JOURNAL OF VIROLOGY, Apr. 1995, p. 2351–2356 Vol. 69, No. 4
0022-538X/95/$04.00+0

Intraspecific Phylogenetics: Support for Dental Transmission of Human Immunodeficiency Virus

KEITH A. CRANDALL*

Department of Zoology, University of Texas at Austin, Austin, Texas 78712-1064

Received 25 August 1994/Accepted 10 January 1995

A new method to estimate within-species gene genealogies was used to establish linkages among individuals associated with the Florida dental human immunodeficiency virus transmission case. Phylogenetic relationships were estimated from 103 nucleotide sequences from the V3 region of the *env* gene representing the Florida dentist, eight of his seropositive patients, and many local controls. The cladogram estimation procedure supports linkages among individuals within the previously described dental clade, whereas local controls and other patients form independent networks or are outliers in the main network, indicating more distant evolutionary relationships. A nested statistical analysis also indicates significant cohesion of the dental clade group.

In a recent correspondence, DeBry et al. (5) commented on the need for an appropriate framework within which one can test the hypothesis of human immunodeficiency virus (HIV) transmission. DeBry et al. (5) criticized the original analysis of an HIV transmission case in Florida for using an inappropriate model of evolution (and phylogeny reconstruction technique) and inadequate sampling of local controls. They reexamined the conclusion reached by Ou et al. (12) that a Florida dentist infected five of his eight HIV type 1 (HIV-1)-seropositive patients using an alternative model of evolution and additional controls. The phylogenetic analysis by DeBry et al. (5) indicated no resolution between the null hypothesis of independent acquisition of HIV-1 versus the alternative of infection via the Florida dentist.

Both the analyses by Ou et al. (12) and DeBry et al. (5) were based on variations of the maximum parsimony procedure for phylogeny reconstruction (18). This procedure establishes relationships based on shared derived characters and was developed for estimating higher-level systematic relationships. Unfortunately, many of the assumptions of the parsimony procedure do not hold for population genetic data. For example, population genetic data are not necessarily strictly bifurcating. They can include reticulations which result in lack of resolution when using a maximum parsimony procedure. Population genetic data, including HIV-1 sequences, can be subject to recombination (14). Traditional phylogeny reconstruction techniques do not take into consideration the possibility of recombination and therefore will give erroneous results if recombination has occurred within the aligned sequences. Finally, individual sequences are typically differentiated by few nucleotide substitutions. Because maximum parsimony requires characters to be shared among variant sequences in order to be informative, the data set available for phylogeny reconstruction is very limited with closely related individuals, resulting in a lack of resolution of phylogenetic relationships.

A method that takes into account these special considerations for within-species cladogram estimation has recently been developed. The intraspecific cladogram estimation procedure of Templeton et al. (21) was designed to have its greatest statistical power when the number of differences between any pair of sequences is few, as is typically the case with intraspecific data sets. Traditional methods (e.g., parsimony) were developed to estimate phylogenetic relationships over greater time spans, in which the number of variable sites is large relative to the number of shared sites. Huelsenbeck and Hillis (11) have shown that these traditional methods (including weighted and unweighted parsimony, UPGMA, and neighbor joining) all perform poorly when few (<100 bp) differences exist among sequences. Crandall (1), however, has shown that the method of Templeton et al. (21) performs well and outperforms bootstrapping with maximum parsimony, even at higher levels of divergence, when few characters are available for analysis. Crandall et al. (4) have detailed the advantages to the intraspecific cladogram estimation procedure relative to traditional methods, including the ability to account for recombination within a data set, accommodation of multifurcating relationships, and the acceptance of extant ancestral sequences. In addition, this procedure provides a statistical framework for testing hypotheses of associations for both continuous and categorical data (20, 22).

The purpose of this paper is to present this alternative procedure for estimating intraspecific gene genealogies and demonstrate its utility in reconstructing relationships of closely related sequences and its facility for hypothesis testing based on the cladogram structure. I have reanalyzed the HIV sequences of Ou et al. (12) with the addition of new sequences from DeBry et al. (5) using this procedure. The resulting cladogram indicates statistical support for the dental clade as originally concluded by Ou et al. (12). Furthermore, a nested statistical analysis gives further support for the dental clade.

