J Mol Evol (1997) 44:199–206

JOURNAL OF MOLECULAR EVOLUTION

Elucidating the Population Histories and Transmission Dynamics of Papillomaviruses Using Phylogenetic Trees

Chi-Keong Ong,[1] Sean Nee,[1] Andrew Rambaut,[1] Hans-Ulrich Bernard,[2] Paul H. Harvey[1]

[1]Department of Zoology, Oxford University, Oxford, OX1 3PS, United Kingdom
[2]Laboratory for Papillomavirus Biology, Institute of Molecular and Cell Biology, National University of Singapore, Kent Ridge, Singapore 0511, Republic of Singapore

Received: 4 June 1996 / Accepted: 12 August 1996

Abstract. Using gene genealogies constructed from gene sequence data, we show that both the mucosal and cutaneous papillomaviruses (PV)—supergroups A and B—appear to have been transmitted through susceptible populations faster than exponentially. The data and methods involved (1) examining the PV database for phylogenetic signal in an L1 open reading frame (ORF) fragment and an E1 ORF segment, (2) demonstrating that the same two fragments have evolved in a way consistent with a molecular clock, and (3) applying methods of phylogenetic tree analysis that test different scenarios for the dynamics of viral transmission within populations. The results indicate increases in PV populations of both supergroups A and B in the recent past. This form of the increases, which fit a null model of population growth with an exponent increasing in time, is compatible with the fact that human populations have grown at a faster than exponential rate, thus increasing the numbers of susceptible hosts for HPVs. There are, however, indications that the population of supergroup A has now stopped increasing in size.

Key words: Papillomaviruses — Phylogeny — Molecular clock — Population history — Transmission dynamics

Correspondence to: P. H. Harvey

Introduction

Human papillomaviruses (HPVs) involved in benign and malignant neoplasias of humans (zur Hausen 1991) are a diverse group of viruses with over 70 recognized types (de Villiers 1994) of which more than half have been isolated from anogenital or oral mucosal lesions. These mucosal viral types are commonly referred to as genital HPVs, some of which have been associated with specific diseases. For example, HPV-16 and HPV-18 are commonly found in cervical carcinomas and severe dysplasias while HPV-6 and HPV-11 are usually associated with benign exophytic and cervical lesions. Aside from the mucosal viruses, many other nongenital HPV types have been identified in lesions of patients suffering from epidermodysplasia verruciformis (EV) and in cutaneous lesions of other pathologies such as common warts. Papillomavirus types have also been identified in mammals and birds (Sundberg 1987), and several have been implicated in animal cancers (Shope and Hurst 1933; Campo et al. 1980).

Large amounts of genomic papillomavirus DNA sequence data have recently become available, and the database is expanding rapidly. Systematic studies incorporating phylogenetic inference have begun in earnest (Chan et al. 1992*a*,*b*, 1995; van Ranst *et al.* 1992; Ong *et al.* 1993; Bernard *et al.* 1994*a*,*b*). However, these studies have not asked what the structures of published trees might reveal about the evolutionary and population processes that led to their genesis. The spread of the virus through contemporary human populations has been analyzed (Chan *et al.* 1992*a;* Ho *et al.* 1993*a*,*b;* Ong *et al.*

1993), although accurate predictions have been hampered by the lack of appropriate tools or models.

In this paper, we use Nee *et al.*'s (1995) methods for analyzing the branching structures of phylogenetic trees to help reveal the population-dynamic history of papillomaviruses. The basic claim is that the distribution of lineage-splitting events across a tree (that is, the branching pattern in a phylogeny), provides information about the relative frequency of transmission events through time. In particular, the temporal distribution of internode intervals in a tree reveals when the viral populations might have grown in size, stayed constant, or declined. The analytical methods used to identify particular types of past population history utilize transformations of the number-of-lineages axis or the time axis in a lineages-through-time plot (Nee *et al.* 1995). These methods have been previously applied to data on the human immunodeficiency virus type 1 (HIV-1), the hepatitis C virus (HCV) (Holmes *et al.* 1995), and flaviviruses (Zanotto *et al.* 1996).

We adopt the taxonomic system of Chan *et al.* (1995). Papillomaviruses (PVs) are divided into five supergroups, A to E. Most of the clinically important viruses belong to supergroup A (mucosal and genital papillomaviruses) or B (EV papillomaviruses). Supergroups C and D are the ungulate PVs, containing, respectively, all of the fibropapillomaviruses and those viruses that cause true papillomas. Supergroup E is the most poorly defined of the five groups and contains a mix of animal and human cutaneous PVs. Here, using phylogenies constructed from genomic fragments of the L1 and the E1 open reading frames (ORFs), we examine the broad population histories of supergroups A and B, and also the more recent intratype histories of two viruses within supergroup A: HPV-16 and HPV-18.

There is no use in performing a phylogenetic analysis on data that have lost their phylogenetic signal, so we tested for that signal before constructing our phylogenies. Nee *et al.*'s (1995) methods currently require phylogenies built under the assumption of a constant rate of nucleotide substitutions (a molecular clock), because this allows for a direct projection on to a relative-time axis. We therefore tested that assumption in the data set. Once we had satisfied ourselves that we were only using data that had phylogenetic signal, and which had evolved in a way consistent with a molecular clock, we constructed the phylogenies and applied Nee *et al.*'s methods to yield a picture of the past epidemiology of papillomaviruses.

