California, Berkeley
Vigilant L, Stoneking M, Harpending H, Hawkes K, Wilson AC (1991) African populations and the evolution of human mitochondrial DNA. Science 253:1503–1507
Wakeley J (1993) Substitution rate variation among sites in hypervariable region 1 of human mitochondrial DNA. J Mol Evol 37:613–623
Ward RH, Frazier BL, Dew-Jager K, Pääbo S (1991) Extensive mitochondrial diversity within a single Amerindian tribe. Proc Natl Acad Sci USA 88:8720–8724
Watson E, Bauer K, Aman R, Weiss G, von Haeseler A, Pääbo S (1996) mtDNA sequence diversity in Africa. Am J Hum Genet 59:437–444

Address for correspondence and reprints: Prof. Dr. Hans-Jürgen Bandelt, Mathematisches Seminar, Universität Hamburg, Bundesstraße 55, D-20146 Hamburg, Germany.

0002-9297/97/6104-0029$02.00

Am. J. Hum. Genet. 61:983, 1997

Reply to Bandelt and Forster

To the Editor:

It is well known (and is clearly stated in our article [Watson et al. 1996]) that patterns in mismatch distributions can be influenced by many evolutionary scenarios other than population growth. In our study, we therefore applied two other approaches to the analysis of demographic history, a graphical method developed by Nee et al. (1995) and a statistical test developed by Tajima (1989). The results of both approaches were compatible with the hypothesis that the food-producing populations have expanded their size, whereas the other populations have not.

Concerning other issues discussed by Bandelt and Forster, we refer the reader to our original article (Watson et al. 1996), which we believe clarifies all relevant points.

GUNTER WEISS, ARNDT VON HAESELER, AND SVANTE PÄÄBO

Institute of Zoology
University of Munich
Munich

References

Nee S, Holmes EC, Rambaut A, Harvey PH (1995) Inferring population history from molecular phylogenies. Phil Trans R Soc Lond B 349:25–31
Tajima F (1989) The effect of change in population size on DNA polymorphism. Genetics 123:597–601
Watson E, Bauer K, Aman R, Weiss G, von Haeseler A, Pääbo S (1996) mtDNA sequence diversity in Africa. Am J Hum Genet 59:437–444

Address for correspondence and reprints: Dr. Svante Pääbo, Institute of Zoology, University of Munich, Luisenstrasse 14, P.O. Box 202136, D-80021 Munich, Germany. E-mail: paabo@zi.biologie.uni-muenchen.de

0002-9297/97/6104-0030$02.00

Am. J. Hum. Genet. 61:983–986, 1997

mtDNA Mutation Rates—No Need to Panic

To the Editor:

Readers of the recent paper by Howell et al. (1996) might be forgiven for thinking that, after all the controversy surrounding the reconstruction of the original mitochondrial gene trees (e.g., see Maddison 1991; Templeton 1993), the field was once again in difficulties because of (*a*) a serious underestimation of the mutation rate by a factor of almost nine and (*b*) the resulting misdating of past divergences. We believe that such an interpretation would be unduly pessimistic.

Conventional approaches have calibrated the mutation rate by reference to the divergence between humans and chimpanzees. For the phylogenetically informative first hypervariable segment of the control region (HVSI),