MATERIALS AND METHODS

HIV sequences. A total of 103 HIV nucleotide sequences from the V3 region of the *env* gene were obtained from GenBank for phylogenetic analysis. These sequences include 86 variant sequences listed by Ou et al. (12) and 17 additional local controls offered by DeBry et al. (5). The sequences were aligned with CLUSTAL V. Because of the high levels of variability within the V3 region, positional homology was suspect in a number of aligned positions. These positions were eliminated from analysis, resulting in total of 240 nucleotide positions from the V3 region. The alignment used in this analysis is available from the author upon request. This represents the largest data set used to analyze the Florida dental transmission hypothesis to date. Because of the resolving power of the method of Templeton et al. (21), multiple sequences per individual can be

* Mailing address: Department of Zoology, University of Texas, Austin, TX 78712-1064. Phone: (512) 471-5661. Fax: (512) 471-9651. Electronic mail address: crandall@phylo.zo.utexas.edu.

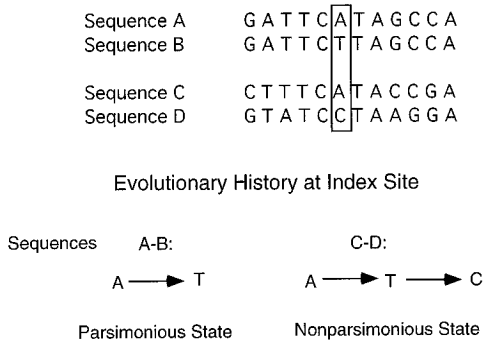

FIG. 1. A hypothetical example demonstrating the parsimony concept used by the cladogram estimation procedure. Given a nucleotide difference between sequences, noted in the box, the probability that a multiple mutation (the nonparsimonious state) has occurred is greater for sequence pairs that share fewer sites, pair C and D, than for those that share more sites, pair A and B.

used without a loss of resolution. Thus, all complete sequences available for each patient and the dentist were used in this study.

Cladogram estimation. The cladogram estimation procedure of Templeton et al. (21) estimates the set of cladograms that portray connections among variant sequences where these connections have a high cumulative probability ($\geq$0.95) of being true, thereby constructing a plausible set of alternative cladograms. This estimation algorithm is based on a parsimony criterion with a statistical procedure to evaluate the limits of the parsimony assumption. Under consideration by the method is the number of mutational events causing a single nucleotide difference between two variant sequences, with a difference caused by a single mutational event being the parsimonious state and a difference cause by multiple mutational events being the nonparsimonious state (Fig. 1). Thus, the term parsimony refers to the minimum number of mutations separating two individual sequences rather than a global minimum tree length based on shared derived characters.

The statistical evaluation of the limits of parsimony asks the following question: given that two variant sequences differ at j nucleotide positions and share m positions, what is the probability that multiple mutational events have occurred within one or more of the j sites differentiating the two sequences? If the probability of multiple mutations is high (i.e., $P > 0.95$), then the parsimony criterion is invalid. Templeton et al. (21) have developed a population genetic model to evaluate this probability. They first condition their model on an index site, the oldest polymorphic site. Then, the total probability that two sequences differ at the index site, differ at $j - 1$ other sites, and share in common m sites is approximated by

$$L(j,m) = (2q_1)^{j-1}(1-q_1)^{2m+1}\left(\frac{1-q_1}{br}\right)\left[\frac{2-q_1(br+1)}{br}\right]^{j-1}\left[1-2q_1\left(\frac{1-q_1}{br}\right)\right] \quad (1)$$

where q_1 is the probability of a nucleotide change within a block of r nucleotides in the two variant sequences since their respective lineages diverged at the index site and b is a mutational bias factor such that $b = 3$ for no bias, indicating three alternative states for a nucleotide substitution, and $b = 1$ for extreme bias, indicating only a single alternative state (21). This model is generalized to be compatible with both restriction site and nucleotide sequence data. With restriction site data, r is the length of the endonuclease recognition sequence, and with sequence data, r is simply one. One can then use equation 1 as a posterior probability distribution of the data given q_1 and estimate q_1 using the Pitman estimator of q_1 and a uniform prior of (0,1):

$$\hat{q}_1 = \frac{\int_0^1 q_1 L(j,m)\,dq_1}{\int_0^1 L(j,m)\,dq_1} \quad (2)$$

When j is >1, deviations from parsimony may still occur at these other sites. One can then estimate q_2 by replacing j with $j - 1$ in equation 2. This iterative procedure results in a set of estimators (q_1 to q_j). Although the relative age order of alleles was used to develop this model, the iterative estimates depend only upon the observable values of j and m. The probability that two sequences