Materials and Methods

Origins of Genomic Sequences of HPV Types. The two genomic fragments used were (1) a ~300-bp fragment from the L1 ORF which has become a widely used genomic segment for the typing, taxonomy, and phylogenetic reconstruction of papillomaviruses (Bernard *et al.* 1994*b;* Chan *et al.* 1995), and (2) a 367-bp fragment of the E1 ORF that spans the region from 2178 to 2544 relative to the published HPV-16 sequence. The L1 sequences were kindly supplied by Dr. S.Y. Chan or were obtained by anonymous ftp from HPV database (ftp atlas@lanl.gov). The E1 sequences were obtained from Genbank (release 94.0). In both cases, the sequences could be aligned unambiguously. In the 30-taxa supergroup A comparison of E1 and L1 fragments, sequence data for both fragments were obtained from the same isolate. Sequences from the following 30 types were used: HPV-2a, 3, 6b, 7, 10, 11, 13, 16, 18, 26, 27, 30, 31, 32, 33, 34, 35, 39, 40, 42, 45, 51, 52, 53, 56, 57, 58, 59, RPV (rhesus papillomavirus), and PCPV (pygmy chimpanzee papillomavirus). A supergroup A L1 ORF phylogeny of 47 types was created, using (additionally) HPV-28, 43, 44, 54, 55, 61, 62, 64, 66, 67, 68, 69, 70, 72, 73, and two putative new HPV types, tentatively designated MM4 and iso39. The supergroup B L1 phylogeny was created using the following sequences: HPV-4, 5b, 8, 9, 12, 14, 15, 17, 19, 20, 21, 22, 23, 24, 25, 36, 37, 38, 47, 48, 49, 50, 60, and 65.

Further information on the isolation of HPV types numbered 1 to 69 is contained in three reviews (Delius and Hoffmann 1994; Myers *et al.* 1994; van Ranst *et al.* 1994). Chan *et al.* (1995) provides information about HPVs numbered higher than 69 and on MM4 and iso39. Published sequences of the noncoding 364- and 321-bp long control region (LCR) fragments for HPV-16 (Chan *et al.* 1992*a;* Ho *et al.* 1993*b*) and HPV-18 (Ong *et al.* 1993) were used in the construction of intratype trees. Ten additional HPV-18 LCR sequences were used in this study and their Genbank accession numbers are U59156–U59165.

Analysis of Data and the Degree of Saturation of Transitions at the Various Codon Positions. The level of phylogenetic signal was assessed by first identifying evidence for saturated change at particular sites and then finding whether those sites with near-saturated change were still phylogenetically informative. Saturated transition substitutions at each codon position were identified by plotting the numbers of transitions of all pairwise comparisons for each individual codon position against the total number of transversions in all three codon positions. A *t*-test was carried out to see if the second-order term of a polynomial regression differed from zero. The magnitude of the second-order polynomial term indicates the degree of saturation. For the 300-bp L1 ORF fragment, the analysis was conducted on all possible pairwise comparisons between 24 sequences available for the supergroup B papillomaviruses. For the 367-bp E1 ORF fragment, the same supergroup was analyzed again. However, out of the original 24 analyzed in the L1, only 13 sequences were available for the E1 analysis.

The degree to which any observed saturation obscures the phylogenetic signal was performed using the distribution of tree lengths of all bifurcating tree topologies (135,135 unrooted trees) for more than 20 randomly selected sets of nine taxa samples from all available PV types (Hillis and Huelsenbeck 1992), under a parsimony model of tree reconstruction. Tree-length distributions for both the L1 and E1 fragments were estimated using PAUP 3.1.1. Each of the constituent codon base positions of the L1 and E1 fragments were analyzed as a separate data matrix. Random selections of no more than nine taxa at a time enabled us to make a rapid assessment of the general phylogenetic signal inherent in the entire PV database at each codon position. Using this test, data matrices with phylogenetic signal should produce tree-length distributions that are strongly left skewed whereas those comprised of random noise should approximate a symmetrical distribution. Permutation tail probability (PTP) tests (Archie 1989*a,b;* Faith and Cranston 1991) were used as an additional way of determining whether phylogenetic signal was present and produced results similar to those of Hillis and Huelsenbeck's procedure.

Analysis of Phylogenetic Tree Structure. The L1, E1, and HPV-16 and HPV-18 LCR nucleotide sequences were aligned by eye. Distance matrix phylogenies (program KITSCH) and maximum likelihood phylogenies with and without the assumption of a molecular clock (DNAMLK and DNAML) were reconstructed using PHYLIP 3.5 (Felsenstein (1982) and the fastDNAml version 1.1.1a program (Olsen *et al.* 1994). Maximum parsimony analyses were conducted using PAUP 3.1.1 (Swofford 1993).

