Support for dental HIV transmission

SIR — On the basis of a phylogenetic analysis of HIV sequences, Ou *et al.*[1] concluded that a Florida dentist infected five of his eight known HIV-1 seropositive patients. These authors used bootstrap resampling[2] to test the reliability of their finding, and found that the HIV sequences from the dentist and infected patients formed a monophyletic group in 79% of the replicates in parsimony analysis. DeBry *et al.*[3] in Scientific Correspondence questioned the conclusion of dental transmission, however, because a bootstrap analysis (based on threshold parsimony) of independently sequenced HIV variants clustered only one of the patient sequences with a dental sequence in the majority-rule consensus tree. DeBry *et al.* concluded that their analyses "...show that the available data are consistent with both the dental transmission hypothesis and the null hypothesis (the patients were independently infected from the local community) and do not distinguish between the two." But both studies[1,3] used an analysis of the bootstrap results that may not be the most appropriate method for this case. We have reanalysed the two datasets, as well as sequences from new patients and new local controls, and find strong support for trees consistent with HIV transmission between the dentist and six of ten of his seropositive patients.

Both Ou *et al.* and DeBry *et al.* used the standard systematic practice of summarizing the bootstrap results by showing the percentage of time particular clades were found in a majority-rule consensus tree of the replicates[2]. Both studies considered patient HIV sequences that clustered more closely with the dental sequences than with any local controls to be consistent with the dental transmission hypothesis. However, given that the question of dental transmission is asked separately for each patient, the majority-rule consensus tree does not summarize the appropriate information. Individual patient sequences may be consistent with the dental transmission hypothesis in every bootstrap replicate, and yet not be present in the majority-rule consensus tree. To demonstrate this fact, consider an hypothetical example (shown in the figure) in which the majority-rule consensus tree provides no support for any of the patient sequences to group with the dental sequence. However, all of the trees are consistent with the dental transmission hypothesis for the sequence from patient A (because the sequences from patient A are always more closely related to those from the dentist than to any of the local controls). Clearly, then, the majority-rule consensus tree is not adequate for identifying the patients who may have been infected by the dentist. Whether or not a tree is consistent with the dental hypothesis for any given patient should be dependent only on the placement of the patient sequence with respect to the dental sequences and local controls, not other patients.

An alternative method of summarizing the bootstrap results poses the question of the dental transmission hypothesis separately for each patient. Thus, for the example above, the support is 100% for patient A, 67% for patients B and C, and 0% for patient D. This is a more appropriate way of summarizing the results, because there is no *a priori* way to hypothesize the extent of the 'dental clade' and the dental transmission hypothesis should be tested for each patient separately. The table shows the results of such an analysis, based on reanalyses of refs 1 and 3, and a new dataset collected by the Centers for Disease Control and Prevention (CDC; Marcia Kalish, personal communication). The latter dataset is the same as that analysed by Ou *et al.*, but includes sequences from three new patients as well as five local controls from the same county as the dental practice. The results in the table show that all the datasets support the conclusions of Ou *et al.* even more strongly than was originally suggested, with the exception of patient G in the DeBry *et al.* dataset (which shows lower support than the 79% bootstrap proportion suggested by Ou *et al.*). The lower consensus bootstrap values reported by DeBry *et al.* appear to be largely the result of the more ambiguous placement of patient G among their trees; three of the purportedly infected patients (A, B, E) actually cluster more strongly with the dentist in the DeBry *et al.* dataset than in the original study. Ou *et al.* concluded that their tree was consistent with the dental transmission hypothesis for patients A, B, C, E, G; at best, the DeBry *et al.* data cast doubt only on patient G. Our analyses of the new CDC data support the original conclusion, and also include patient I in the dental clade (in agreement with a separate analysis by CDC[4]).