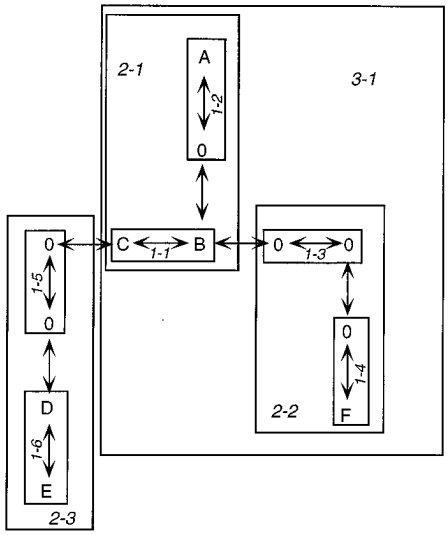

FIG. 2. A demonstration of the nesting procedure with missing intermediates. The nesting begins with the tip variants, those with only a single mutation connection, and proceeds by nesting n-step clades separated by $n + 1$ mutational steps in the interior direction. Each variant sequence forms a zero-step clade. Six one-step clades are formed by joining the zero-step clades separated by one mutational difference (each arrow represents a single mutational step). The procedure is repeated recursively until one reaches the level at which the next round of nesting would unite the entire cladogram into a single category.

differing by j sites and sharing m sites have a parsimonious relationship (i.e., no unobserved mutations at any site) is then estimated by

$$\hat{P}_j = \prod_{i=1}^{j}(1-q_i) \quad (3)$$

To implement this method, pairwise distances from the set of aligned HIV sequences were calculated with PAUP's "show distance matrix" option (17). Equation 3 was then used to calculate the probability of parsimonious connections between variant sequences given the minimum number of differences between pairs, j, and the number of shared sites (with sequence data, $m = n - j$, where n is the total number of nucleotides). When parsimonious connections could not be justified with equation 3, nonparsimonious connections of one additional mutation were considered on the basis of the following relationship, which calculates the probability that y or fewer of the x site mutations are not parsimonious (21):

$$\sum_{i=0}^{y}\sum_{I}\prod_{k=1}^{i} q_{j(k)} \prod_{k=i+1}^{x}[1-q_{j(k)}] \quad (4)$$

where I refers to the set of all permutations of the x age ranks and $j(k)$ refers to the kth permutation in the set I (21). Equations 3 and 4 were calculated with a computer package in Mathematica (23).

Nested analyses. After cladogram estimation, statistical hypothesis testing was performed using a nested analysis for categorical data (22). This procedure utilizes the fact that the evolutionary history of a gene region forms a nested hierarchy which can be exploited for hypothesis testing. The nested hierarchy represents the relative age order of groups of variant sequences within the cladogram. For example, five-step clades (those variants separated by five nucleotide substitutions) are older than the average four-step clades nested within them. Thus, the nested series of clades corresponds to a nested series of relative average ages.

The nesting procedure consists of nesting n-step clades within $n + 1$-step clades, where n refers to the number of transitional steps used to define the clade. By definition, each variant sequence is a zero-step clade. The $n + 1$-step clades are formed by the union of all n-step clades that can be joined together by $n + 1$ mutational steps. The nesting procedure begins with tip clades, i.e., those clades with a single mutational connection, and proceeds to interior clades (Fig. 2). In previous analyses based on restriction site data, missing intermediates were ignored in the nesting procedure, as they were inconsequential to these analyses (20, 22). However, with nucleotide sequence data there are many more missing intermediates, as variant sequences are typically differentiated by more than a single nucleotide substitution. These missing intermediates must be considered

TABLE 1. Connection probabilities for parsimonious and parsimonious plus one connections

No. of mutational differences (j)	Probability of parsimony[a]	Plus one probability[b]
1	0.998	—
2	0.994	—
3	0.988	—
4	0.979	—
5	0.969	—
6	0.957	—
7	0.943	0.999
8	0.927	0.998
9	0.910	0.996
10	0.891	0.995
11	0.870	0.992
12	0.848	0.989
13	0.824	0.985
14	0.800	0.980
15	0.774	0.975
16	0.748	0.968
17	0.721	0.960
18	0.693	0.951
19	0.665	0.940

[a] The probability of parsimony was calculated with equation 3. Probabilities were calculated with j (the number of nucleotides by which two variant sequences differ), m (the number of nucleotides that two variant sequences share) = 240 (the total number of nucleotides sequenced per individual) − j, u (the upper bound of the uniform prior) = 1, r (length of the recognition sequence) = 1, and b (transition bias factor) = 1 (HIV data are known to have an extreme substitutional bias [10]).