The logarithm of the number of lineages in the phylogeny was plotted against time, or in this case genetic distance, which represents relative time when the molecular clock assumption holds. The shape of this semilogarithmic plot gives some indication of population-dynamic history and is used to suggest further transformations of the number-of-lineages axis or the time axis to provide tests of more specific population-dynamic histories (Nee *et al.* 1995). These transformations distinguish among populations that have increased or decreased in size or remained constant and, for growing populations, they can distinguish among those growing exponentially at constant, accelerating, or decelerating rates. The methods are designed for small population samples, and simulation studies show that they are informative when $\leq 0.0625\%$ of the total extant population has been sampled (Ong *et al.* 1996). We apply these methodologies to the various PV data sets; for more detailed accounts of the procedures used see Harvey *et al.* (1995) and Harvey (1996). Two computer packages were used to implement these analyses: the ENDEMIC-EPIDEMIC analytical package (Rambaut *et al.* in press *a*) and the BIRTH-DEATH simulation package (Rambaut *et al.* in press *b*). These packages are available by anonymous ftp from ftp://evolve.zps.ox.ac.uk/packages. In the analysis of intratype trees of HPV-16 and HPV-18, only sequences that differed from one another by two or more substitutions were included. Sequences that differed by only one substitution are noninformative for the creation of lineages-through-time plots (Nee *et al.* 1995). The KITSCH algorithm was used to reconstruct these modified trees.

Results

Data Analysis

Determining Which Classes of Transition Substitutions Were Saturated

Figure 1 illustrates the analysis that we carried out for the papillomaviruses classed as supergroup B viruses for both the L1 and E1 ORF fragments. The plots from Fig. 1A and B indicate that increases in the first and third codon position transitions are nonlinear in comparison with the overall increase in transversions. Only the second base positions of both E1 and L1 fragments suggested that a linear regression was as good a fit as a quadratic regression. Therefore Fig. 1A and B suggests that there was no saturation in second codon position transitions while the first position was approaching saturation for both the E1 and L1 ORF fragments. The third positions appear heavily saturated for both fragments.

Are the First and Third Codon Positions Still Phylogenetically Informative?

To test for significant phylogenetic structure in our data matrices, we used tabulated 95% and 99% values of a skewness test statistic, g_1, for data sets comprising nine taxa and four-state characters of 10–100 characters (from Hillis and Huelsenbeck 1992).

We present the analysis of three data sets (Table 1) whose results are typically representative. For the L1 analyses, we separate results from an analysis of nine randomly chosen HPV types from supergroup B from those which look only at the animal papillomaviruses and contain the whole of supergroup C, bovine papillomavirus type-3 (BPV-3), *Mastomys natalensis* PV (MnPV), and cotton-tail rabbit PV (CRPV). For the E1 analysis, we include the result of nine randomly chosen PV types that represent a broad spectrum of PV types, comprising the elk papillomavirus (EPV), BPV-4, and HPV-3, 9, 16, 18, 25, 32, and 49. Table 1 lists the number of variable sites for each individual codon positions for the three data sets and their g_1 values. The first and second codon positions of all three data sets demonstrate strong left skew while the third codon positions approximate a symmetrical distribution. This indicates that third codon positions are not significantly different from that of random noise. While regression analysis indicates that the first codon was approaching saturation, it nevertheless appears to be phylogenetically informative. Faith and Cranston's (1991) PTP test was also used on the data and it corroborated the informative nature of the first and second codon positions and the noninformative nature of the third codon position.

Did Papillomaviruses Evolve to a Strict Molecular Clock?

The clock assumption may be tested by comparing the likelihood values calculated with (l_0) and without (l_1) restriction of rate constancy amongst the lineages (Felsenstein 1981; Goldman 1993; Yang *et al.* 1995*a,b*). In the study on supergroup A, 30 taxa where sequences were available for both the E1 and L1 fragments were compared, used in tree reconstruction, and used in a test for a molecular clock. Supergroup B is also assessed for a molecular clock, although the work is confined to the L1 fragment since a reasonable number of sequences (24) for this supergroup were available only for that fragment. We also examined intratype phylogenies of two viruses within supergroup A, HPV-16 and HPV-18, for evidence of molecular clocks.

Table 2 lists the differences in log-likelihood values of maximum likelihood trees reconstructed with and without the assumption of a constant-rate molecular clock for two supergroups of papillomaviruses (A and B). FastDNAml 1.1.1 and the DNAMLK program in the PHYLIP package were used to generate maximum likelihood trees with and without a clock assumption. The degree of freedom corresponding to the appropriate χ^2 value is the difference between the number of branches (α) in the unrooted tree topology and the number of branching nodes (β) in the rooted tree. If n taxa are used, then $\alpha = 2n - 3$ and $\beta = n - 1$. The χ^2 value is compared against twice the difference ($2\Delta l$) between l_0 and l_1. The nonsignificant results suggest that the molecular clock assumption is reasonable.

Analysis of Phylogenetic Tree Structure

We used trees constructed under the constant rate assumption as the basis for lineages-through-time plot analyses using the procedures given by Nee *et al.* (1995).

