

Statistical cloud over African Eden

Reconstructions of the family tree of modern people by phylogenetic analysis based on extant mitochondrial DNA appear unexpectedly difficult. An African Eden seems not yet proven.

GENESIS offers many puzzles, one of which is the appearance of entire populations from thin air at key points in the narrative. It is a cause of some wonder, for example, whence the murderous Cain recruited the inhabitants of Enoch, the city he is said to have founded after his exile, when the human population at that time had fallen sharply — from four to three. Biblical imagery, of course, has a potency that logic seems powerless to confine.

Could it be the same with science? Four years ago, the late Allan Wilson and his colleagues at the University of California at Berkeley produced a family tree of human origins based on restriction enzyme maps of mitochondrial DNA (mtDNA) from more than 130 people of diverse racial type and dispersed geographical location (Cann *et al. Nature* **325**, 31–36; 1987). They concluded that all human mtDNA genomes now extant derive from a single ancestral mtDNA molecule in sub-Saharan Africa about 200,000 years ago. Because mtDNA molecules are inherited maternally, the authors wrote that the ancestral mtDNA must have been present in “one woman”, by implication the ancestor of all humanity.

Inevitably, this form of words brought biblical imagery bubbling to the surface of the public consciousness. The popular evocations of Eve in her African Eden were such that one would have been forgiven for thinking that she had been interviewed in person for her comments on the *Nature* paper.

Now, the serpents have moved in. Despite a radical revision by the Berkeley group, using sequence rather than restriction data and more rigorous hypothesis testing (Vigilant *et al., Science* **253**, 1503–1507; 1991), the idea of an African genesis has suffered what may be a mortally venomous bite. In a letter in last week's *Science*, Alan Templeton from Washington University, St Louis, demolishes the concept in four curt paragraphs (**255**, 737; 1992). In reply, the survivors of the Berkeley group (now all at Pennsylvania State University) succumb: Eden in Africa is unproven.

The words inscribed on the serpent's apple are “maximum parsimony”, the criterion for deciding among possible phylogenetic reconstructions by supposing that the best hypothesis has the fewest assumptions. Although it may be impossible to determine the degree to which features in related organisms are the conse-

quences of convergent evolution rather than indicative of shared common ancestry, the most parsimonious evolutionary track has a special place as a kind of lower bound to what natural selection has done. The maximum parsimony phylogeny is that inferred from data with the fewest discrete changes of character states in the course of evolution. Applied to nucleotide sequences, the most parsimonious solutions are those with the smallest numbers of base changes.

Maximum parsimony is popular for two reasons. First, character-state distributions among extant organisms suggest hypotheses about character-state distributions in even extinct ancestors (whence the enthusiasm of palaeontologists). Second, there are several desktop software packages that will generate phylogenies on these principles. In their original analysis, Cann *et al.* used a version of the program PAUP (Phylogenetic Analysis Using Parsimony).

The original data showed a greater degree of diversity among African mtDNAs (few though they were) than in any other subset of human mtDNAs classified by continent of origin. The PAUP tree reflected this in a bifurcation, one branch leading exclusively to African mtDNAs, the other to a larger subset from Africa and elsewhere. The implication was that modern human mtDNA had its home in Africa.

But neither PAUP nor any other parsimony routine can generate a unique most-parsimonious tree standing leaf-and-branch above the rest. There are probably about 7.13×10^{35} possible trees for a dataset of 134 separate items (in this case, mtDNA types) such as that compiled by Cann *et al.* Faced with such vertiginous choice, the authors concentrated on trees with an African branch, sensible enough given the independent evidence for an African origin, if over-welcoming of the conclusion.

Even that argument is incomplete. To specify geographical origin is to make geography a phylogenetic character. Last year, David R. Maddison showed that there are at least 10,000 phylogenetic trees more parsimonious than Cann *et al.*'s, most of which do not imply an African origin (*Syst. Zool.* **40**, 355–363; 1991). The Berkeley group (Wilson was still alive at the time) dismissed this analysis, citing an observation that parsimony methods neglecting sample size may produce the ‘wrong’ results (Wilson *et al. Syst. Zool.* **40**, 363–365, 1991). Yet the original analy-

sis could thus have been affected; the number of individuals sampled from the source population (with African mtDNAs) was much smaller than that from the supposedly derived population.

The paper from Vigilant *et al.* filled this gap, using new statistical techniques to adjust for sample size and much more detailed sampling of African populations. A PAUP analysis generated 100 equally parsimonious trees of 528 steps, themselves a selection from a number running into thousands. Nevertheless, all implied a relatively great age for African mtDNAs. Eden in African was still on the cards.

Templeton's new challenge is a double blow. First, he emphasizes that the myriad of possible trees would accommodate all possible homelands for humanity: the choice of Africa without some kind of assessment of the others is not sound. Second, he points to a procedural trap of which uninitiated users of PAUP should beware; when PAUP is used to generate phylogenies for a small number of items, the data are entered in sequence and the trees are built up as the program runs. But, Templeton says, simple sequence addition is unsuitable for large data sets (such as the sequence information compiled by Vigilant *et al.*) because the results reflect the order in which data are entered. For really big data sets, data must be entered at random and many runs performed to achieve consensus. An example shows that Africa is no more likely to have been Eden than anywhere else.

In reply (*Science* **255**, 737–739; 1992), the Penn State group concedes that its original analysis was flawed, but poses an intriguing question: what kind of data may solve the problem of human origins if an mtDNA dataset as comprehensive as that now available cannot?

The answer may simply be that the researchers were using methods inappropriate for the job. As they say, when the number of items (136 mtDNA sequences) exceeds the number of informative nucleotide positions (117), it is not surprising that the program output is so much garbage. There are 10^{267} possible trees for the latest data set: the number of maximally parsimonious trees is unknown, but is certainly “much larger than 1 billion”. Yet nobody disputes that African mtDNA is, indeed, more diverse than that from elsewhere. At least the fossils seem to indicate an African origin, even if the numbers lag behind.

Henry Gee