Molecular Biology and Evolution

Quantifying the Phylodynamic Forces Driving Papillomavirus Evolution

Marc Gottschling,[1] Markus Göker,[2] Alexandros Stamatakis,[3] Olaf R.P. Bininda-Emonds,[4] Ingo Nindl,[5] and Ignacio G. Bravo*[,6,7]

[1]Department of Biology, Systematic Botany and Mycology, Ludwig-Maximilians-University Munich, Germany

[2]German Collection for Microorganisms and Cell Cultures GmbH (DSMZ), Braunschweig, Germany

[3]The Exelixis Lab, Scientific Computing Group, Heidelberg Institute for Theoretical Studies, Heidelberg, Germany

[4]AG Systematics and Evolutionary Biology, Institut für Biologie und Umweltwissenschaften (IBU)—Faculty V, Carl von Ossietzky University, Oldenburg, Germany

[5]Department of Dermatology, German Cancer Research Centre (DKFZ)—Charité Cooperation, Skin Cancer Center Charité, University Hospital of Berlin, Berlin, Germany

[6]Genomics and Health, Centre for Public Health Research (CSISP), Valencia, Spain

[7]CIBER Epidemiology and Public Health (CIBERESP), Barcelona, Spain

***Corresponding author**: E-mail: bravo_ign@gva.es.

Associate editor: Daniel Falush

Research article

Abstract

The associations between pathogens and their hosts are complex and can result from a variety of evolutionary processes including codivergence, lateral transfer, or duplication. Papillomaviruses (PVs) are double-stranded DNA viruses ubiquitously present in mammals and are a suitable target for rigorous statistical tests of potential virus–host codivergence. We analyze the evolutionary dynamics of PV diversification by comparing robust phylogenies of PVs and their respective hosts using different statistical approaches to assess topological and branch-length congruence. Mammalian PVs segregated into four diverse major clades that overlapped to varying degrees in terms of their mammalian host lineages. The hypothesis that PVs and hosts evolved independently was globally rejected ($P = 0.0001$), although only 90 of 207 virus–host associations (43%) were significant in individual tests. Virus–host codivergence accounted roughly for one-third of the evolutionary events required to reconcile PV–host evolutionary histories. When virus–host associations were analyzed locally within each of the four viral clades, numerous independent topological congruencies were identified that were incompatible with respect to the global trees. These results support an evolutionary scenario in which early PV radiation was followed by independent codivergence between viruses within each of the major clades and their hosts. Moreover, heterogeneous groups of closely related PVs infecting non-related hosts suggest several interspecies transmission events. Our results argue thus for the importance of alternative events in PV evolution, in contrast to the prevailing opinion that these viruses show a high degree of host specificity and codivergence.

Key words: codivergence, cophylogeny, host–parasite association, Parafit, phylogenomics, treefitter, treemap.

Introduction

Vertebrates are continuously invaded by myriads of different parasites, and one of the most basic and important questions in evolutionary biology is the degree, to which diversification of parasites is linked to the speciation pattern of their hosts (Klassen 1992; Osche 1966; Johnson et al. 2003). Parasites comprise a phylogenetically heterogeneous and diverse assemblage of multi- and unicellular biological entities characterized by close ecological interactions with their hosts. In this regard, pathogens such as viruses exhibit many parasite-like traits (Mindell et al. 2004; Bamford et al. 2005): they are mostly host specific, much smaller than their host, frequently exhibit a high degree of specialization, and reproduce more rapidly and in larger numbers than their hosts.

Under the assumptions that viruses are transmitted only vertically and that they are host specific, the phylogeny of viruses should be topologically congruent with that of their hosts—Fahrenholz's rule, strict codivergence (Fahrenholz 1913). Human pathogens would then be of primate origin, implying that their ancestors infected the last common ancestor of *Homo sapiens* and its sister group. More generally, closely related viruses are expected to be found in humans, apes, and other primates. In the real world, the complex dynamics of viral infections are characterized more frequently by exceptions to rather than by agreement with these assumptions. Several evolutionary mechanisms including lateral transfer (e.g., host switch, recombination) and duplication (e.g., during colonization of new ecological niches by adaptive radiation on the same host species) can disrupt the topological congruence between the phylogenetic trees of viruses and their hosts (Lyal 1986; Page 1994; Page and Charleston 1998; Jackson 2005).

Interspecies virus transmission is a broadly accepted phenomenon, but it is still debated whether it requires adaptation to a new host species during the early stages of

http://mbe.oxfordjournals.org/

infection or whether transmission itself is largely a random and frequent process, with successful colonization involving the founder effect of the transfer of a viral strain with the necessary genetic properties (Dennehy et al. 2006; Holmes and Drummond 2007). In this respect, the concept of an "ecological license" is key, describing a "previously not utilized unit of the environment that is suitable for becoming an ecological dimension of an organism's [or a virus'] niche" (Osche 1966). The successful colonization of new hosts and the establishment of persistent infections depend on various factors subject to variation and selection, including the individual duration, persistence, and virulence of the infection; the infection rate between new hosts; the ability of the host to raise a protective immunity; and host population density, size, and structure permitting the pathogen's regional persistence (Antia et al. 2003; Wolfe et al. 2007; Elena et al. 2009). The genotypic and/or phenotypic changes associated to these evolutionary processes constitute the viral adaptations during the realization of the new ecological niche.

Many major human infectious diseases may be of animal origin (Wolfe et al. 2007), presumably because of the similar licenses provided by humans and by related mammals. Indeed, two-thirds of the approximately 1,500 known human pathogens have close relatives in nonhuman vertebrate hosts (Woolhouse et al. 2005; Heeney 2006), and there is good evidence for zoonotic transmission events in a number of major human viruses. One example is that of *Human immunodeficiency virus 1* (*Retroviridae*, *Lentivirus*), which is derived from *Simian immunodeficiency virus* and has adapted to humans during the past half-century (Gao et al. 1999; Heenay et al. 2006; Plantier et al. 2009). Another example is *Influenzavirus A* (*Orthomyxoviridae*), one of the next major threats to human public health (Wolfe et al. 2007). The principle hosts of this virus are waterfowl, but viruses in this nonhuman reservoir may exhibit an increased penetration potential due to reassortment with avian, porcine, and/or human strains (Hay et al. 2001; Ghedin et al. 2005; Russell et al. 2008; Gibbs et al. 2009). It is also known that the proximate hosts of human-specific *Measles virus* (*Paramyxoviridae*, *Morbilivirus*) are cattle, but the ultimate hosts are likely to be wild ruminants (Wolfe et al. 2007). Similarly, zoonotic events may have given rise to the (re)emergence of RNA viruses, such as *Hantavirus*, *Henipavirus*, and *Flavivirus* (Woolhouse et al. 2005; Mansfield et al. 2009; Weingartl et al. 2009).

By contrast, DNA viruses such as hepatitis B viruses, *Herpesviridae*, and papillomaviruses (*Papillomaviridae*, PVs) are considered to have closely codiverged with their mammalian hosts (McGeoch et al. 2000; Bernard et al. 2006; Kay and Zoulim 2007). Related PVs infect related host species—"Clay's rule," a more generalized formulation of Fahrenholz's rule: (Lyal 1986)—, with Delta-PV infections of ruminants being a frequently quoted example. PVs also infect humans, apes, and monkeys, and the presence of human PVs is generally regarded as a result of ancestral primate inheritance (Chan et al. 1995; Van Ranst et al. 1995; Halpern 2000; Bernard 2005).

The main interest in human PVs arises from the association of specific viruses with cervical cancer and their potential for malignant transformations in mucosal tissue (Gissmann and zur Hausen 1980; zur Hausen 2002; Muñoz et al. 2003). Human PVs have been extensively studied, and more than 150 genomes have already been completely sequenced (Bernard et al. 2010). However, animal PVs are still poorly sampled, and viruses from less than 50 nonhuman host species have been isolated and (partly) sequenced (Breitbart and Rohwer 2005; Gottschling et al. 2008; Woolhouse et al. 2008). This situation reflects the general trend that our knowledge about the closest relatives of human pathogens is often extremely limited (Mahy and Brown 2000; Nunn 2004). For example, humans are the only host in which *Mumps virus* (*Paramyxoviridae, Rubulavirus*) infection is known to be productive (Muhlemann 2004), and close relatives of the *Rubella virus* (*Togaviridae*, *Rubivirus*) affecting only humans can additionally infect animals (Zhou et al. 2007).

General conclusions about PV evolution may be premature and largely speculative as long as the knowledge about (particularly nonhuman) PV diversity is unbalanced and limited (Bravo et al. 2010). If strict codivergence with hosts had driven PV evolution, then congruence between PV and respective host phylogenies can be expected. This hypothesis would be rejected by significantly inconsistent tree topologies, either globally or with respect to specific associations between PV and their hosts. In cases of topological incongruence, alternative explanations including lateral transfer events—(e.g., host switch and recombination: Narechania et al. 2005; Varsani et al. 2006; Gottschling, Stamatakis et al. 2007; Bogaert et al. 2008; Rector et al. 2008; Gottschling et al. 2010)—and duplication on the same host—during realization of new ecological niches by adaptive radiation: García-Vallvé et al. 2005; Bravo et al. 2010)—need to be considered.