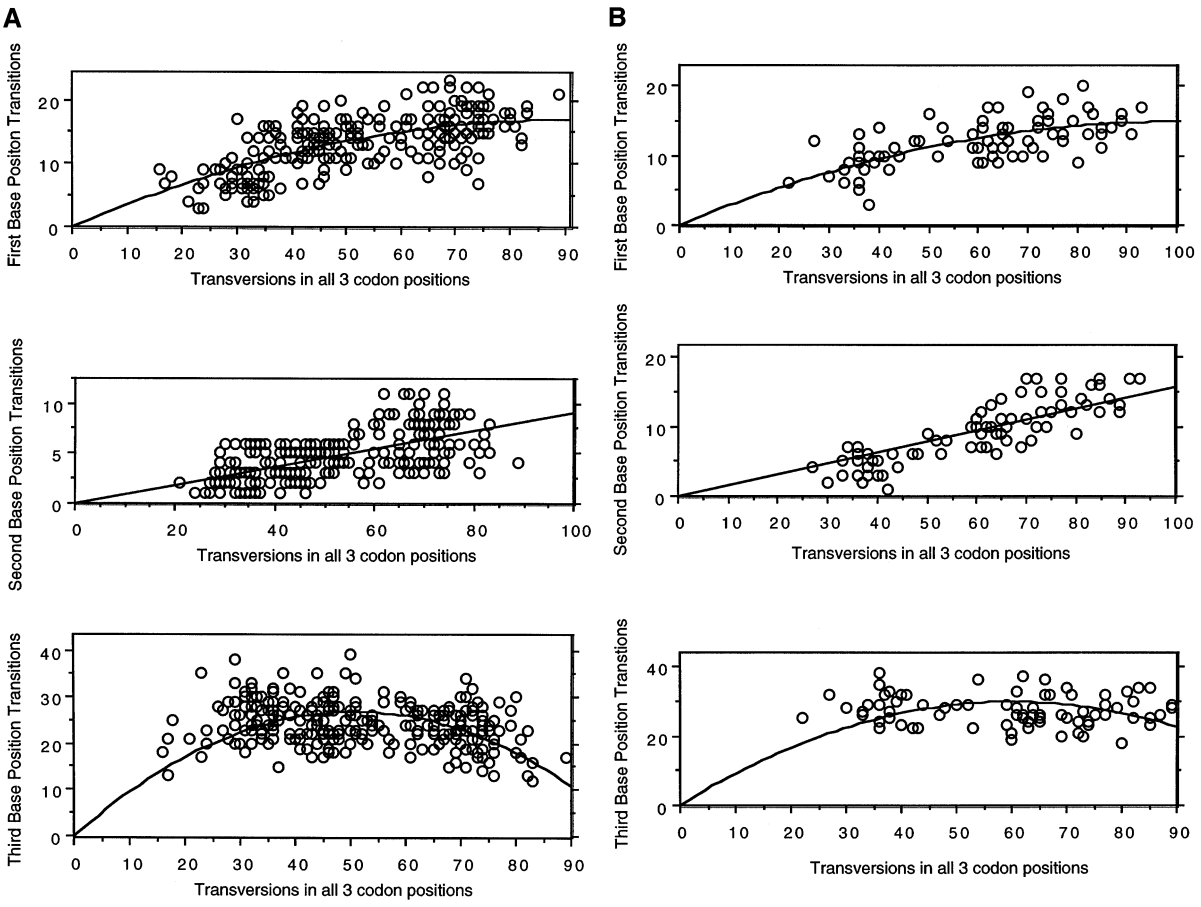

Fig. 1. Plots of the numbers of transitions and transversions at the first, second, and third codon positions for **(A)** 24 L1 sequences from supergroup B and **(B)** 13 E1 sequences from the same group against the total number of transversions for all three codon positions. One-tailed *t*-test statistics for the second-order term in a polynomial regression (with degrees of freedom corrected for nonindependence, by conservatively restricting the degrees of freedom to the absolute numbers of independent comparisons) are $t_{21} = 9.00$, $P < 0.05$, for the first position and $t_{21} = 28.96$, $P < 0.05$, for the third position of the L1 regressions and $t_{10} = 4.72$, $P < 0.05$, for the first position and $t_{10} = 12.78$, $P < 0.05$, for the third position of the E1 regressions. All other *P* values are >0.05, indicating that a linear model is sufficient for transitions at the second codon position.

Table 1. g_1 skewness values, number of variable sites (in parentheses), and associated statistical significance ($P < 0.01$, represented by **, ns = not significantly skewed)

Codon positions	L1 from supergroup B	L1 from animal PVs	E1 of various PVs
1	−0.76 (66) **	−0.78 (73) **	−0.47 (80) **
2	−0.78 (41) **	−0.96 (51) **	−0.47 (60) **
3	−0.13 (100) ns	−0.19 (94) ns	−0.14 (115) ns

Table 2. Differences in log likelihood values of maximum likelihood trees constructed with and without the assumption of a molecular clock

	Degrees of freedom (*df*)	$2\Delta l$	$\chi^2_{0.05}$
30 types supergroup A E1	28	12.120	41.337
30 types supergroup A L1	28	5.435	41.337
24 types supergroup B L1	22	15.496	33.924
HPV-16 LCR	47	59.830	64.001
HPV-18 LCR	33	40.178	44.985

The supergroup semilogarithmic plots became shallower toward the present, indicating that populations may have been growing exponentially. As a consequence, the so-called ''epidemic'' transformation was applied to the number-of-lineages axis to determine whether the exponent had stayed constant or had been increasing or decreasing through time (see Fig. 2–5).

We reconstructed a DNAMLK tree (not shown) using all 47 L1 sequences currently available for supergroup A using only the first two codon base positions. Figure 2 displays a similar pattern to that observed for plots derived from the 30 taxa (types) comparative studies of the L1 and E1 fragments (Figs. 3 and 4): There is a steepening toward the present. This similarity suggests that these plots are not artefacts of looking only at certain regions of the genome and that the 30 types we have used in our comparative studies are representative of the total data set.

