Virology Journal

Research

Evidence of recombination within human alpha-papillomavirus

Manuel Angulo[1] and Antonio Carvajal-Rodríguez*[2]

Address: [1]Department of pathology. Complejo Hospitalario Universitario de Vigo, 36204 Vigo, Spain and [2]Departamento de Bioquímica, Genética e Inmunología. Universidad de Vigo, 36310 Vigo, Spain

Email: Manuel Angulo - mangcou1@sergas.es; Antonio Carvajal-Rodríguez* - acraaj@uvigo.es

* Corresponding author

Abstract

Background: Human papillomavirus (HPV) has a causal role in cervical cancer with almost half a million new cases occurring each year. Presence of the carcinogenic HPV is necessary for the development of the invasive carcinoma of the genital tract. Therefore, persistent infection with carcinogenic HPV causes virtually all cervical cancers. Some aspects of the molecular evolution of this virus, as the putative importance of recombination in its evolutionary history, are an opened current question. In addition, recombination could also be a significant issue nowadays since the frequency of co-infection with more than one HPV type is not a rare event and, thus, new recombinant types could be currently being generated.

Results: We have used human alpha-PV sequences from the public database at Los Alamos National Laboratory to report evidence that recombination may exist in this virus. A model-based population genetic approach was used to infer the recombination signal from the HPV DNA sequences grouped attending to phylogenetic and epidemiological information, as well as to clinical manifestations. Our results agree with recently published ones that use a different methodology to detect recombination associated to the gene L2. In addition, we have detected significant recombination signal in the genes E6, E7, L2 and L1 at different groups, and importantly within the high-risk type HPV16. The method used has recently been shown to be one of the most powerful and reliable procedures to detect the recombination signal.

Conclusion: We provide new support to the recent evidence of recombination in HPV. Additionally, we performed the recombination estimation assuming the best-fit model of nucleotide substitution and rate variation among sites, of the HPV DNA sequence sets. We found that the gene with recombination in most of the groups is L2 but the highest values were detected in L1 and E6. Gene E7 was recombinant only within the HPV16 type. The topic deserves further study because recombination is an important evolutionary mechanism that could have high impact both in pharmacogenomics (i.e. on the influence of genetic variation on the response to drugs) and for vaccine development.

Background

Presence of the carcinogenic human papillomavirus (HPV) is necessary for the development of the invasive carcinoma of the genital tract [1]. Persistent infection with carcinogenic HPV causes virtually all cervical cancers. According to the latest global estimates this cancer is the

second most common in women after breast cancer [2]. Almost half a million new cases of cervical cancer occur each year among women world-wide causing 274.000 deaths, 85% of them happening in underdevelopment countries[3].

The HPV genome has three different regions: two coding (E – early and L – late expression) and a regulatory non-coding region. The early region codifies regulatory, transforming and replication proteins, among which E6 and E7 are known to act like oncoproteins in high risk virus types [4,5]. The late region (L) contains two coding genes, L1 and L2, which encode viral capsid proteins.

Among more than 100 types of HPV known today, approximately 30 infect the genital tract. Within these, HPV16 and HPV18 are the two types with the highest oncogenic power. A prospective way of fighting cervical cancer is with an anti-HPV vaccine. Phase III vaccine trials are being developed by Merck, GlaxoSmithKline and the National Cancer Institute [5,6]. Besides, research is ongoing on different aspects of HPV biology, such as the mechanisms of down-regulation through which HPV causes cell transformation, the evaluation of biomarkers for risk progression, the role of environmental co-factors and the determinants of immune response to the viral infection [1]. However, to have a better understanding of the HPV, it is key to gain an improved insight from an evolutionary perspective [7,8]. The possibility for HPV recombination was first suggested by several facts: The biological viability of artificial HPV strains with chimerical proteins [9]. The appearance of some HPV16 variants that seemed mosaics between different established types [10]. The isolation of a novel HPV type (HPV77) with an unusual pattern of sequence similarity over the E6, E7 and L1 regions [11,12]. The plurality of HPV types [13] and also by the frequent observed co-infection (M. Angulo, personal observation). The latter will be specially important in AIDS patients which are often co-infected with very diverse mixtures of human papillomavirus (HPV) [14-17].

Because of HPV extreme diversity, the occurrence of recombination was initially discarded, in part because of the technical difficulties for aligning extremely diverse sequences, and in part because of the less accurate methods available for the researchers until the past decade. Nevertheless, recently reported phylogenetic incongruence at the putative high-risk ancestor node, showing that one or more presumed old recombination events should explain a non monophyletic evolution of oncogenic HPVs [18,19], has provided new convincing support of recombination in α-PVs. In addition, a recent rigorous analysis, using several recombination estimation methods, has

provided fresh evidence of ancient recombination in papillomavirus, especially for the L2 gene [20].

However, the methods used to assess the presence of recombination signals were either phylogenetic based or substitution based. No model-based method was used. Hence, the difficulty of aligning all currently sequenced PVs imposes an additional challenge. Here we addressed the existence of recombination in HPV using a very efficient composite likelihood method [21,22]. The advantages of this kind of method, which is a model-based method, over the model-free ones, are well-known [23] including the fact that with model-free ones the true level of recombination that has occurred is greatly underestimated.

We have centered on human alpha PV sequences, which alignment is much more reliable. Our goal was to get recombination estimates of different genes (E6, E7, L1 and L2) in different groups. We defined three groups (GI, GII and GIII) attending to their phylogenetic relationships but also to their epidemiology and clinical outcome. The identification of the specific recombinant sequences or the recombination break points is a more complex problem and requires different algorithms and software, being out of the scope of the present paper.