A comprehensive evaluation of the phylogenetic associations of PVs to their vertebrate hosts is still wanting. Well-resolved trees are a necessary prerequisite to test hypothesis of codivergence (Page et al. 1996), and large molecular data sets are available for both viruses and hosts. Molecular and morphological studies have improved our knowledge on the phylogenetic relationships within placental mammals during the past decade (Murphy et al. 2001; Reyes et al. 2004; Springer et al. 2004; Bininda-Emonds et al. 2007). For PVs, we have derived a comprehensive and robust phylogeny based on the E1–E2–L1 gene combination encompassing nearly 50% of the viral genome (Gottschling et al. 2007). The analyses identify four high-level assemblies or crown groups, which we consistently name here by the constituent taxa that are most distantly related as follows: Alpha + Omikron-PVs (infecting Carnivora, Cetacea, Primates, and Suina), Beta + Xi-PVs (infecting Ruminantia, Carnivora, Eulipotyphla, Primates, and Rodentia), Delta + Zeta-PVs (infecting Perissodactyla and Ruminantia), and Lambda + Mu-PVs (infecting Carnivora, Lagomorpha, Primates, and Rodentia). In the present study, we compare a PV phylogeny comprising the known viral diversity with

a multigene phylogeny of their mammalian hosts using different statistical methods to identify putative cophylogenetic structures between the viruses and their hosts. We identified global and local instances of virus–host codivergence as well as discrepancies/incongruences that are better explained by alternative evolutionary forces.

Materials and Methods

Phylogenetic Analyses Of Papillomaviruses and Their Hosts

We analyzed the concatenated amino acid (aa) sequences of the E1–E2–L1 genes from 207 PV genomes (supplementary table S1, Supplementary Material online), excluding the highly variable E4 gene region nested within the E2 gene. A previous phylogenetic analysis (Gottschling et al. 2007) identified this gene combination as being optimal for minimizing the degree of data-inherent noise. We aligned aa sequences using MAFFT v6.523 (Katoh et al. 2005). The final data matrix is freely available at http://htcc.pt-dlr.de/dateien/GottschlingParaFit.fasta. The LG protein substitution matrix (Le and Gascuel 2008) was identified as the best-suited evolutionary model for our data set using ProtTest v2.4 (Abascal et al. 2005).

Maximum Likelihood (ML)-based phylogenetic analyses were conducted using RAxML v7.2.6 (Stamatakis 2006) and the LG + Γ4 substitution model under three partitions corresponding to each of the genes, and we computed 5,000 bootstrap replicates. Bayesian phylogenetic analyses were performed with PhyloBayes (Lartillot and Philippe 2004) under the LG substitution model. We completed two independent Monte Carlo Markov Chains, allowing each of them to reach stationarity, checked for convergence between them (maximum discrepancy across all bipartitions = 0.0949; mean discrepancy across all bipartitions = 0.000934) and sampled the posterior distributions every 100th generation to obtain 1,000 values. PV trees obtained after ML and Bayesian reconstructions were compared regarding topological congruence using RAxML v7.2.6 and topd/ftms (Puigbo et al. 2008) as well as pairwise distances using K-TreeDist (Soria-Carrasco et al. 2007). The trees were rooted with the sauropsid PVs based on the E1 tree topology (García-Vallvé et al. 2005).

The host tree was obtained by extracting sequence data for host species from the molecular 68-gene data set underlying the mammalian supertree in Bininda-Emonds et al. (2007). The base data set (supplementary fig. S1, Supplementary Material online) restricted to exclude the transfer RNAs and also to include only the well-sampled genes among our host species. Using the perl script GenBankStrip.pl (freely available at http://www.molekularesystematik.uni-oldenburg.de/33997.html), we updated the data set to include homologous sequence data for the nonmammals as well as for missing host species that had been sequenced in the meantime. ML-based phylogenetic analyses were also conducted using RAxML, with 1,000 rapid bootstraps and a general time reversible + Γ substitution model. For the supermatrix analysis, we partitioned the model on a per-gene basis for each of the 35 genes and used all nonmammalian sequences as a monophyletic outgroup.

Analysis Of Cophylogenetic Associations: TreeMap

Several methods for testing codivergence hypotheses are available, most of which are reviewed in Paterson and Banks (2001) and Stevens (2004). Three representative methods, as implemented in the TreeMap, TreeFitter, and ParaFit programs, were deployed to analyze the host–parasite interactions based on the associations between PVs and their hosts. In all analyses, a P value of 0.05 was used.

The program TreeMap 1.0 (freely available at http://taxonomy.zoology.gla.ac.uk/rod/treemap.html) uses four types of events to explain the origin of a given association: codivergence (C), host switching (H), duplication or intrahost divergence of the parasite (D), and sorting or extinction of the parasite lineage (S; Johnson et al. 2003). The program explores how the parasite tree fits onto the host tree by adequately mixing the four basic types of codivergence events (Page 1994). More recent versions of the program (TreeMap 2.0 and TreeMap 3.0, both of which are beta versions and freely available at http://www.it.usyd.edu.au/~mcharles/) can compute all optimal solutions by exhaustive search using the Jungles algorithm (Charleston 1998). Both versions also implement heuristic searches and provide the option to assess the significance of the results using permutation tests based on random host and parasite trees. The fraction of permutations, for which the number of codivergence events is greater or equal to than that in the real data set, being derived from randomized data sets, defines the threshold for significance.

Execution times are prohibitive for all versions, even for moderately sized data sets. This is due to the large number of equally optimal solutions that yield equally "good" reconstructions with equal scores, especially if a high number of host-switch events was allowed. Even when reducing our data set to a single representative PV sequence per monophyletic lineage occurring on a single host species (using the newick.tcl script available at http://www.goeker.org/mg/distance/), TreeMap 1.0 was still searching for an optimal reconciliation with a maximum of five host switches before being terminated after a week of run-time, and TreeMap v2.0 crashed when trying to compute solutions with more than four host switches. Thus, we chose to explore optimal reconstructions by separately analyzing the four more restricted data sets corresponding to the four viral crown groups and their hosts. Even under these conditions, we were forced to restrict our searches to a maximum of eight host switches in the case of the Alpha + Omikron-PVs crown group. Costs for noncodivergence events, namely duplications, losses, and host switches, were considered to be equal. We executed 1,000 permutation test replicates on the parasite tree to assess significance using the actual combinations of values for the different evolutionary events to evaluate how often the random trees fitted the hypothesis as well as the reference tree being tested.

Analysis Of Cophylogenetic Associations: TreeFitter

A further technique for analyzing the fit between the host and a parasite tree is generalized parsimony (Ronquist 1995). TreeFitter is also topology based, but less computationally expensive. Therefore, it allows the exploration of different cost combinations for each of the four types of events described above (Ronquist 1995; Page and Charleston 1998; Paterson and Banks 2001; Begerow et al. 2004). Given a combination of costs for each of these events, the optimality criterion implemented in TreeFitter v1.1 (freely available at http://www.ebc.uu.se/systzoo/research/treefitter/treefitter.html) strives to minimize the global cost.

We employed a permutation-based approach to solve the problem of determining the optimal combination of different costs for each event that best explains the data. Following the procedure outlined in Ronquist (2003), who presented the results of this permutation-based approach for six hypothetical evolutionary patterns, the globally best combinations of event costs are those that yield the lowest probability of the null hypothesis. In our analyses, codivergence and sorting events were assigned zero and unit costs (1.0), respectively, whereas switch and duplication costs were varied between 0.0 and 10.0 in increments of 1.0. For each combination of costs, 10,000 permutations of the original associations (not the trees) were computed. The analysis was performed with the pruned PV and host trees and considering only a single PV representative for each monophyletic group of viruses that was present on a specific host.

Ronquist (2003) investigated a number of exemplars for hypothetical patterns characterized by a certain combination of dominant events. Because this author only included "cospeciation–duplication" pattern with late duplications, we here modified his exemplars to obtain a pattern in which pure cospeciation followed zero to four duplications at the base of the parasite tree (a duplication–cospeciation pattern). These five artificial data sets were analyzed in the same way than the empirical virus data, and the outcome was compared (see the Supplementary Material online for details). As in TreeMap, permutation tests were used to determine significance through the number of times an equally low or lower total cost was found for randomized associations compared with the real associations.

Analysis Of Cophylogenetic Associations: ParaFit

An alternative method to assess host–parasite codivergence is ParaFit (Legendre et al. 2002). Whereas TreeMap and TreeFitter are topology-based, ParaFit uses pairwise or patristic (path-length) distances to test the "global null hypothesis" ("GH_0" in the following) that "the similarity between the trees is not higher than expected by chance" given the actual associations between hosts and parasites (Legendre et al. 2002). ParaFit also uniquely estimates the contribution of each individual host–parasite association to the global fit between the matrices to test the "individual null hypothesis" ("IH_0") that "any given contribution is not different from random," such that the association could as well be omitted. Associations, for which IH_0 was accepted, will be termed "nonsignificant" in the following, and "significant" otherwise. Both tests are based on converting the distances to eigenvectors and multiplying the resulting matrices with the association matrix (Legendre et al. 2002). Significance testing is based on random permutations of the rows of the association matrix (i.e., input trees are not randomly resampled). In contrast to most other cophylogenetic tests, type I and type II error rates of ParaFit have been explored in extensive simulation studies (Legendre et al. 2002).