Application of the epidemic transform to a supergroup B lineages-through-time plot (Fig. 5) indicates that this population of viruses has also undergone an increase in

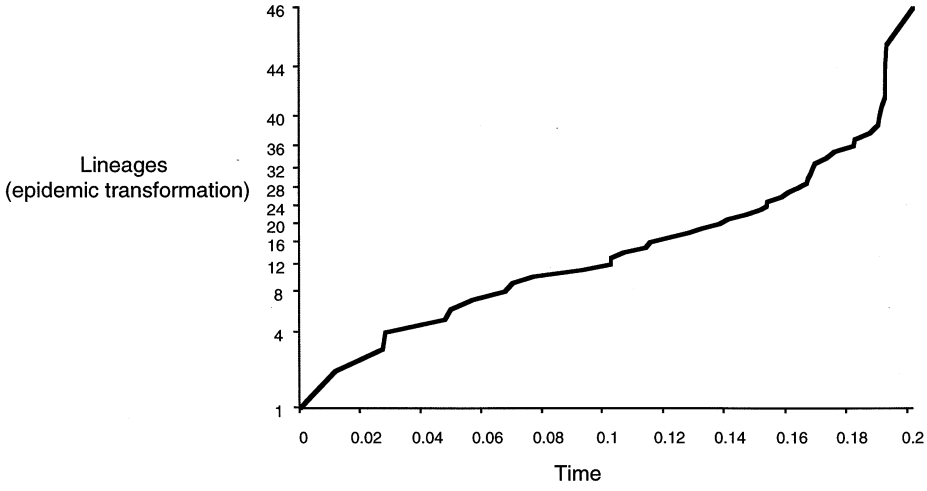

Fig. 2. Epidemically transformed lineages-through-time plot derived from a 47-type supergroup A L1 DNAMLK phylogeny. Time is measured as substitutions per base from the root of the tree to the present. Significant departure from linearity was assessed using a Wilcoxon signed-rank test. The proportion of points in the plot amongst the most recent coalescences that visually contributed toward a steepening was assessed for supergroup A and was found to be approximately 24%. Therefore, points from the most recent 24% of points (11 samples) and from the middle 24% of the plot were sampled and compared. Reciprocals of the gradients (in order to eliminate infinite gradients, since the time between any two coalescence events can be zero) of the two sets of sampled points were calculated with the gradient reciprocals of the middle portion of the curve being subtracted from those at the end portion of the curve. A one-tailed test (since the line turns upward) at the $p = 0.05$ level, where only the positive values were flagged, was performed. The test statistic, $T = 8$, $P < 0.05$, indicated that the line departed significantly from linearity.

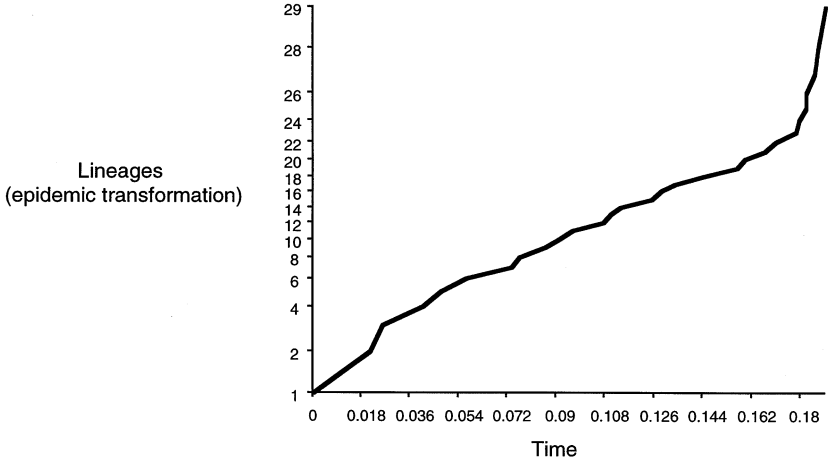

Fig. 3. Epidemically transformed lineages-through-time plot derived from a 30-type supergroup A L1 DNAMLK phylogeny. Points from the most recent 24% of points (six samples) and from the middle 24% of the plot were sampled and compared. The test statistic $T = 0$, $P < 0.05$, indicated that the line departed significantly from linearity.

population size in the relatively recent past and that they are being transmitted with an exponent that is increasing with time.

Previously, Ong *et al.* (1993) showed that the distance in an LCR fragment between HPV-45 and its closest neighbor HPV-18 was approximately 2.6 times that between typical HPV-18 variants. Given the constant-rate maximum likelihood E1 phylogeny of supergroup A (not shown) used to derive Fig. 4 and assuming a similar relationship exists between HPV-18 and HPV-45 in the E1, we are able to show that the distance representing typical variants within a HPV-18 phylogeny would encompass not more than 10% of the most recent coalescences from the supergroup A phylogeny.

The analyses of the intratype phylogenies of both HPV-16 and HPV-18 therefore offer a ''higher-resolution'' view of the recent steepening in the transformed plots of supergroup A. This in effect represents an expanded study of the coalescences from the most recent past for specific types within a supergroup A phylogeny. The semilogarithmic lineages-through-time plots (not shown) for both HPV-16 and HPV-18 indicated that the endemic transform was suitable, since the plots were steepening toward the present instead of shallowing off (Nee *et al.* 1995). Application of the endemic transformation suggested that both HPV-16 and HPV-18 populations have been roughly constant over the period of time covered by these coalescences because the plots do not differ significantly from a straight line (Figs. 6 and 7).