[b] The plus one probability, that is, the probability of a parsimonious connection or a connection of one additional mutational step, was calculated with equation 4 with the same parameters as described above. —, not applicable.

in the nesting procedure to ensure overall consistency. Because many of the variant sequences are differentiated by a number of nucleotide substitutions, the nesting procedure results in a number of empty nests. These empty nests are required for the consistency of the nesting procedure to form higher-level nests but can be ignored during subsequent statistical analyses since they contain no observations. The nesting procedure results in hierarchical nests with nesting level directly correlated to evolutionary time, i.e., the lower the nesting level is, the more recent are the evolutionary events relative to higher nesting levels.

Once established, the nesting design is then used to test for significant associations of phenotype and genotype by either a nested analysis of variance for continuous data (20) or a permutation chi-square contingency test for categorical data (13, 22). In this case, there is the categorical "phenotype" of being classified within the Florida dental clade or not. Thus, I first tested the question put forth in the original paper by Ou et al. (12) and later asked by DeBry et al. (5): is there evidence for a dental clade? Put into a hypothesis-testing framework, the comparison is the null hypothesis of independent acquisition versus the alternative hypothesis of acquisition of HIV from the dentist. These hypotheses give very different predictions relating to the nesting level at which heterogeneity in the bivariate label of dental clade-not dental clade will occur. The null hypothesis is rejected if all the patients presumed to have been infected by the dentist nest together before significant heterogeneity exists between local controls or non-dentist-infected patients and dentist-infected patients. This hypothesis was tested to demonstrate this hypothesis-testing framework relative to that used in these previous papers. However, Hillis and Huelsenbeck (9) correctly point out that the appropriate test is for the association of each patient in turn with the dentist, since there is no a priori way to hypothesize the extent of dental clade. Therefore, the transmission hypothesis was tested for each patient within the dental clade. Once again, the null hypothesis of no transmission is rejected if all the sequences from the patient under consideration nest with the dentist sequences before they nest with any control sequences.

RESULTS

Cladogram estimation. The application of the above-described intraspecific cladogram estimation procedure to the HIV-1 data results in the probability of parsimonious connections being supported ($P \geq 0.95$) for variant sequences that differ by six or fewer nucleotide substitutions (Table 1). All the variant sequences previously reported as members of the dental clade are connected into a single parsimony network except for sequences from patient G, whose nearest sequence (sequence PGD) to the main network is seven mutational steps from a patient C sequence (sequence C12) (Fig. 3). Parsimonious connections plus one additional mutation are supported for variant sequences that differ by 18 or fewer mutations (Table 1). This allows the connection of the patient G network to the main network, as sequence PGD is seven steps from C12 and eight steps from variant sequences CD20 and C6. Some local controls are also connected to the main network with this relaxed criterion. The closest control sequence (sequence 6911) is nine steps from a patient E sequence (sequence PE4). Network I shows the main network of variant sequence connections, with patients designated with a P followed by the patient letter. Sequences from the dentist are prefixed with a D, and controls are prefixed by a Q or designated by the last four digits in the accession number (DeBry sequences). Sequences from patients D, F, and H form their own independent networks (Fig. 3, networks II, III, and IV) whose connection to the main network cannot be justified statistically. Likewise, some local control sequences form independent networks (Fig. 3, networks V and VI). The following variant sequences (all local controls) could not be connected to any network because the minimum number of mutational steps to any other variant sequence was greater than 18: DSPID, Q1015, 6903, 6905, 6906, 6909, 6912, 6913, and 6915.

Nested analyses. To apply the nested statistical analysis, I used the maximum parsimony network (Fig. 3, network I). Although there is a degree of uncertainty of these connections (i.e., closed loops within the network), the nested statistical analysis is robust to even large levels of cladogram uncertainty (22). To test the hypothesis of heterogeneity within the dental clade, each variant sequence is assigned a value of one if it is an a priori member of this clade and zero if not. The permutation test identifies significant heterogeneity among clades at the six-step level ($\chi^2 = 60.1$; $P < 0.0001$), which nests the dental clade plus a single local control sequence (6908) with the remaining local controls (Fig. 4). An alternative connection within the plausible set also exists for variant 6908, 16 steps from variant 6911. This connection would eliminate any control sequence from nesting with any dentist or dental clade patient sequence before the entire dental clade is nested together. The nesting design reveals that all the variant sequences from the dental clade are nested together from one- to four-step nests before the inclusion of local control 6908, indicating significant separation in relative age from the local controls. Furthermore, all individual patient sequences nest with at least one dentist sequence before nesting with any local controls, supporting the dental transmission hypothesis for each patient in turn. This is because the only local control that nests with the dental clade (6908) is connected distantly through a dentist variant (D7), not a patient sequence.