ParaFit-based analyses were performed using CopyCat (Meier-Kolthoff et al. 2007), which incorporates a graphical user interface and wrapper for AxParafit (Stamatakis et al. 2007). AxParafit is a significantly faster implementation of ParaFit that yields numerically identical results to the original program. Through AxPcoords (Stamatakis et al. 2007), CopyCat also conducts all necessary conversions (e.g., converting trees to patristic distances and computing the eigenvectors of the distance matrices) to prepare the data for analysis with AxParafit. In each AxParafit-based analysis, we computed 9,999 permutations of the association matrix using CopyCat.

We also corrected for three potential issues with ParaFit-based analyses. First, to rule out that differences between the PV trees as inferred from the two phylogenetic approaches used (Bayesian and ML) yield different results, we conducted all ParaFit analyses for both PV trees. Second, to account for the differential sampling of the host species with respect to their PV diversity, we selected single PVs to represent monophyletic lineages associated with a specific host (as recommended in the ParaFit manual). This selection process mainly affected PVs isolated from humans ($n = 132$), *Macaca fascicularis* ($n = 10$), cows ($n = 10$), and dogs ($n = 7$). Third, global tests in ParaFit are affected by differences in the hierarchical pattern of the host–parasite associations. The associations of predominantly congruent "local" host–parasite subtrees may be determined as being nonsignificant, if these subtrees occur at incompatible positions in the full tree (and vice versa). This is because ParaFit tests their impact on the "overall" fit between host and parasite trees (Stamatakis et al. 2007). Hence, the major PV crown groups (Gottschling, Stamatakis et al. 2007) were individually tested for congruence between virus and host trees. In particular, we tested the following set of associations with CopyCat/ParaFit, pruning the host and parasite trees as needed by appropriately adapting the input associations table in CopyCat:

1. all known PVs present on specific hosts;
2. a single PV genome as representative for each monophyletic lineage of viruses present on a single host species (e.g., HPV32 as representative for Alpha-PV species 1 infecting humans; HPV95 as representative for Gamma-PVs infecting humans; and BPV6 as representative for Xi-PVs infecting cattle) to counteract host taxa that are overrepresented with respect to the number of their associations;
3. the Alpha + Omikron-PV crown group with Carnivora, Cetacea, Primates, and Suina;

4. the Beta + Xi-PV crown group with Carnivora, Eulipotyphla, Marsupialia, Primates, Rodentia, and Ruminantia;
5. the Delta + Zeta-PV crown group with Perissodactyla and Ruminantia; and
6. the Lambda + Mu-PV crown group with Carnivora, Lagomorpha, Primates, and Rodentia.

The small size of some of these subsets was likely to decrease the power of the associated tests, and the test of individual host–parasite links in particular (Legendre et al. 2002). We therefore explored the minimum tree sizes necessary to reject the null hypothesis in the case of identical topologies and 1:1 pathogen–host associations (i.e., a single pathogen per host and vice versa). For data sets of four through ten terminal taxa, all possible unrooted topologies were constructed using PAUP* v4.0b10 (Swofford 2002) and a dummy data matrix under the "ALLTREES" command. We then reduced these to a set of trees corresponding to all distinct unlabeled topologies. For these tests, it was only necessary that the labels were in the corresponding position in host and parasite trees, the labels themselves did not affect the results otherwise. All branch lengths were set to 1.0. AxParafit was run with 99 permutations for each case, and the probabilities of GH_0 and IH_0 were recorded.

Results

Papillomavirus and Hosts Phylogenies Are Well Resolved

The aligned aa sequence data consisted of 1,828 positions containing 827, 321, and 608 distinct alignment site patterns in each of the three partitions E1–E2–L1, respectively. Bayesian and ML trees were topologically largely congruent. Only six of 408 possible differences were detected (Robinson–Foulds distance: 0.0147; *K*-score: 0.141). Pairwise branch lengths were also similar, with a scale factor of 0.832 between both trees. Figure 1 shows the best-scoring ML tree, with ML bootstrap values (LBS) and Bayesian posterior probabilities (BPP) indicating strong, if not maximal, support for the vast majority of internal nodes. The mammalian PVs segregated into four highly supported crown groups, plus eight PVs (i.e., CPV3, CPV4, CPV5, FdPV2, MnPV1, OaPV3, RaPV1, and TmPV1), whose detailed phylogenetic relationships could not be disentangled (4% of all 207 PV-types). The host tree (supplementary fig. S1, Supplementary Material online) likewise showed high statistical support for most branches. Those few nodes with low statistical support (e.g., that linking carnivores, bats, and ungulates) reflected relationships that are known to be problematic but were not relevant for the exploration of cophylogenetic structures as conducted here.

Congruent Topologies Between Papillomaviruses and Their Hosts at Derived Phylogenetic Positions

Global congruence between PV and mammal phylogenies was not observed, with the same host (in, e.g., carnivores, primates, rodents, and ruminants) often being infected by several polyphyletic PV taxa. Beta + Xi-PVs and Lambda + Mu-PVs were particularly heterogeneous with respect to their hosts, with the species distributed throughout mammalian taxa at high taxonomic ranks, such as Laurasiatheria and Euarchontoglires. Numerous PV lineages isolated from humans, dogs, and cattle did not constitute monophyletic groups with respect to their host species but were instead scattered throughout the full viral tree in a highly polyphyletic pattern. Completely congruent topologies between PVs and their hosts (when the latter numbered more than three) were only observed in a small number of assemblages: 1) the two *Pan* PVs + HPV13 and Hominidae; 2) Pi-PV and Muridae (as far as the weakly resolved nodes in both viral and host trees allowed for this observation); 3) felid Lambda-PV and Felidae (but see below); and 4) Kappa + Mu-PV and Euarchontoglires.

At shallower levels within viral crown groups, several well-supported PV clades (LBS $>$ 95, 1.00 BPP) corresponded to their respective host taxa: Alpha-PVs and Primates; Beta-PVs and Primates; Delta + Epsilon-PVs and Ruminantia; Lambda-PVs and Carnivora; and Omikron + Epsilon-PVs and Cetacea. However, the internal topologies of the PV and host clades were partly incongruent. Some notable inconsistencies include: 1) Human PVs comprised numerous distantly related lineages and also were paraphyletic within Alpha- and Beta-PVs, respectively; 2) *Pan* and *Macaca* PVs showed derived instead of basal phylogenetic positions within Alpha- and Beta-PVs; 3) Delta + Epsilon-PVs infecting Bovidae were paraphyletic and comprised three distantly related lineages; 4) European elk PV (EEPV) (Delta-PV), isolated from *Alces alces*, was the closest relative of reindeer PV (RPV) isolated from *Rangifer tarandus*, and not of deer PV (DPV), isolated from *Odocoileus virginianus*, as would have been expected if the viral and host phylogenies were congruent; and 5) the different Cetacean PVs did not cluster in agreement with the phylogeny of their host species within Omikron + Upsilon-PVs.

Statistical Tree Reconciliation Detects a Variety of Significant and Non-significant Associations

TreeFitter

Results of the permutation test with generalized (event-cost) parsimony are shown in supplementary figure S2, Supplementary Material online. The null hypothesis was rejected ($P \leq 0.0001$) for all investigated event/cost combinations, except for those with zero switch cost and those with a switch cost of 1.0 combined with duplication costs of at least 8.0, and this resembled a hypothetical "pure cospeciation" pattern. It had, however, a higher similarity to a duplication–cospeciation pattern with two basal duplications, yielding three parasite groups, which would have codiverged with their hosts in parallel (supplementary fig. S2, Supplementary Material online and Supplementary Material online).