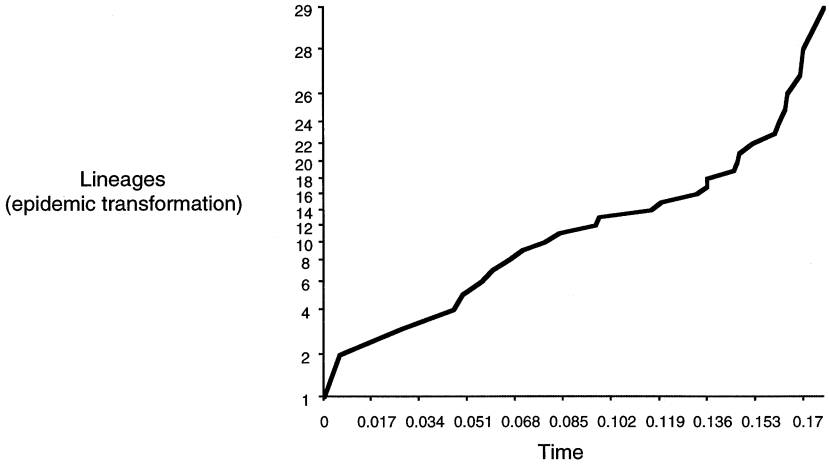

Fig. 4. Epidemically transformed lineages-through-time plot derived from a 30-type supergroup A E1 DNAMLK phylogeny. Points from the most recent 24% of points (six samples) and from the middle 24% of the plot were sampled and compared. The test statistic $T = 0$, $P < 0.05$, indicated that the line departed significantly from linearity.

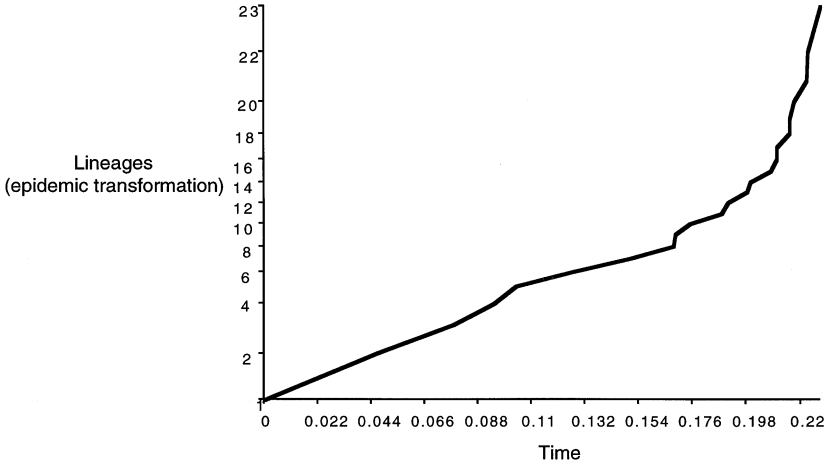

Fig. 5. Epidemically transformed lineages-through-time plot derived from a 24-type supergroup B (EV) L1 DNAMLK phylogeny. Approximately one-third of points in the plot contributed to the steepening. Therefore, points from the most recent 33% of points (seven samples) and from the middle 33% of the plot were sampled and compared. The test statistic $T = 0$, $p < 0.05$, indicated that the line departed significantly from linearity.

Discussion

The methods of Nee *et al.* (1995) seem well suited for analyzing the papillomavirus database, which offers different time scales that may be explored. Take supergroup A as an example. We have shown that epidemic transformed plots of the intertype phylogenies demonstrate that there has been an explosion in the viral population in the past. We are also able to discern the very recent population history of HPV-18 and possibly of HPV-16 if we assume that both their intratype phylogenies represent the coalescences of the most recent past from a supergroup A phylogeny. Both the semilogarithmic lineages-through-time plots of the intratype phylogenies and the endemically transformed plots suggested that HPV-16 and HPV-18 populations have not been growing exponentially in the recent past but that they fit the null model, indicating that these viruses have been infecting a roughly constant number of hosts in the recent past.

At the present, HPV-16 is the predominant genital HPV type found in invasive cancers of all populations worldwide (Syrjaenen *et al.* 1987; Orth 1994). HPV-16 accounts for 50–60% of all such cases, with HPV-18

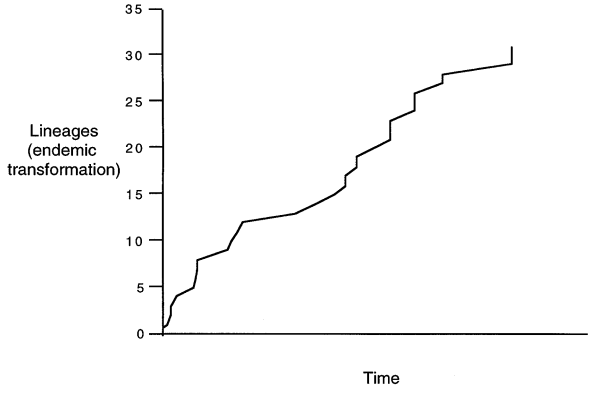

Fig. 6. Endemically transformed lineages-through-time plot derived from a HPV-16 KITSCH phylogeny. Because the time axis is formed by concatenating transformed intervals, it is not possible to put a scale to the axis for both Fig. 6 and 7 (Nee *et al.* 1995). Linearity of line indicates that HPV-16 has been infecting a constant population in the recent past. Significant departure from linearity was tested using the uniform conditional test (Cox and Lewis, 1966), which is used as a test statistic for determining whether time intervals between a series of events fit a Poisson distribution. A two-tailed test has $Z = 1.75$, $P > 0.05$; thus the line does not depart significantly from linearity. This test was implemented in the END-EPI program.