DISCUSSION

This analysis shows strong support for the previously hypothesized dental clade in multiple ways. First, the cladogram estimation procedure itself interconnects all members of the dental clade into a single network with a high degree of certainty (all parsimonious connections supported at $P \geq 0.95$ except for patient G, whose parsimonious connection has a probability of 0.943). When parsimony plus one additional mutation connections are considered, the patient G network is connected exclusively to the dental clade. The local controls included in the main network form a subclade with only a single parsimonious connection to the dental clade, with the

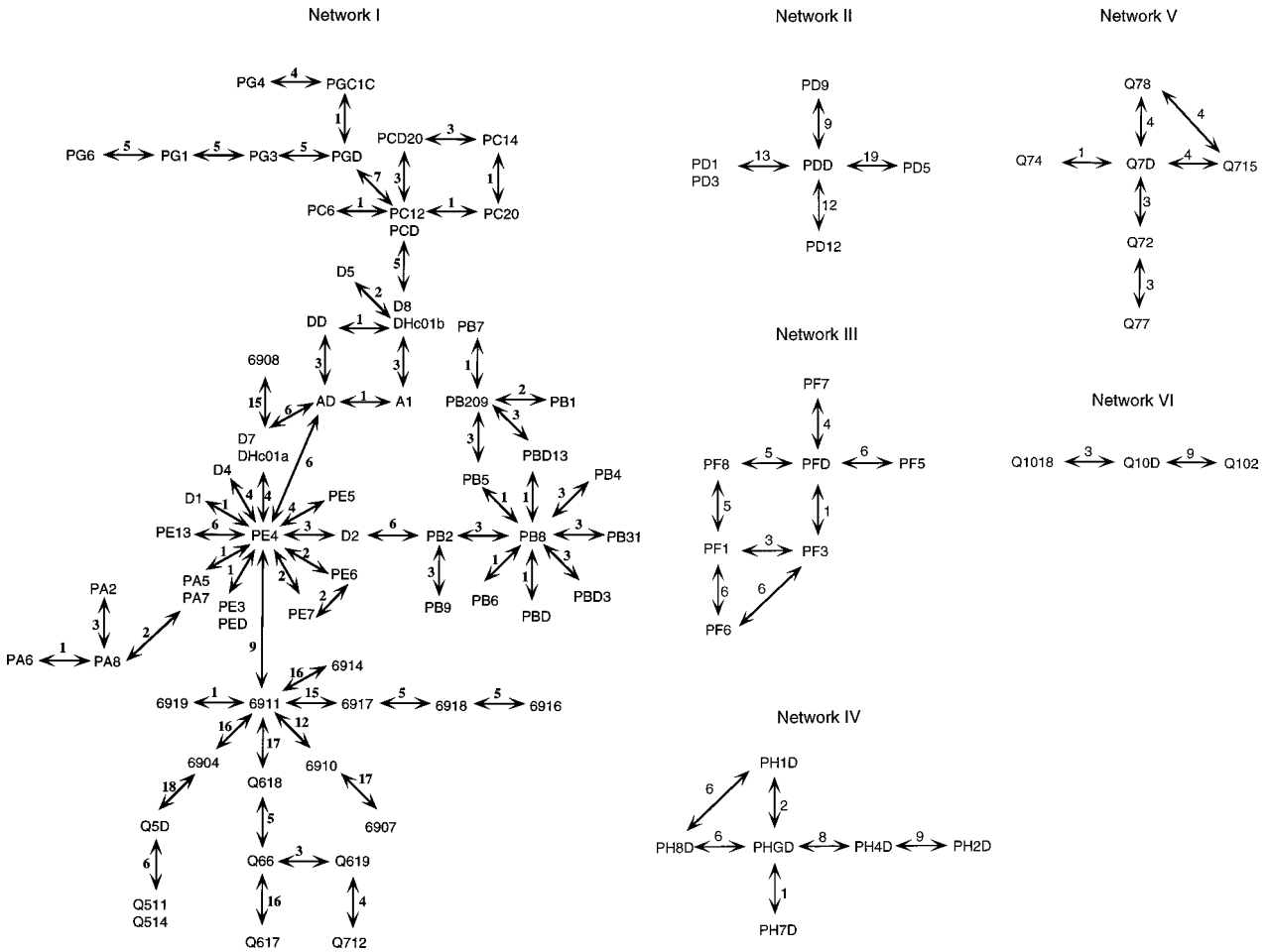

FIG. 3. Networks connecting the HIV variant sequences by the method of Templeton et al. (21). For legibility, only parsimonious connections are shown. These networks are based on HIV sequences from a dentist (D), eight of his HIV-seropositive patients (PA to PH), and individuals from the local population (designated by the prefix Q or by the last four digits of the accession number). The number of nucleotide substitutions separating variant sequences is given on the connecting arrows. Network I forms the 95% plausible set of cladograms which include the dentist sequences and all members of the dental clade as well as some local controls. Other patients and local controls form independent networks which lie statistically outside the dental clade.