TreeMap

The host and the pruned virus trees, considering only a single PV representative for a monophyletic group of viruses present on a specific host, were compared via a tanglegram

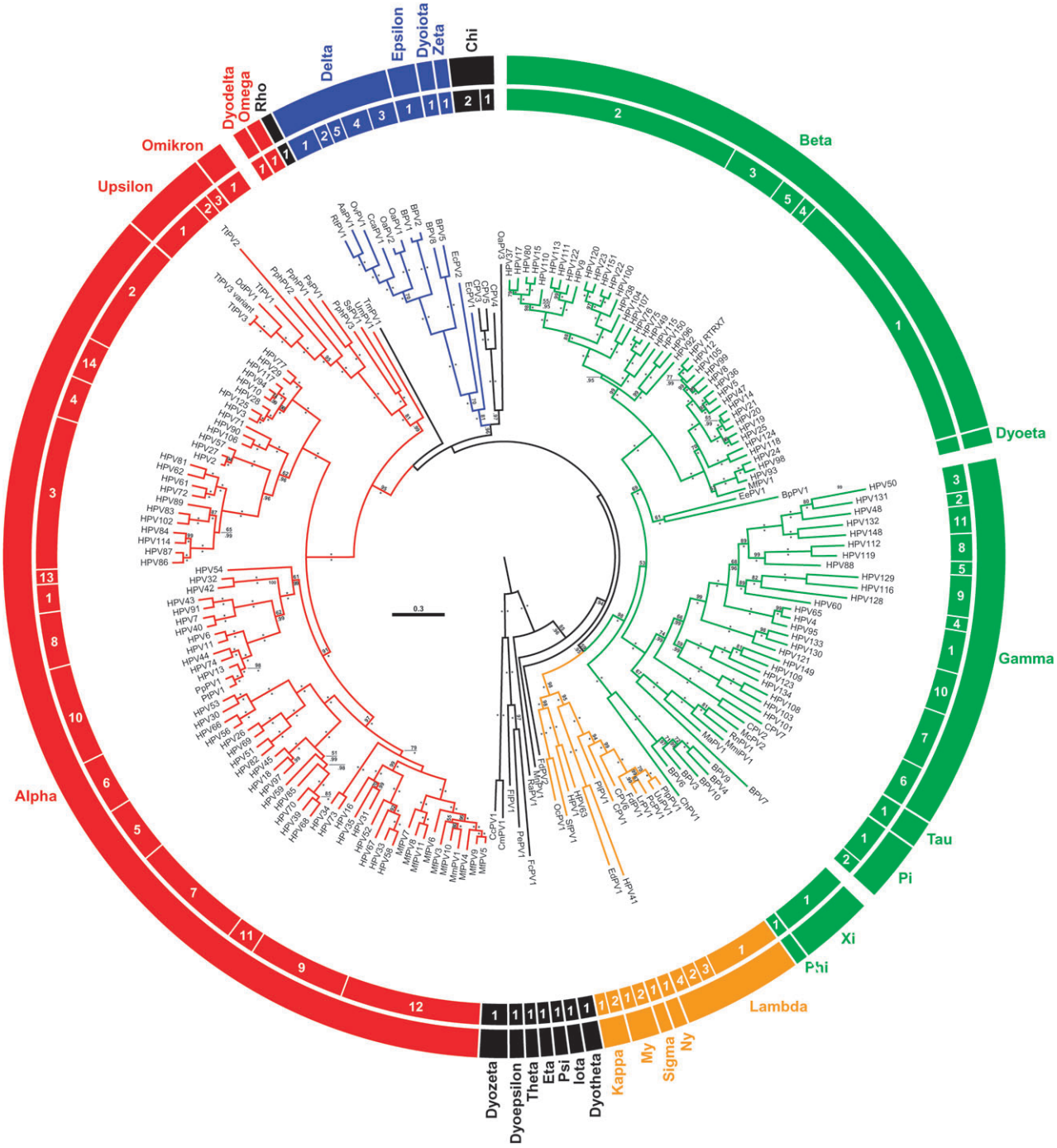

Fig. 1. The papillomavirus tree is well resolved. ML tree of 207 PVs as inferred from a combined E1–E2–L1 amino acid sequence analysis. PV taxonomic units are indicated in Greek letters ("genera") and numbers ("species": Bernard et al. 2010). The crown groups are colored red (Alpha + Omikron-PVs), green (Beta + Xi-PVs), blue (Delta + Zeta-PVs), and ochre (Lambda + Mu-PVs), respectively. Branch lengths are drawn to scale, with the scale bar indicating the expected number of amino acid substitutions per site. Numbers on branches are ML bootstrap support values (above) and Bayesian probabilities (below). Values under 50 and 0.90, respectively, are not shown, and asterisks indicate maximal support values.

in figure 2. Reconstructions identified a series of duplication events at the base of the mammalian PV tree that have given rise to the ancestors of each of the different PV crown groups. Reconciliation of the PV and the vertebrate phylogenies was also studied separately for each of the viral crown groups and their hosts (fig. 3 and table 1). In these crown group analyses, duplication events were also predominant at basal nodes of the PV tree, whereas codivergence events were observed close to the tips. The number of codivergence events was significantly larger than that for random trees ($P = 0.001$), but almost two-thirds of all events are better explained by alternative evolutionary mechanisms. Codivergence events often appeared in reconciliations for the Delta + Zeta-PV and Lambda + Mu-PV crown

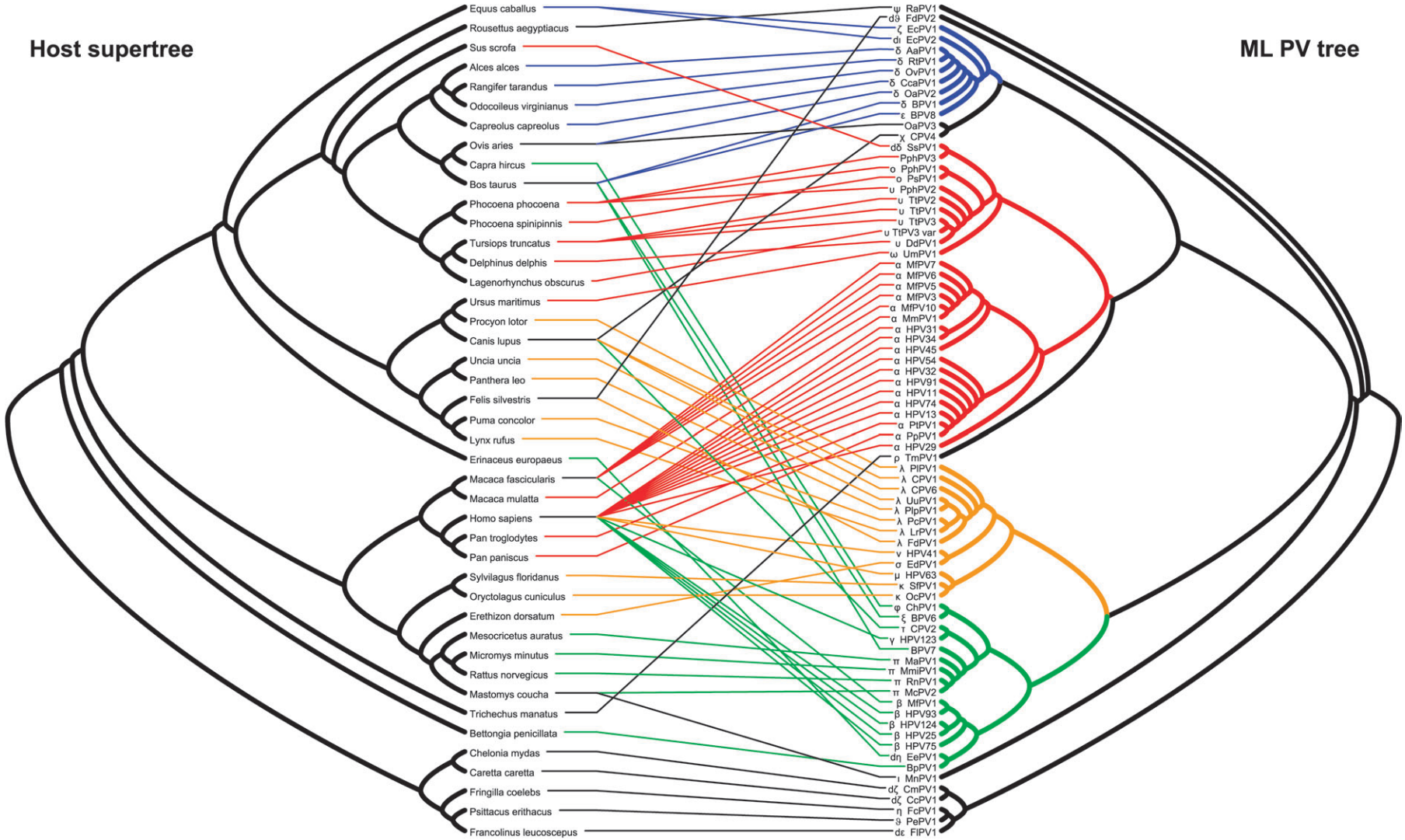

Fig. 2. Congruence between PV (right) and vertebrate (left) phylograms is rather found in the distal parts of the tree. Tanglegram (i.e., pair of binary trees, whose terminal taxa sets are in correspondence) linking the cladograms of PVs and of their hosts (the central linking lines represent the specific virus–host associations). Associations are shown disentangled using TreeMap 3.0.