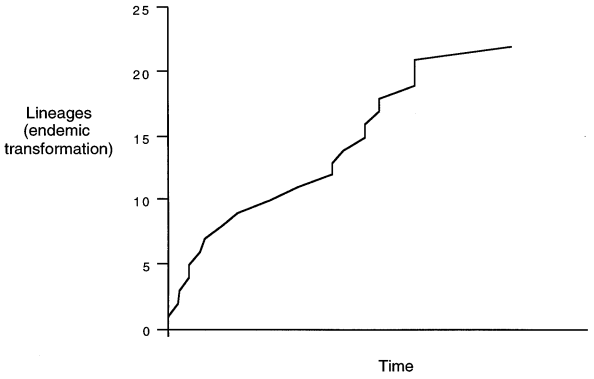

Fig. 7. Endemically transformed lineages-through-time plot derived from a HPV-18 KITSCH phylogeny. A two-tailed test has $Z = 1.01$, $P > 0.05$; thus the line does not depart significantly from linearity.

accounting for 10–20% and HPV-31, 33, 35, 39, 45, 51, 52, 56, and 58 collectively accounting for another 10–20% (Orth 1994). Just because HPV-16 and HPV-18 populations appear to have remained constant in the recent past says nothing about the very recent epidemic history of the rest of the supergroup A population. Any HPV types associated with invasive carcinomas that are found to be transmitting exponentially right up to the present day would be candidates to replace both HPV-16 and HPV-18 as the most prevalent cancer-associated HPV of the future. If the pattern observed with HPV-16 and HPV-18 is found to be consistent in analyses of the intratype phylogenies of all other members of supergroup A, this would imply that supergroup A viruses were transmitting exponentially until the very recent past, when population growth stabilized, unlike that of the human population.

These observations would be compatible with previous predictions that, for the present, (1) PV researchers may be ''asymptotically approaching the identification of all HPVs that infect the genital tract'' and that (2) we have identified most if not all of the cancers associated HPVs (Galloway 1994).

Conclusions

Applying the methods of Nee *et al.* (1995) we have been able to discern the histories of different HPV populations and their transmission profiles. The demonstration of a molecular clock in the evolution of various papillomaviruses overcomes one of the most controversial assumptions that must be made when applying these methods (Holmes *et al.* 1995). Our results indicate that faster than exponential increases in PV populations of both supergroups A and B apparently occurred in the recent past. This would be compatible with the fact that human populations have grown at a faster than exponential rate, thus increasing the numbers of susceptible host for HPVs. There are, however, indications that, for supergroup A, the population size may have stabilized.

Acknowledgments. We would like to thank Dr. E.C. Holmes for helpful discussion. This work was funded by the BBSRC grant no. GR/H53655 (PHH), and the Wellcome Trust grant no. 038468 (PHH).

References

Archie JW (1989*a*) A randomisation test for phylogenetic data in systematic data. Syst Zool 38:239–252

Archie JW (1989*b*) Phylogenies in plant families: a demonstration of phylogenetic randomness in DNA sequence data derived from proteins. Evolution 43:1796–1800

Bernard HU, Chan SY, Delius H (1994*a*) Evolution of Papillomaviruses. Curr Top Microbiol Immunol 186:33–54

Bernard H-U, Chan SY, Manos MM, Ong CK, Villa LL, Delius H, Peyton CL, Bauer HM, Wheeler CM (1994*b*) Identification and assessment of known and novel human papillomaviruses by polymerase chain reaction, restriction digest fingerprinting, nucleotide sequence and phylogenetic algorithms. J Infect Dis 170:1077–1085

Campo MS, Moar MH, Jarrett WHF, Laird HM (1980) A new papillomavirus associated with alimentary cancer in cattle. Nature 286: 180–182

Chan SY, Ho L, Ong CK, Chow V, Drescher B, Duerst M, ter Meulen J, Villa LL, Luande J, Mgaya HN, Bernard H-U (1992*a*) Molecular variants of human papillomavirus-type 16 from four continents suggest ancient pandemic spread of the virus and its coevolution with humankind. J Virol 66:2057–2066

Chan SY, Bernard H-U, Ong CK, Chan SP, Hoffman B, Delius H (1992*b*) Phylogenetic analysis of 48 papillomaviruses and 28 subtypes and variants: a showcase for the molecular evolution of DNA viruses. J Virol 66:5714–5725

Chan SY, Delius H, Halpern AL, Bernard H-U (1995) Analysis of genomic sequences of 95 papillomavirus types: uniting typing, phylogeny, and taxonomy. J Virol 69:3074–3083

Cox DR, Lewis PAW (1966) The statistical analysis of series of events. Methuen, London

de Villiers E-M (1994) Human pathogenic papillomavirus types: an update. Curr Top Microbiol Immunol 186:1–12

Delius H, Hoffmann B (1994) Primer-directed sequencing of human papillomavirus types. Curr Top Microbiol Immunol 186:13–31