exception of sequence 6908. The estimation procedure indicates statistical support for the parsimonious connection of this variant sequence to the dentist sequence D7 and the nonparsimonious connection to local control sequence 6911. Other local controls and patients not suspected of acquiring HIV through contact with the dentist form independent networks whose connection to the main dental clade network cannot be statistically justified. The second line of evidence supporting the dental clade and the dental HIV transmission hypothesis comes from the nested analysis. This analysis indicates that each dental clade patient is nested with sequences from the dentist before nesting with any local control. Furthermore, the entire dental clade nests together before the nesting of local controls except for local control sequence 6908.

While the dental clade and the hypotheses of transmission from dentist to various patients are well supported by this and previous analyses (9, 12), the issue of appropriate local controls has been raised as well (5). This analysis utilizes the largest data set of local controls available to date and still supports the conclusions reached previously. The structure of mutational changes between variant sequences from the dental clade is different from that of the local controls. Notice that within the dental clade, all connections are made with seven or fewer mutational steps (Fig. 3). Connections among the local controls, on the other hand, have a much broader range of mutational substitutions connecting variant sequences, from a single mutational step to 16 mutational steps within network I (Fig. 3) to a minimum of 27 mutational steps connecting sequence 6912 to any other variant sequence in the data set. This increase in genetic variation among local controls relative to the dental clade is also suggestive of the adequacy of the local controls and support for the dental clade. Future work will attempt to quantify levels of variation within data sets of known transmission relative to data sets of independent infection. Using the cladogram estimation procedure and nested analysis framework presented here, one is able to partition effects of population structure from population history (19). This approach will provide quantitative guidelines for assessing the adequacy of local controls in testing hypotheses of transmission.

The cladogram estimation procedure utilized in this study has a number of advantages over the parsimony procedures used in previous analyses. The main advantage of this method over traditional approaches to phylogeny reconstruction (e.g., parsimony or neighbor joining) is the greater resolution achieved by the method of Templeton et al. (21). Figure 3

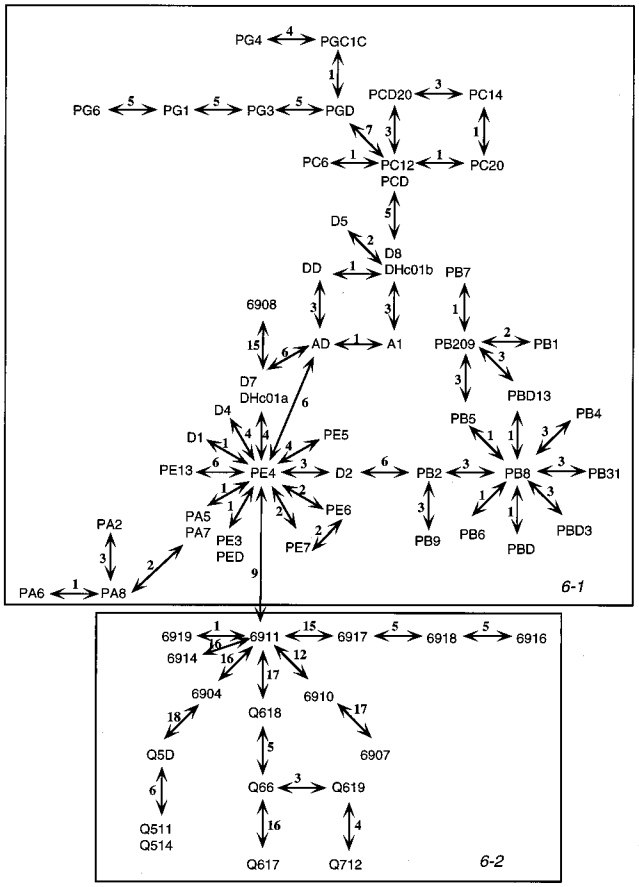

FIG. 4. Nesting design as calculated by the procedures outlined by Templeton and Sing (22). Pairs of numbers in italics refer to the nesting level, with the first number indicating the number of nucleotide substitutions between subclades and the second number indexing the number of clades at that level. Significant heterogeneity in the dental clade–non-dental clade variable is found only at that six-step level. At this point, only a single local control is nested within the dental clade nest (*6-1*), while every sequence from patients previously associated with the dental clade is nested within this category.