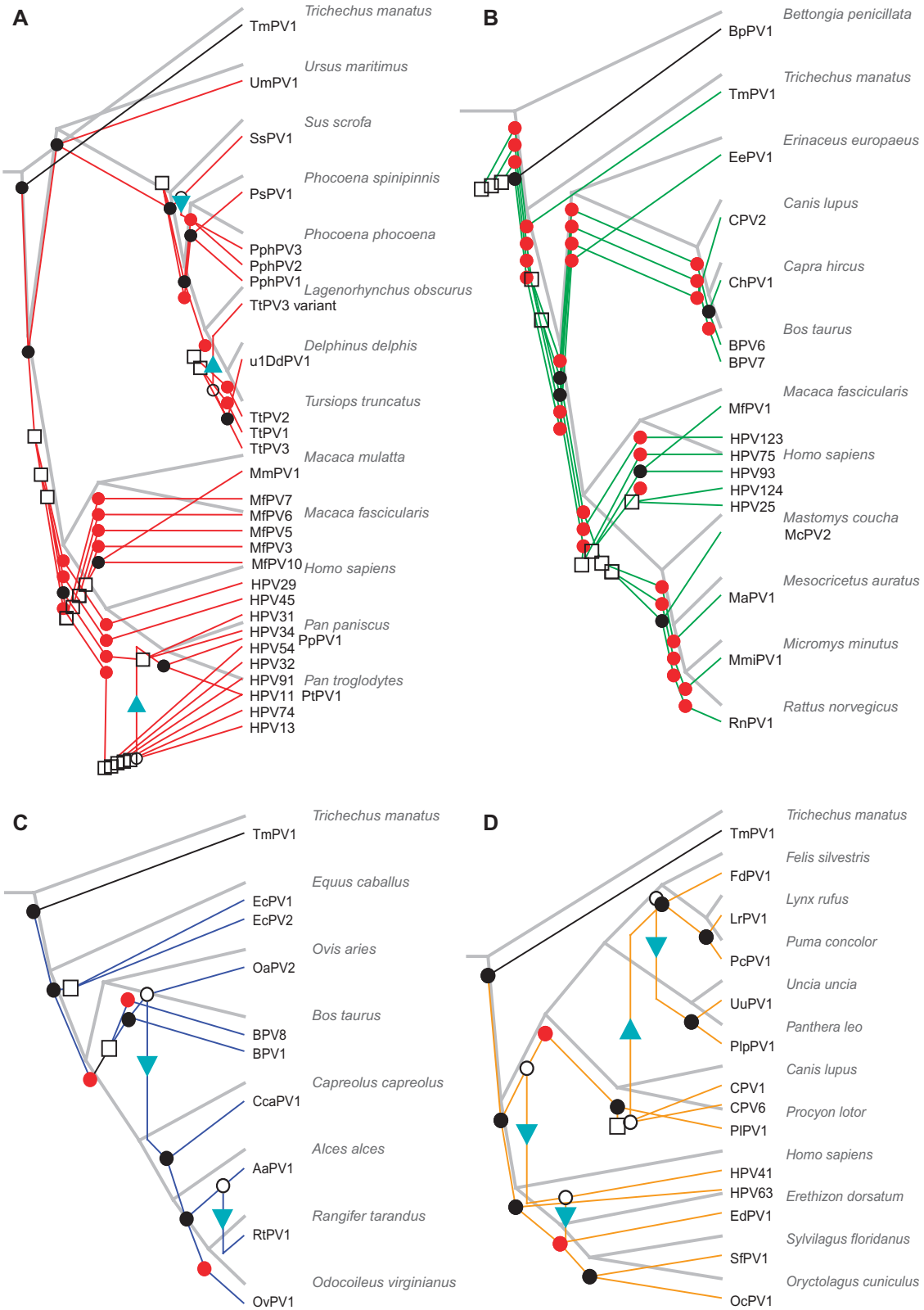

Fig. 3 Multiple evolutionary processes drive papillomavirus evolution. Reconciliation of PV (colored) and vertebrate phylogenies (gray lines, italic labels) with TreeMap 2.0, separately derived for the four PV crown groups (A, Alpha + Omikron-PVs; B, Beta + Xi-PVs; C, Delta + Zeta-PVs; and D, Lambda + Mu-PVs). The frequencies of the different evolutionary events are summarized in table 1. The symbols used for the events are: Black circles, codivergence; red circles, lineage sorting; turquoise arrowheads, switching; and white squares, duplication.

Table 1. Papillomavirus Diversification Results from Different Evolutionary Mechanisms.

	Number of Optimal Solutions	Number of Codivergence Events	Number of Noncodivergence Events	Number of Duplications	Number of Losses	Number of Host Switches
Alpha + Omikron-PVs	**169**	**6–22**	**53–60**	**36–52**	**3–27**	**0–8***
Beta + Xi-PVs	**24**	**16**	**28**	**16**	**6–8**	**4–6**
Delta + Zeta-PVs	**8**	**10**	**13–16**	**8**	**2–8**	**0–3**
Lambda + Mu-PVs	**11**	**12–18**	**14–20**	**8–14**	**0–8**	**0–6**
Total 4 crown groups	**222**	**34–64**	**108–124**	**68–90**	**11–51**	**4–23**

Frequencies of the different evolutionary events per optimal solution invoked for reconciliation of PVs and host phylogenetic trees, using TreeMap 2.0, for each of the four crown groups separately.

* Only solutions with a maximum of eight host switches were explored.

groups and accounted for almost half of all such events across PVs. Codivergence events contributed less to reconciliations of the Alpha + Omikron-PV and Beta + Xi-PV crown groups with their larger host diversity.

ParaFit

We explored the minimum data set size necessary to reject either global or individual null hypotheses based on topology comparison (supplementary fig. S3, Supplementary Material online): For testing a global link, more than four associations were necessary (supplementary fig. S3*a*, Supplementary Material online), whereas this number rose to seven when testing individual links, even in the case of 1:1 associations, identical topologies, and uniform branch lengths (supplementary fig. S3*b*, Supplementary Material online). Accordingly, none of the data sets contained less than four associations and less than eight PVs in the biological analyses.

ParaFit analyses using viral ML and Bayesian trees yielded comparable results. For all PVs, GH_0 was rejected ($P = 0.0001$), indicating that PV evolution can be at least partly explained by codivergence with the hosts they infect. However, only 90 of the 207 individual associations (43%) had a significantly larger impact than random associations on the global congruence between the host and parasite trees (supplementary table S2, Supplementary Material online). Of the 43 vertebrate species contributing to the PV–host associations, 26 (60%) were not among the significant associations (increasing to 16/24 at the hosts family level, 67%). Thus, considering the full PV diversity, a considerable number of associations were apparently incongruent with the host phylogeny. The ParaFit analyses on the pruned trees (considering only a single PV representative per monophyletic viral lineage on a specific host) yielded similar results to the comprehensive tree analysis (supplementary table S2, Supplementary Material online). Twenty-nine of 43 (67%) host species were not among the significant PV–host associations, representing 17 of 24 (71%) host taxa at the family level.

In each of the four analyses for the well-supported PV crown groups, GH_0 was globally rejected with the exception of the Beta + Xi-PVs (See Supplementary Material online). Thus, PV evolution within the crown groups is frequently not independent of the host phylogeny. A significant global test also alleviated the interpretation of the tests of the individual links, which could have inflated type I errors if GH_0 was accepted. Individual codivergence hypotheses IH_0 in the data set composed of Alpha + Omikron-PVs for Carnivora, Cetacea, Primates, and Suina were rejected for all associations (only Alpha-PVs in the full tree analysis). Moreover, IH_0 was rejected for all associations with the exception of *Macaca* PVs in separate Alpha-PV analyses. Within the Beta + Xi-PVs, significant associations included Xi-PVs, BqPV1, and EdPV1 (none in the full tree analysis). For GH_0 of the Delta + Zeta-PVs with Perissodactyla and Ruminantia, we obtained significant associations for the two horse PVs at the base of the viral tree and their host *Equus caballus* (Linnaeus 1758), exactly complementary to the full tree. Finally, individual significant associations within Lambda + Mu-PV analyses were present for felid PVs, Kappa-, and Mu-PVs (none in the full tree analysis).

Only the associations involving the Alpha-PVs and PsPV1 were significant in all analyses. By contrast, some PVs belonged to associations that were never significant, namely the orphan MnPV1, RaPV1, TmPV1, and all canine PVs as well as Beta- and Gamma-PVs from the Beta + Xi-PV crown group and PlPV1 and HPV41 from the Lambda + Mu-PV crown group.

Discussion

The Importance of Codivergence Between Papillomaviruses and Their Hosts

The phylogenetic analysis of 207 completely sequenced PVs based on the E1–E2–L1 gene combination yields a topology with four well-supported PV crown groups, each of which corresponds to certain mammalian host taxa. This result confirms and expands previous studies about PV phylogeny and evolution (Chan et al. 1995; García-Vallvé et al. 2005; Gottschling, Stamatakis et al. 2007; Rector et al. 2007; Bravo et al. 2010). It is worth noting the high statistical support for the close relationships of both Alpha + Omikron-PVs as well as of Delta + Zeta-PVs. This result might be due to our analyses of the largest possible PV taxon sample to date, including all known types, given that taxon sampling has repeatedly been shown to be an important prerequisite for improved resolution of organismal phylogenetic trees (Dunn et al. 2008; Heath et al. 2008; Sanderson 2008).