Faith DP, Cranston PS (1991) Could a cladogram this short have arisen by chance alone?: On a permutation test for cladistic structure. Cladistics 7:1–28

Felsenstein J (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach. J Mol Evol 17:368–376

Felsenstein J (1982) Numerical methods for inferring evolutionary trees. Q Rev Biol 57:379–404

Galloway DA (1994) Editorial: navigating the descent into papillomavirus hell. J Infect Dis 170:1075–1076

Goldman N (1993) Statistical tests of models of DNA substitution. J Mol Evol 36:182–198

Harvey PH (1996) Phylogenies for ecologists: the 1995 Tansley lecture. J Anim Ecol 65:255–263

Harvey PH, Holmes EC, Rambaut A, Nee S (1995) How to reveal viral population history from gene sequence data. In: Schwyzer M, Ackermann M (eds) Immunobiology of viral infections. Fondation Mérieux, pp 1–8

Hillis DM, Huelsenbeck JP (1992) Signal, noise, and reliability in molecular phylogenetic analyses. J Hered 83:189–195

Ho L, Tay SK, Chan SY, Bernard H-U (1993*a*) Sequence variants of human papillomavirus type 16 from couples suggest sexual transmission with low infectivity and polyclonality in genital neoplasia. J Infect Dis 168:803–809

Ho L, Chan SY, Burk RD, Das BC, Fujinaga K, Icenogle JP, Kahn T, Kiviat N, Lancaster W, Mavromara-Nazos P, Labropoulou V, Mitrani-Rosenbaum S, Norrild B, Pillai MR, Stoerker J, Syrjaenen K, Syrjaenen S, Tay SK, Villa LL, Wheeler CM, Williamson A-L, Bernard H-U (1993*b*) The genetic drift of human papillomavirus type 16 is a means of reconstructing prehistoric viral spread and the movement of ancient human populations. J Virol 67:6413–6423

Holmes EC, Nee S, Rambaut A, Garnett GP, Harvey PH (1995) Revealing the history of infectious disease epidemics through phylogenetic trees. Philos Trans R Soc Lond Biol 349:33–40

Myers G, Bernard H-U, Delius H, Favre M, Icenogle J, van Ranst M, Wheeler C (1994) Human papillomaviruses 1994. In: Wheeler C (ed) A compilation and analysis of nucleic and amino acid sequences. Los Alamos National Laboratory, Los Alamos, NM, pp III-3

Nee S, Holmes EC, Harvey PH (1995) Inferring population history from molecular phylogenies. Philos Trans R Soc Lond Biol 349: 25–31

Olsen GJ, Matsuda H, Hagstrom R, Overbeek R (1994) fastDNAml: a tool for construction of phylogenetic trees of DNA sequences using maximum likelihood. Comput Appl Biosci 10:41–48

Ong CK, Chan SY, Campo MS, Fujinaga K, Mavromara-Nazos P, Labropoulou V, Pfister H, Tay SK, ter Meulen J, Villa LL, Bernard H-U (1993) Evolution of human papillomavirus type 18: an ancient phylogenetic root in Africa and intratype diversity reflect coevolution with human ethnic groups. J Virol 67:6424–6431

Ong CK, Nee S, Rambaut A, Harvey PH (1996) Inferring the population history of an epidemic from a phylogenetic tree. J Theor Biol 182:173–178

Orth (1994) Papillomaviruses- human. In: Webster RG, Granoff A (eds) Enclyclopedia of virology, vol 2. Academic Press, London, pp 1013–1021

Rambaut A, Harvey PH, Nee S (in press *a*) End-Epi: an application for reconstructing phylogenetic and population processes from molecular sequences. Comput Appl Biosci

Rambaut A, Grassly N, Nee S, Harvey PH (in press *b*) Bi-De: an application for simulating phylogenetic processes. Comput Appl Biosci

Shope RE, Hurst EW (1933) Infectious papillomatosis of rabbits. J Exp Med 68:607–624

Sundberg JP (1987) Papillomavirus infections in animals. In: Syrjaenen K, Gissmann L, Koss LG (eds) Papillomaviruses and human disease. Springer Verlag, Berlin, pp 40–103

Swofford DL (1993) Phylogenetic analysis using parsimony: PAUP, version 3.1. Illinois Natural History Survey, Champaign, IL

Syrjaenen KJ, Gissman L, Koss LG (1987) Papillomaviruses and human disease. Springer-Verlag, Berlin

van Ranst M, Kaplan JB, Burk RD (1992) Phylogenetic classification of human papillomaviruses: correlation with clinical manifestations. J Gen Virol 73:2653–1660

van Ranst M, Tachezy AR, Burk RD (1994) Human papillomavirus sequences: what is in stock? Papillomavirus Rep 5:65–75

Yang ZH, Goldman N, Friday A (1995*a*) Maximum likelihood trees from DNA sequences: a peculiar statistical estimation problem. Syst Biol 44:384–399

Yang ZH, Luder IJ, Lin HG (1995*b*) Molecular evolution of the Hepatitis B virus genome. J Mol Evol 41:587–596

Zanotto PM de A, Gould AG, Gao GF, Harvey PH, Holmes EC (1996) Population dynamics of flaviviruses revealed by molecular phylogenies. Proc Natl Acad Sci USA 93:548–553

zur Hausen H (1991) Viruses in human cancers. Science 254:1167–1173