illustrates that many of the sequences in this study (especially sequences within individuals and the dental clade itself) are differentiated by few nucleotide substitutions. This cladogram estimation procedure has its statistical power when there are few substitutions differentiating variant sequences, the opposite of traditional phylogeny reconstruction techniques (1). Note that all the connections made in network 1 (Fig. 3) are supported by the statistical criteria of equation 3 or 4 at the $P \geq 0.95$ level. The application of bootstrapping with maximum parsimony in previous analyses has not been able to achieve the resolution shown here (5, 9, 12). This supports the conclusions drawn by a previous analysis of viral sequences of known phylogenetic relationship that the method of Templeton et al. (21) establishes more connections with greater confidence than does bootstrapping with maximum parsimony when few characters are available for analysis (1). Thus, the conclusion here is not that different methods yield different results based on similar sets of data; rather, it is that different methods give different degrees of resolution of the underlying genealogical relationships. The method of Templeton et al. (21), designed specifically for the intraspecific level, gives the greatest resolution when few characters differentiate variant sequences.

While no recombination was detected in these sequences, recombination has been documented in HIV sequences (14) as well as other gene regions (21). Thus, the ability for this algorithm to detect products of recombination is especially important in population genetic analyses. If recombination is present, further ambiguity in phylogenetic relationships will result if the region of recombination cannot be identified accurately (21).

The method of Templeton et al. (21) provides a quantitative assessment of cladogram uncertainty as part of the estimation procedure. Unlike traditional parsimony analysis, which often results in numerous most parsimonious trees when applied to the population genetic level, this estimation procedure provides probabilities to assess the reliability of each connection. These resulting probabilities can then be used to define a plausible set of alternative relationships with a cumulative probability of ≥ 0.95, thus providing a quantitative assessment of deviations from parsimony. They can also be used to assess the relative weights given to alternative connections when performing hypothesis testing. Additionally, these probabilities can be refined on the basis of the topological status of variant sequences and their frequencies by using arguments from population genetic theory (3, 4).

Finally, the resulting cladogram provides a powerful framework for testing hypotheses of associations based on evolutionary relationship. In traditional parsimony analyses, often a bootstrap majority-rule consensus tree is used to assess the reliability of a clade (6, 8, 24, 25). However, Hillis and Huelsenbeck (9) have pointed out the inadequacies of the majority-rule consensus tree relative to this hypothesis-testing framework. The nested analysis presented here combined with the cladogram estimation procedure provides a clear and powerful hypothesis-testing framework. Unlike typical comparative methods which rely on a single estimate of phylogenetic relationships (7), the nested analysis can accommodate ambiguities in the estimated relationships (22). Thus, hypotheses can be tested over a class of plausible phylogenies. The cladogram estimation procedure and associated nested analysis have been used to test a variety of hypotheses, from interspecies viral transmission of primate T-cell leukemia/lymphoma viruses (2) to the identification of causal mutations associated with risk factors for coronary artery disease (15, 16). Because of the statistical power at the population genetic level and the ability to account for population genetic phenomena (e.g., recombination), this method is ideally suited for analysis of HIV sequences involving recent transmission and other viral sequences with lower substitution rates (e.g., human T-cell leukemia/lymphoma viruses).

ACKNOWLEDGMENTS

I thank David Hillis, Gerry Myers, and Alan Templeton for helpful discussions concerning the analyses presented and helpful comments on the manuscript.

This work was supported by National Science Foundation grant DEB-9303258 and by a grant from the Alfred P. Sloan Foundation.