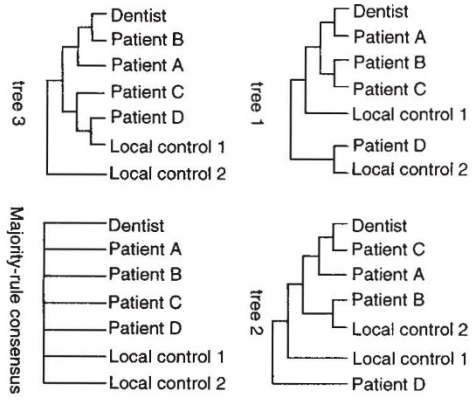
As noted by Ou *et al.* and DeBry *et al.*, the interpretation of significance levels for bootstrap proportions is a matter of some debate[5-7]. However, it is clear that bootstrap proportions (as usually formulated) considerably underestimate phylogenetic accuracy under most conditions[5]. We have shown that the probability of resolving the dental clade (the dentist plus patients A, B, C, E and G from the original study) is greater than 99% for various parsimony and distance methods, given the observed divergence and substitution frequencies[8]. Therefore, the individual bootstrap values shown in the table are also underestimates of phylogenetic accuracy, especially given the problem of multiple comparisons

INDIVIDUAL BOOTSTRAP SUPPORT (BASED ON 1,000 REPLICATES) FOR THE DENTAL TRANSMISSION HYPOTHESIS FOR SEQUENCES FROM 10 PATIENTS (A–J).

Patient	Other HIV risk factors*	Data set: Ou *et al.* BP (*n*)	Data set: DeBry *et al.* BP (*n*)	Data set: New CDC data BP (*n*)
A	No	83–99 (2)	89–98 (4)	91–100 (2)
B	No	78–79 (2)	89–90 (3)	84–86 (2)
C	Unconfirmed	98 (2)	88 (4)	99 (2)
D	Yes	1 (2)	Not examined	1 (2)
E	No	79 (2)	94–95 (4)	86 (2)
F	Yes	0 (2)	0 (2)	0 (2)
G	No	94–96 (2)	63–65 (2)	96–98 (2)
H	Yes	Not examined	Not examined	8 (1)
I	No	Not examined	Not examined	89 (1)
J	Yes	Not examined	Not examined	0–13 (2)

* Sources: refs 4, 10; Carol Ciesielski, Centers for Disease Control and Prevention, personal communication.

In each study, multiple sequences were analysed from most individuals; the range of bootstrap proportions (BP, given as a percentage) among sequences is shown for each individual, with the number of sequences examined shown in parentheses. The Ou *et al.* dataset[1] includes six local control sequences (the local control consensus sequence was not used here because it does not represent an individual viral sequence), two dental sequences and an outgroup. The DeBry *et al.* dataset[3] includes 13 local control sequences (from 10 individuals) and six dental sequences. The new CDC dataset includes all the earlier data from ref. 1 plus five new local control sequences (see text). Parsimony analyses were based on unordered characters with all state changes weighted equally[9] (information on analysis based on alternative weighting and ordering assumptions, as well as details of the analyses and searching methods, is available on request from D. M. H.).



across patients.

Our analyses agree with the conclusions of Ou *et al.*: there is significant support for phylogenetic trees that are consistent with the dental transmission hypothesis for patients A, B, C, E, G (and I), and significant evidence that patients D and F (as well as H and J) were infected from another source. These results are consistent with what is known about the existence of other risk factors for HIV among the 10 patients. DeBry *et al.* may turn out to be correct that regions other than the highly variable C2-V3 region will be even more informative about HIV transmission, but the C2-V3 data are nonetheless highly informative in this case.

David M. Hillis
John P. Huelsenbeck
Department of Zoology,
University of Texas,
Austin, Texas 78712, USA.