The assumption of a broad and general codivergence between PVs and their hosts has been historically inferred from related PV-types that appear to infect related host species (Van Ranst et al. 1995; Halpern 2000; Bernard 2005; Shah et al. 2010). However, a rigorous evaluation

of cophylogenetic relationships between PV and vertebrate trees had not been conducted so far. Our results show that virus–host codivergence plays an important evolutionary role in PVs, but it is not the only event that has shaped PV diversification. The ParaFit analyses indicate that more than half of the associations between PV and hosts are equally well explained by chance. Furthermore, TreeMap results attribute only one-third of all evolutionary events to codivergence, with both TreeMap and TreeFitter analyses indicating that alternative evolutionary processes such as duplications, lineage sorting, and switches also play important roles in PV evolution (see below).

There are no unambiguous examples of strict virus–host codivergence in our analyses. Nonrandom virus–hosts associations can be identified only for assemblages of more than four related viruses infecting more than four related hosts. In our sequence data set, only four PV "genera" encompass more than three different host species: Alpha-PVs infecting primates, Delta-PVs infecting ruminants, Lambda-PVs infecting carnivores, and Pi-PVs infecting rodents. The topologies of Alpha- and Delta-PVs are clearly inconsistent with their corresponding host trees. ParaFit analyses support this observation because several PV–host associations do not differ significantly from random. Global congruence between virus and host trees depends strongly on Lambda-PVs infecting carnivores (Rector et al. 2007), but even this case is not unambiguous because of the poorly resolved phylogeny of Felidae. The felid phylogeny has long been a matter of debate, and the molecular tree of the felid hosts provided by (Rector et al. 2007) neither agrees with a comprehensive molecular phylogeny of cats (Johnson et al. 2006) nor with the phylogeny obtained in this study. The internal phylogeny of Pi-PVs is also not entirely congruent to the rodent host tree topology (Schulz et al. 2009). A solid case of cophylogenetic relationships between PVs and their mammalian hosts supported by identical tree topologies, even on a local scale, is thus still wanting.

We have analyzed global PV–host codivergence at two different levels by testing both the comprehensive PV tree and the four well-supported large crown groups. The fact that the ParaFit results differ between the two analyses with respect to the global hypothesis is not due to inconsistencies of the method, but rather to the underlying shift of scale of the null hypothesis for each data set. ParaFit tests the impact of each association on the "global" fit between the trees (Legendre et al. 2002). Thus, congruent host–parasite subtrees that are significant when considered separately can occur at incompatible positions in the global analysis, potentially rendering the corresponding associations nonsignificant at this more inclusive level. A cogent example is provided by the cetacean PVs and Lambda-PVs. Both sets of associations are significant when considering the Alpha + Omikron- and the Lambda + Mu-PV crown groups, respectively but become significant when using the comprehensive ML and Bayesian trees. The same argument applies in reverse for internally incongruent subtrees that appear at compatible positions in the global trees and may therefore be significant. This is the case for Alpha-PVs, whose associations were identified as being significant in the comprehensive trees as they encompass viruses infecting primates but nonsignificant within each of the crown groups where they occur. Additional examples of this scaling behavior are found in the analyses of the other PV crown groups (fig. 2).

We interpret the evidence for congruent subtrees at incompatible positions in the full tree as support for the hypothesis that early PV radiation has been followed by independent codivergence between viruses of the major crown groups and their hosts (García-Vallvé et al. 2005; Gottschling, Stamatakis et al. 2007; Bravo et al. 2010). Such biphasic evolution is also supported by the results of the TreeMap analysis, which show an accumulation of duplication events at basal nodes of the global PV tree (fig. 3) and a strong resemblance to a simulated scenario of duplication followed by codivergence (supplementary fig. S2, Supplementary Material online). Thus, noncodivergence at the base of the PV seems justified, but the exact form of the evolutionary events remains somewhat unclear as the global associations for the complete PV and host trees could only be tested by restricting the number of host switch events. This may have inflated the postulated number of duplication events because given similar topologies, reconstructions yielding the same maximum codivergence score will either comprise few switches or few duplications and sorting events. Overall, results of the TreeFitter analyses indicate that host switching of PVs is less frequent than expected by chance, otherwise zero switch costs would not result in accepting the null hypothesis.

Additional Evolutionary Mechanisms Driving Papillomavirus Diversification

The radiation of vertebrates has undoubtedly influenced PV evolution. However, our increased knowledge about PV diversity, coupled with the use of well-supported molecular trees, suggests that evolutionary mechanisms other than codivergence have also significantly contributed to PV diversification. Mammalian species are infected not only by a single PV type but most probably by myriads of PV types. This fact is best documented for humans where more than 100 different PV genomes have been sequenced (Bernard et al. 2010) but also for other hosts, such as dogs and cows. This high intraspecies diversity cannot be explained by strict codivergence alone (Lyal 1986), but alternative mechanisms including adaptive radiations ("duplications"), and the realization of new ecological niches (García-Vallvé et al. 2005; Gottschling, Stamatakis et al. 2007) also need to be considered to explain the numerous, paralogous PV lineages on the same host species (Jackson 2005). This general scenario is supported by all statistical tests in this study, which indicate that codivergence rather occurred at a more local scale, resulting in multiple infections of a host lineage by different PV types.

Papillomaviruses infecting the same host species are not only diverse but frequently appear to be paraphyletic or polyphyletic. A plausible hypothesis to explain such topologies are cross-infections between different host species (Myers

et al. 1996; Chen et al. 2009), although PVs are traditionally considered to be host specific. The similarity of licenses between the hosts because of their close evolutionary relationship may facilitate viral infections of a new host, albeit at low reproduction rates (Turner and Elena 2000; Wolfe et al. 2007). Similar licenses could also arise in distantly related hosts from dramatic changes in host behavior, as it has been the case during human evolution (e.g., change from a nomadic lifestyle, animal and plant domestication: Woolhouse et al. 2005; Hoberg and Brooks 2008).

The common assumption of host specificity in PVs persists in spite of experimental evidence for host multiplicity. The cottontail rabbit SfPV1 isolated from *Sylvilagus floridanus* can infect the domestic rabbit *Oryctolagus cuniculus* (Amella et al. 1994; Salmon et al. 2000; Campo 2002) despite some abortive infections in which the productive development of the virus is not completed. Another example of host plasticity includes BPV1 and -2 (Delta-PVs), isolated primarily from lesions in cattle. These viruses infect closely related species including water buffaloes (Silvestre et al. 2009) but also more distantly related species, such as tapirs (Kidney and Berrocal 2008), horses (Bogaert et al. 2008), and zebras living either in zoos (Löhr et al. 2005) or in the wild (van Dyk et al. 2009). It remains to be determined whether the ability of BPV1 to infect distantly related hosts is an ancient or a more recent phenotypic acquisition that is driven by vector-mediated interspecies transmission (Finlay et al. 2009) and/or putatively linked to the human domestication of cattle and horses. In the latter case, the concomitant ecological changes in the different hosts may have increased their susceptibility to BPV1 cross-infection and/or have simply increased the frequency of physical contact between them to grant BPV1 improved access to a potential new host.

Impact of Sampling Error

The outcome of statistical tests as applied here strongly depends on tree size and on the amount of information near the tips, which in our case represents the host diversity (Page et al. 1996; Legendre et al. 2002). The continuous discovery of novel PVs suggests their widespread occurrence in vertebrates. However, there is an inherent bias in taxon sampling, arising from the historical interest in human medicine. Other host species such as cattle or dogs appear to comprise a similar diversity, but we lack information about PVs infecting other major lineages within mammals. For many other mammalian hosts, only single sequences are available (e.g., the chiropteran RaPV1 or the sirenian TmPV1). Our knowledge of PV diversity (particularly from field research) remains definitely fragmentary.

In addition to potentially influencing the accuracy of the inferred PV tree, our incomplete knowledge of PV diversity may influence the evolutionary inferences based on that tree. The lack of close viral relatives can hinder the reliable resolution of deep nodes through reconstruction artifacts such as long-branch attraction. For instance, the "real" absence of a specific PV associated with a host (e.g., the "absence" of a *Pongo pygmaeus*, Linnaeus 1760, PV within Alpha-PVs) could derive from sorting events (Paterson and Banks 2001) including the extinction of the ancestral PV in the host lineage ("drowning on arrival") or the absence of the ancestral PV in the host's ancestor ("missing the boat"). The majority of the 50 duplication events identified by TreeMap are probably true evolutionary events, but many of the inferred sorting events might reflect simply our lack of knowledge about viral diversity due to unbalanced sampling toward medically important PVs.

Conclusion

The presence of PVs in diverse vertebrates, together with the local congruence between virus and host phylogenies, indicates that codivergence is an important phylodynamic force of PV evolution. However, this mechanism alone cannot explain the origin of PV diversity, and barely half of the known host–parasite associations appear to derive from cophylogenetic events. Alternative mechanisms such as within-host virus duplication, viral sorting, or viral adaptation after a host switch, may therefore contribute considerably to PV diversification. Finally, our knowledge about PV evolution cannot improve without a more thorough and systematic broad sampling of PV diversity. Understanding the underlying evolutionary mechanisms of PV speciation via improved nonhuman taxon sampling will allow us to better address medically important questions, such as the putative presence of different PV niches or targets within the human skin, and the time of origin of PVs causing malignant transformations. Based on this improved knowledge, it is crucial to determine whether transforming mechanisms present in only distantly related viral lineages exemplify convergent evolution or rather the exploitation of different breaches in the host immune response.