REFERENCES

1. **Crandall, K. A.** 1994. Intraspecific cladogram estimation: accuracy at higher levels of divergence. Syst. Biol. **43:**222–235.
2. **Crandall, K. A.** Multiple interspecies transmissions of STLV-I and HTLV-I sequences. Submitted for publication.
3. **Crandall, K. A., and A. R. Templeton.** 1993. Empirical tests of some predictions from coalescent theory with applications to intraspecific phylogeny reconstruction. Genetics **134:**959–969.
4. **Crandall, K. A., A. R. Templeton, and C. F. Sing.** 1994. Intraspecific phylogenetics: problems and solutions, p. 273–297. *In* R. W. Scotland, D. J. Siebert, and D. M. Williams (ed.), Models in phylogeny reconstruction. Clarendon Press, Oxford.
5. **DeBry, R. W., L. G. Abele, S. H. Weiss, M. D. Hill, M. Bouzas, E. Lorenzo,**

F. Graebnitz, and L. Resnick. 1993. Dental HIV transmission? Nature (London) **361:**691.
6. **Felsenstein, J.** 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution **39:**783–791.
7. **Harvey, P. H., and M. D. Pagel.** 1991. The comparative method in evolutionary biology. Oxford University Press, Oxford.
8. **Hillis, D. M., and J. J. Bull.** 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Syst. Biol. **42:**182–192.
9. **Hillis, D. M., and J. P. Huelsenbeck.** 1994. Support for dental HIV transmission. Nature (London) **369:**24–25.
10. **Hillis, D. M., J. P. Huelsenbeck, and C. W. Cunningham.** 1994. Application and accuracy of molecular phylogenies. Science **264:**671–677.
11. **Huelsenbeck, J. P., and D. M. Hillis.** 1993. Success of phylogenetic methods in the four-taxon case. Syst. Biol. **42:**247–264.
12. **Ou, C.-Y., C. A. Ciesielski, G. Myers, C. I. Bandea, C.-C. Luo, B. T. M. Korber, J. I. Mullins, G. Schochetman, R. L. Berkelman, A. N. Economou, J. J. Witte, L. J. Furman, G. A. Satten, K. A. MacInnes, J. W. Curran, H. W. Jaffe, L. I. Group, and E. I. Group.** 1992. Molecular epidemiology of HIV transmission in a dental practice. Science **256:**1165–1171.
13. **Roff, D. A., and P. Bentzen.** 1989. The statistical analysis of mitochondrial DNA polymorphisms: chi-square and the problem of small samples. Mol. Biol. Evol. **6:**539–545.
14. **Sabino, E. C., E. G. Shpaer, M. G. Morgado, B. T. M. Korber, R. Diaz, V. Bongertz, S. Cavalcante, B. Galvao-Castro, J. I. Mullins, and A. Mayer.** 1994. Identification of human immunodeficiency virus type 1 envelope genes recombinant between subtypes B and F in two epidemiologically linked individuals from Brazil. J. Virol. **68:**6340–6346.
15. **Sing, C. F., M. B. Haviland, A. R. Templeton, K. E. Zerba, and S. L. Reilly.** 1992. Biological complexity and strategies for finding DNA variations responsible for inter-individual variation in risk of a common chronic disease, coronary artery disease. Ann. Med. **24:**539–547.
16. **Sing, C. F., M. B. Haviland, K. E. Zerba, and A. R. Templeton.** 1992. Application of cladistics to the analysis of genotype-phenotype relationships. Eur. J. Epidemiol. **8:**3–9.
17. **Swofford, D. L.** 1993. PAUP: phylogenetic analysis using parsimony. Version 3.1.1. Smithsonian Institution, Washington, D.C.
18. **Swofford, D. L., and G. J. Olsen.** 1990. Phylogeny reconstruction, p. 411–501. *In* D. M. Hillis and C. Moritz (ed.), Molecular systematics. Sinauer Associates, Inc., Sunderland, Mass.
19. **Templeton, A. R.** 1993. The "Eve" hypothesis: a genetic critique and reanalysis. Am. Anthropol. **95:**51–72.
20. **Templeton, A. R., E. Boerwinkle, and C. F. Sing.** 1987. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. I. Basic theory and an analysis of alcohol dehydrogenase activity in Drosophila. Genetics **117:**343–351.
21. **Templeton, A. R., K. A. Crandall, and C. F. Sing.** 1992. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. Genetics **132:**619–633.
22. **Templeton, A. R., and C. F. Sing.** 1993. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. IV. Nested analyses with cladogram uncertainty and recombination. Genetics **134:**659–669.
23. **Wolfram, S.** 1991. Mathematica. Addison-Wesley, Redwood City, Calif.
24. **Zharkikh, A., and W.-H. Li.** 1992. Statistical properties of bootstrap estimation of phylogenetic variability from nucleotide sequences. I. Four taxa with a molecular clock. Mol. Biol. Evol. **9:**1119–1147.
25. **Zharkikh, A., and W.-H. Li.** 1992. Statistical properties of bootstrap estimation of phylogenetic variability from nucleotide sequences. II. Four taxa without a molecular clock. J. Mol. Evol. **35:**356–366.