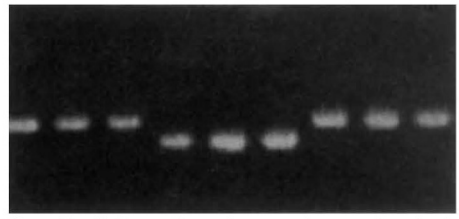
1. Ou, C.-Y. *et al. Science* **256**, 1165-1171 (1992).
2. Felsenstein, J. *Evolution* **39**, 783-791 (1985).
3. DeBry, R.W. *et al. Nature* **361**, 691 (1993).
4. CDC *Mort. Morbid. Wkly Rep.* **42**, 329-331, 337 (1993).
5. Hillis, D. M. & Bull, J.J. *Syst. Biol.* **42**, 182-192 (1993).
6. Felsenstein, J. & Kishino, H. *Syst. Biol.* **42**, 193-200 (1993).
7. Zharkikh, A. & Li, W.-H. *Molec. Biol. Evol.* **9**, 1119-1147 (1992).
8. Hillis, D.M. *et al. Science* (in the press).
9. Swofford, D. L. *PAUP 3.1: Phylogenetic Analysis Using Parsimony* (Illinois Biological Survey, Urbana, 1993).
10. Ciesielski, C. *et al. Ann. intern. Med.* **116**, 798-805 (1992).

■ The HIV sequence data reported in ref. 3 are now publicly available. Genbank accession numbers are U06872 – U06919.

DNA from ancient Easter Islanders

SIR — Easter Island is one of the most remote inhabited places on Earth and there is considerable speculation about the origins of its prehistoric inhabitants. Early accounts of Easter Island are sketchy, but the population diverged from the pattern of Polynesian settlement and culture as a result of isolation and ecological degradation. The island was almost completely depopulated after European contact, and the present population of about 2,000 individuals is mostly Chilean. We report genetic evidence for the Polynesian origin of prehistoric Easter Islanders based on amplification and sequencing of mitochondrial DNA (mtDNA) from ancient human bones.

We analysed DNA from prehistoric skeletal material recovered from two discrete archaeological contexts, Ahu Tepeu on the west coast (1100–1680 AD), and Ahu Vinapu on the southern side of the island (1680–1868 AD). The bones were excavated by Thor Heyerdahl during his Norwegian Archaeological Expedition of 1955–56 (ref. 1) and kept in the Museo Nacional de Historia Natural, Santiago de Chile. The relatedness of the individuals could not be inferred from the spatial distribution of graves as the skeletons were from secondary burials.



PCR products of a segment of the mtDNA COII/tRNA^{lys} region amplified from ancient human bones. Six outside bands, PCR products generated from archaeological bones of European origin; middle three lanes, from DNA of prehistoric Easter Islanders exhibiting the 9-bp-length mutation frequently found in Asians. Bone fragments consisted of wedges of right femur, 2-3 g in weight, carefully removed with a hacksaw. The pieces were cleaned and pulverized, and DNA was extracted from 0.7 g of bone powder, as described previously, with care to avoid contamination by modern DNA[2]. Extractions were carried out in parallel with blanks containing no tissue, as well as control extractions of bone samples from different geographical origins. Approximately 100 μl DNA extract were obtained from 0.7 g bone powder. Aliquots of 1.5 μl, representing 1–2% of the DNA extracted, were amplified by the method of Perkin-Elmer-Cetus in 25-μl reaction volumes containing 2 units of *Taq* polymerase and 4 μg BSA, and using primers A and B[3] which define a 120-bp fragment of the human mtDNA reference sequence containing two 9-bp tandem repeats. The amplified band is 111 bp in length in individuals with the 9-bp deletion. Amplifications consisted of 35 cycles of PCR (94 °C for 1 min, 55 °C for 1 min and 72 °C for 1 min). Electrophoresis of the PCR products was on 3% NuSieve/1% Sigma type 1 agarose gel, followed by ethidium bromide staining.