Supplementary Material

Supplementary material, tables S1 and S2, and figures S1–S3 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

I.G.B. is funded by the Spanish Ministerio de Ciencia e Innovación, MICINN (Ramón y Cajal program, CGL2009-08350 and CGL2010-16713). A.S. was funded under the auspices of the Emmy-Nother program by the German Science Foundation (DFG). Help from G. Grimm (Naturhistorika Rijksmuseet, Stockholm), with Supplementary figure S2, Supplementary Material online is gratefully acknowledged.

References

Abascal F, Zardoya R, Posada D. 2005. ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* 21:2104–2105.

Amella CA, Lofgren LA, Ronn AM, Nouri M, Shikowitz MJ, Steinberg BM. 1994. Latent infection induced with cottontail rabbit papillomavirus. A model for human papillomavirus latency. *Am J Pathol.* 144:1167–1171.

Antia R, Regoes RR, Koella JC, Bergstrom CT. 2003. The role of evolution in the emergence of infectious diseases. *Nature* 426:658–661.

Bamford DH, Grimes JM, Stuart DI. 2005. What does structure tell us about virus evolution? *Curr Opin Struct Biol*. 15:655–663.

Begerow D, Göker M, Lutz M, Stoll M. 2004. About the evolution of smut fungi on their hosts. In: Agerer R, Blanz P, Piepenbring M, editors. Frontiers in basidiomycote mycology. Eching (Germany): IHW. p. 81–98.

Bernard H-U. 2005. The clinical importance of the nomenclature, evolution and taxonomy of human papillomaviruses. *J Clin Virol*. 32S:S1–S6.

Bernard H-U, Burk RD, Chen Z, van Doorslaer K, zur Hausen H, de Villiers E-M. 2010. Classification of papillomaviruses (PVs) based on 189 PV types and proposal of taxonomic amendments. *Virology* 401:70–79.

Bernard H-U, Calleja-Macias IE, Dunn ST. 2006. Genome variation of human papillomavirus types: phylogenetic and medical implications. *Int J Cancer*. 118:1071–1076.

Bininda-Emonds ORP, Cardillo M, Jones KE, MacPhee RDE, Beck RMD, Grenyer R, Price SA, Vos RA, Gittleman JL, Purvis A. 2007. The delayed rise of present-day mammals. *Nature* 446:507–512.

Bogaert L, Martens A, Van Poucke M, Ducatelle R, De Cock H, Dewulf J, De Baere C, Peelman L, Gasthuys F. 2008. High prevalence of bovine papillomaviral DNA in the normal skin of equine sarcoid-affected and healthy horses. *Vet Microbiol*. 129:58–68.

Bravo IG, de Sanjosé S, Gottschling M. 2010. The clinical importance of understanding the evolution of papillomaviruses. *Trends Microbiol*. 18:432–438.

Breitbart M, Rohwer F. 2005. Here a virus, there a virus, everywhere the same virus? *Trends Microbiol*. 13:278–284.

Campo MS. 2002. Animal models of papillomavirus pathogenesis. *Virus Res*. 89:249–261.

Chan S-Y, 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.

Charleston MA. 1998. Jungles: a new solution to the host/parasite phylogeny reconciliation problem. *Math Biosci*. 149:191–223.

Chen Z, van Doorslaer K, de Salle R, Wood CE, Kaplan JR, Wagner JD, Burk RD. 2009. Genomic diversity and interspecies host infection of alpha12 Macaca fascicularis papillomaviruses (MfPVs). *Virology*. 393:304–310.

Dennehy JJ, Friedenberg NA, Holt RD, Turner PE. 2006. Viral ecology and the maintenance of novel host use. *Am Nat*. 167:429–439.

Dunn CW, Hejnol A, Matus DQ, et al. (15 co-authors). 2008. Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature* 452:745–749.

Elena SF, Agudelo-Romero P, Lalic J. 2009. The evolution of viruses in multi-host fitness landscapes. *Open Virol J*. 3:1–6.

Fahrenholz H. 1913. Ectoparasiten und Abstammungslehre. *Zool Anz*. 41:371–374.

Finlay M, Yuan Z, Burden F, Trawford A, Morgan IM, Campo MS, Nasir L. 2009. The detection of Bovine Papillomavirus type 1 DNA in flies. *Virus Res*. 144:315–317.

Gao F, Bailes E, Robertson DL, Chen Y, et al. (9 co-authors). 1999. Origin of HIV-1 in the chimpanzee *Pan troglodytes troglodytes*. *Nature*. 397:436–441.

García-Vallvé S, Alonso A, Bravo IG. 2005. Papillomaviruses: different genes have different histories. *Trends Microbiol*. 13:514–521.

Ghedin E, Sengamalay NA, Shumway M, et al. (16 co-authors). 2005. Large-scale sequencing of human influenza reveals the dynamic nature of viral genome evolution. *Nature* 437:1162–1166.

Gibbs AJ, Armstrong JS, Downie JC. 2009. From where did the 2009 'swine-origin' influenza A virus (H1N1) emerge? *Virol J*. 6:207.

Gissmann L, zur Hausen H. 1980. Partial characterization of viral DNA from human genital warts (*Condylomata acuminata*). *Int J Cancer*. 25:605–609.

Gottschling M, Bravo IG, Schulz E, Bracho Lapiedra MA, Deaville R, Jepson PD, Van Bressem M-F, Stockfleth E, Nindl I. 2010. Modular organizations of novel cetacean papillomaviruses. *Mol Phylogenet Evol*. doi:10.1016/j.ympev.2010.12.013.

Gottschling M, Stamatakis A, Nindl I, Stockfleth E, Alonso A, Bravo IG. 2007. Multiple evolutionary mechanisms drive papillomavirus diversification. *Mol Biol Evol*. 24:1242–1258.

Gottschling M, Wibbelt G, Wittstatt U, Stockfleth E, Nindl I. 2008. Novel papillomavirus isolates from *Erinaceus europaeus* (Erinaceidae, Insectivora) and the Cervidae (Artiodactyla), *Cervus timorensis* and *Pudu puda*, and phylogenetic analysis of partial sequence data. *Virus Genes*. 36:281–287.

Halpern AL. 2000. Comparison of papillomavirus and immunodeficiency virus evolutionary patterns in the context of a papillomavirus vacine. *J Clin Virol*. 19:43–56.

Hay AJ, Gregory V, Douglas AR, Lin YP. 2001. The evolution of human influenza viruses. *Philos Trans R Soc Lond B Biol Sci*. 356:1861–1870.

Heath TA, Zwickl DJ, Kim J, Hillis DM. 2008. Taxon sampling affects inferences of macroevolutionary processes from phylogenetic trees. *Syst Biol*. 57:160–166.

Heeney JL. 2006. Zoonotic viral diseases and the frontier of early diagnosis, control and prevention. *J Intern Med*. 260:399–408.

Heeney JL, Dalgleish AG, Weiss RA. 2006. Origins of HIV and the evolution of resistance to AIDS. *Science* 313:462–466.

Hoberg EP, Brooks DR. 2008. A macroevolutionary mosaic: episodic host-switching, geographical colonization and diversification in complex host-parasite systems. *J Biogeogr*. 35:1533–1550.

Holmes EC, Drummond AJ. 2007. The evolutionary genetics of viral emergence. In: Childs JE, Mackenzie JS, Richt JA, editors. Wildlife and emerging zoonotic diseases: the biology, circumstances and consequences of cross-species transmission. Berlin (Germany): Springer. p. 51–66.

Jackson AP. 2005. The effect of paralogous lineages on the application of reconciliation analysis by cophylogeny mapping. *Syst Biol*. 54:127–145.

Johnson KP, Adams RJ, Page RD, Clayton DH. 2003. When do parasites fail to speciate in response to host speciation? *Syst Biol*. 52:37–47.

Johnson WE, Eizirik E, Pecon-Slattery J, Murphy WJ, Antunes A, Teeling E, O'Brien SJ. 2006. The late Miocene radiation of modern Felidae: a genetic assessment. *Science*. 311:73–77.

Katoh K, Kuma K, Toh H, Miyata T. 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res*. 33:511–518.

Kay A, Zoulim F. 2007. Hepatitis B virus genetic variability and evolution. *Virus Res*. 127:164–176.

Kidney BA, Berrocal A. 2008. Sarcoids in two captive tapirs (*Tapirus bairdii*): clinical, pathological and molecular study. *Vet Dermatol*. 19:380–384.

Klassen GJ. 1992. Coevolution: a history of the macroevolutionary approach to studying host-parasite associations. *J Parasitol*. 78:573–587.

Lartillot N, Philippe H. 2004. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol Biol Evol*. 21:1095–1109.

Le SQ, Gascuel O. 2008. An improved general amino acid replacement matrix. *Mol Biol Evol*. 25:1307–1320.