We extracted DNA from bone samples of 12 adult individuals as previously described[2] and used it for polymerase chain reaction (PCR) amplification with primers specific for two anthropologically informative regions of the human mitochondrial genome. The first informative mtDNA fragment is in a small intergenic region between the cytochrome oxidase II (COII) and the lysyl transfer RNA genes that has two tandemly repeated copies of the 9-base-pair (bp) sequence CCCCCTCTA (ref. 3). A length mutation consisting of a deletion of one of the two repeats is often found (5–40%) in present-day individuals of Asian origin, including Pacific islanders, and has reached fixation in some Polynesian islands[4]. All the prehistoric Easter Islanders were found to carry the 9-bp deletion, detected by electrophoresis of the PCR fragments through high-percentage agarose gels (see figure).

The second sequence analysed is a 228-bp fragment of the hypervariable control region of mtDNA, known to contain most of the sequence variability in the human mitochondrial genome. PCR products were sequenced directly after amplification as described previously[2], and the sequences read between mtDNA bases 16,215 and 16,410 of the published human mtDNA reference sequence[5]. All the ancient Easter Island DNA samples had identical base substitutions, differing from the reference sequence at positions 16,217 (T→C), 16,247 (A→G) and 16,261 (C→T). One individual had an additional transition at position 16,271 (T→C) and two other individuals at position 16,292 (C→T).

The three substitutions at mtDNA positions 16,217, 16,247 and 16,261, in conjunction with the COII/tRNA^{lys} 9-bp deletion, are typical of Polynesians. We previously observed these polymorphisms in most present-day Polynesians sampled (unpublished observations), and in prehistoric Polynesians from the Chatham Islands (near New Zealand) and Hawaii[6]. The survival of only one principal mitochondrial DNA lineage throughout remote Oceania is presumably due to population bottlenecks during the human colonization of the region. We have demonstrated that the prehistoric settlers of Easter Island derived from this identical lineage, characterized by the 9-bp deletion and the three base substitutions at mtDNA positions 16,217, 16,247 and 16,261. Our results confirm the Polynesian affinities of the original settlers and support the view that the unique culture that developed on the island resulted from the environmental conditions and several centuries of isolation.

Concerning the question of a possible settlement from South America, recent surveys of the hypervariable region of mtDNA of native Americans have shown much greater mtDNA heterogeneity in America than in Polynesia, reflecting multiple waves of migration and greater depth of prehistory, whereas the Polynesian type that we observed in Easter Island has not been detected[7–10]. Although the 9-bp deletion exists throughout the Americas, Asia and the Pacific, often in association

1. Heyerdahl, T. & Ferdon, E. N. Jr *Monogr. Sch. Am. Res. and Kon-Tiki Mus.* No. 24, pt 2 (Santa Fe, 1965).
2. Hagelberg, E. & Clegg, J. B. *Proc. R. Soc.* B**244**, 45–50 (1991).
3. Wrischnick, L. A. *et al. Nucleic Acids Res.* **15**, 529–542 (1987).
4. Hertzberg, M., Mickleson, K. P. N., Serjeantson, S. W., Prior, J. F. & Trent, R. J. *Am. J. hum. Genet.* **44**, 504–510 (1989).
5. Anderson, S. *et al. Nature* **290**, 457–465 (1981).
6. Hagelberg, E. & Clegg, J. B. *Proc. R. Soc.* B**252**, 163–170 (1993).
7. Ward, R. H., Frazier, B. L., Dew-Jager, K. & Pääbo, S. *Proc. natn. Acad. Sci. U.S.A.* **88**, 8720–8724 (1991).
8. Ginther, C. *et al.* in *DNA Fingerprinting: State of the Science* (eds Pena, S. D. J., Chakraborty, R. & Jeffreys, A. J.) 211–219 (Birkhäuser, Basel, 1993).
9. Horai, H. *et al. Molec. Biol. Evol.* **10**, 23–47 (1993).
10. Torroni, A. *et al. Am. J. hum. Genet.* **53**, 563–590 (1993).