Legendre P, Desdevises Y, Bazin E. 2002. A statistical test for host-parasite coevolution. *Syst Biol*. 51:217–234.

Löhr CV, Juan-Sallés C, Rosas-Rosas A, Parás García A, Garner MM, Teifke JP. 2005. Sarcoids in captive zebras (*Equus burchellii*): association with bovine papillomavirus type 1 infection. *J Zoo Wildl Med*. 36:74–81.

Lyal CHC. 1986. Coevolutionary relationships of lice and their hosts: a test of Fahrenholz's rule. In: Stone AR, Hawksworth DL, editors. Coevolution and systematics. Oxford: Clarendon Press. p. 77–91.

Mahy BW, Brown CC. 2000. Emerging zoonoses: crossing the species barrier. *Rev Sci Tech*. 19:33–40.

Mansfield KL, Johnson N, Phipps LP, Stephenson JR, Fooks AR, Solomon T. 2009. Tick-borne encephalitis virus—a review of an emerging zoonosis. *J Gen Virol*. 90:1781–1794.

McGeoch DJ, Dolan A, Ralph AC. 2000. Toward a comprehensive phylogeny for mammalian and avian herpesviruses. *J Virol*. 74:10401–10406.

Meier-Kolthoff JP, Auch AF, Huson DH, Göker M. 2007. CopyCat: co-phylogenetic analysis tool. *Bioinformatics* 23:898–900.

Mindell DP, Rest JS, Villarreal LP. 2004. Viruses and the tree of life. In: Cracraft J, Donoghue MJ, editors. Assembling the tree of life. New York: Oxford University Press. p. 107–118.

Mühlemann K. 2004. The molecular epidemiology of mumps virus. *Infect Genet Evol*. 4:215–219.

Muñoz N, Bosch FX, de Sanjosé S, Herrero R, Castellsague X, Shah KV, Snijders PJF, Meijer CJLM. 2003. Epidemiologic classification of human papillomavirus types associated with cervical cancer. *N Engl J Med*. 348:518–527.

Murphy WJ, Eizirik E, O'Brien SJ, et al. (8 co-authors). 2001. Resolution of the early placental mammal radiation using Bayesian phylogenetics. *Science* 294:2348–2351.

Myers G, Lu H, Calef C, Leitner T. 1996. Heterogeneity of papillomaviruses. *Semin Cancer Biol*. 7:349–358.

Narechania A, Chen Z, DeSalle R, Burk RD. 2005. Phylogenetic incongruence among oncogenic genital alpha human papillomaviruses. *J Virol*. 79:15503–15510.

Nunn CL. 2004. Parasites and the evolutionary diversification of primate clades. *Am Nat*. 164(5 Suppl):S90–S103.

Osche G. 1966. Grundzüge der allgemeinen Phylogenetik. In: Gessner F, editor. Handbuch der Biologie III, 2. Frankfurt/M (Germany): Athenaion. p. 817–906.

Page RDM. 1994. Parallel phylogenies: reconstructing the history of host-parasite assemblages. *Cladistics* 10:155–173.

Page RDM, Charleston MA. 1998. Trees within trees: phylogeny and historical associations. *Trends Ecol Evol*. 13:356–359.

Page RDM, Clayton DH, Paterson AM. 1996. Lice and cospeciation: a response to Barker. *Int J Parasitol*. 26:213–218.

Paterson AM, Banks J. 2001. Analytical approaches to measuring cospeciation of host and parasites: through a glass, darkly. *Int J Parasitol*. 31:1012–1022.

Plantier JC, Leoz M, Dickerson JE, De Oliveira F, Cordonnier F, Lemee V, Damond F, Robertson DL, Simon F. 2009. A new human immunodeficiency virus derived from gorillas. *Nat Med*. 15:871–872.

Puigbo P, Bravo IG, Garcia-Vallvé S. 2008. E-CAI: a novel server to estimate an expected value of Codon Adaptation Index (eCAI). *BMC Bioinformatics*. 9:65.

Rector A, Lemey P, Tachezy R, et al. (12 co-authors). 2007. Ancient papillomavirus-host co-speciation in Felidae. *Genome Biol*. 8:R57.

Rector A, Stevens H, Lacave G, et al. (10 co-authors). 2008. Genomic characterization of novel dolphin papillomaviruses provides indications for recombination within the *Papillomaviridae*. *Virology* 378:151–161.

Reyes A, Gissi C, Catzeflis F, Nevo E, Pesole G, Saccone C. 2004. Congruent mammalian trees from mitochondrial and nuclear genes using Bayesian methods. *Mol Biol Evol*. 21:397–403.

Ronquist F. 1995. Reconstructing the history of host-parasite associations using generalised parsimony. *Cladistics*. 11:73–89.

Ronquist F. 2003. Parsimony analysis of coevolving species associations. In: Page RD, editor. Tangled trees: phylogeny, cospeciation, and coevolution. Chicago (IL): University of Chicago Press. p. 22–64.

Russell CA, Jones TC, Barr IG, et al. (25 co-authors). 2008. The global circulation of seasonal influenza A (H3N2) viruses. *Science*. 320:340–346.

Salmon J, Nonnenmacher M, Caze S, Flamant P, Croissant O, Orth G, Breitburd F. 2000. Variation in the nucleotide sequence of cottontail rabbit papillomavirus a and b subtypes affects wart regression and malignant transformation and level of viral replication in domestic rabbits. *J Virol*. 74: 10766–10777.

Sanderson MJ. 2008. Phylogenetic signal in the eukaryotic tree of life. *Science*. 321:121–123.

Schulz E, Gottschling M, Wibbelt G, Stockfleth E, Nindl I. 2009. Isolation and genomic characterization of the first Norway rat (*Rattus norvegicus*) papillomavirus and its phylogenetic position within *Pipapillomavirus*, primarily infecting rodents. *J Gen Virol*. 90:2609–2614.

Shah SD, Doorbar J, Goldstein RA. 2010. Analysis of host–parasite incongruence in papillomavirus evolution using importance sampling. *Mol Biol Evol*. 27:1301–1314.

Silvestre O, Borzacchiello G, Nava D, Iovane G, Russo V, Vecchio D, D'Ausilio F, Gault EA, Campo MS, Paciello O. 2009. Bovine papillomavirus type 1 DNA and e5 oncoprotein expression in water buffalo fibropapillomas. *Vet Pathol*. 46:636–641.

Soria-Carrasco V, Talavera G, Igea J, Castresana J. 2007. The K tree score: quantification of differences in the relative branch length and topology of phylogenetic trees. *Bioinformatics*. 23:2954–2956.

Springer MS, Stanhope MJ, Madsen O, de Jong WW. 2004. Molecules consolidate the placental mammal tree. *Trends Ecol Evol*. 19:430–438.

Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*. 22:2688–2690.

Stamatakis A, Auch AF, Meier-Kolthoff J, Göker M. 2007. AxPcoords & parallel AxParafit: statistical co-phylogenetic analyses on thousands of taxa. *BMC Bioinformatics*. 8:405.

Stevens J. 2004. Computational aspects of host-parasite phylogenies. *Brief Bioinform*. 5:339–349.

Swofford DL. 2002. PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4. Sunderland (MA): Sinauer Associates.

Turner PE, Elena SF. 2000. Cost of host radiation in an RNA virus. *Genetics*. 156:1465–1470.

van Dyk E, Oosthuizen MC, Bosman AM, Nel PJ, Zimmerman D, Venter EH. 2009. Detection of bovine papillomavirus DNA in sarcoid-affected and healthy free-roaming zebra (*Equus zebra*) populations in South Africa. *J Virol Methods*. 158:141–151.

Van Ranst M, Kaplan JB, Sundberg JP, Burk RD. 1995. Molecular evolution of papillomaviruses. In: Gibbs A, Calisher CH, García-Arenal F, editors. Molecular basis of virus evolution. Cambridge: Cambridge University Press. p. 455–476.

Varsani A, van der Walt E, Heath L, Rybicki EP, Williamson AL, Martin DP. 2006. Evidence of ancient papillomavirus recombination. *J Gen Virol*. 87:2527–2531.

Weingartl HM, Berhane Y, Czub M. 2009. Animal models of henipavirus infection: a review. *Vet J*. 181:211–220.

Wolfe ND, Dunavan CP, Diamond J. 2007. Origins of major human infectious diseases. *Nature* 447:279–283.

Woolhouse ME, Haydon DT, Antia R. 2005. Emerging pathogens: the epidemiology and evolution of species jumps. *Trends Ecol Evol*. 20:238–244.

Woolhouse ME, Howey R, Gaunt E, Reilly L, Chase-Topping M, Savill N. 2008. Temporal trends in the discovery of human viruses. *Proc R Soc B Biol Sci*. 275:2111–2115.

Zhou Y, Ushijima H, Frey TK. 2007. Genomic analysis of diverse rubella virus genotypes. *J Gen Virol*. 88:932–941.

zur Hausen H. 2002. Papillomaviruses and cancer: from basic studies to clinical application. *Nat Rev Cancer*. 2:342–350.