Results
Evolutionary model and rate variation
Table 1 shows the best-fit models of nucleotide substitution selected for each data set. Different models were selected for the different genes and for the different groups. The simplest models were found within the HPV16 group and the most complex ones (GTR) for the gene L2 in GI and GII groups. L2 had also the most complex model within the HPV16 and the GIII groups (Table 1).

In addition, rate variation among sites has been detected in several of the data sets, though only in few ones the shape value of the gamma distribution was below one indicating an important rate heterogeneity [24]. The significant rate variation (below one) was mainly detected for GII and GIII groups (Table 1).

Population recombination
Using a gene conversion model with a Jukes-Cantor nucleotide evolution model [25] and assuming no rate variation we found significant recombination in all genes and in all groups (Figure 1). The highest value of recombination was found associated with E6 and the highest number of groups with recombination was linked to L2. Recombination was detected in the genes L1, L2 and E6 only for the GI group (high risk). For the gene E7 recombination was detected only within HPV16 type within

Table 1: Evolutionary models for the different data sets

Sequences	Nucleotide Model	Base Freqs	Rates	Invariable Sites	Rate Variation
HPV16_L1	HKY+I	A = 0.32 C = 0.19 G = 0.19 T = 0.30	Ti/tv = 3.4	0.93	No
HPV16_L2	TrN+I	A = 0.31 C = 0.22 G = 0.16 T = 0.31	R(a) = 1.0 R(b) = 6.5 R(c) = 1.0 R(d) = 1.0 R(e) = 3.9 R(f) = 1.0	0.89	No
HPV16_E6	HKY+I	A = 0.34 C = 0.16 G = 0.22 T = 0.28	Ti/tv = 1.2	0.90	No
HPV16_E7	HKY+I	A = 0.31 C = 0.21 G = 0.23 T = 0.25	Ti/tv= 7.0	0.95	No
GI_L1	TVM+I+G	A = 0.31 C = 0.18 G = 0.19 T = 0.32	R(a) = 4.5 R(b) = 7.9 R(c) = 3.4 R(d) = 3.1 R(e) = 7.9 R(f) = 1.0	0.29	1.47
GI_L2	GTR+I+G	A = 0.29 C = 0.21 G = 0.18 T = 0.32	R(a) = 2.4 R(b) = 4.6 R(c) = 2.8 R(d) = 1.7 R(e) = 3.3 R(f) = 1.0	0.17	1.67
GI_E6	TVM+I+G	A = 0.37 C = 0.17 G = 0.21 T = 0.25	R(a) = 4.0 R(b) = 7.1 R(c) = 2.8 R(d) = 5.4 R(e) = 7.1 R(f) = 1.0	0.17	2.41
GI_E7	HKY+G	A = 0.33 C = 0.23 G = 0.22 T = 0.22	Ti/tv = 1.2	0.00	0.85
GII_L1	TVM+G	A = 0.29 C = 0.18 G = 0.22 T = 0.30	R(a) = 3.6 R(b) = 5.4 R(c) = 2.3 R(d) = 1.8 R(e) = 5.4 R(f) = 1.0	0.00	0.56
GII_L2	GTR+G	A = 0.25 C = 0.24 G = 0.20 T = 0.30	R(a) = 2.3 R(b) = 4.8 R(c) = 2.6 R(d) = 1.4 R(e) = 3.2 R(f) = 1.0	0.00	0.84
GII_E6	TVM+G	A = 0.31 C = 0.20 G = 0.23 T = 0.26	R(a) = 3.1 R(b) = 6.1 R(c) = 2.2 R(d) = 2.8 R(e) = 6.1 R(f) = 1.0	0.00	0.69
GII_E7	TVM+G	A = 0.31	R(a) = 3.7	0.00	1.25

Table 1: Evolutionary models for the different data sets *(Continued)*

		C = 0.22	R(b) = 7.5		
		G = 0.24	R(c) = 3.9		
		T = 0.23	R(d) = 5.0		
			R(e) = 7.5		
			R(f) = 1.0		
GIII_L1	TVM+I+G	A = 0.24	R(a) = 2.7	0.29	0.83
		C = 0.23	R(b) = 4.8		
		G = 0.22	R(c) = 1.1		
		T = 0.30	R(d) = 0.9		
			R(e) = 4.8		
			R(f) = 1.0		
GIII_L2	TVM+I+G	A = 0.22	R(a) = 1.5	0.33	1.85
		C = 0.26	R(b) = 3.1		
		G = 0.22	R(c) = 1.7		
		T = 0.30	R(d) = 0.7		
			R(e) = 3.1		
			R(f) = 1.0		
GIII_E6	TrN+G	A = 0.32	R(a) = 1.0	0.00	0.68
		C = 0.22	R(b) = 2.2		
		G = 0.26	R(c) = 1.0		
		T = 0.20	R(d) = 1.0		
			R(e) = 4.0		
			R(f) = 1.0		
GIII_E7	HKY+I	A = 0.27	Ti/tv = 1.1	0.32	Equal
		C = 0.20			
		G = 0.30			
		T = 0.23			

I: Invariable sites. G: Rate variation. R(a):A-C. R(b): A-G. R(c): A-T. R(d): C-G. R(e): C-T. R(f): G-T. Ti/tv: Transition/transversion rate.

which, however, no other recombinant gene were detected.

When using the best-fit models of nucleotide substitution and considering the estimated rate variation the obtained results were qualitatively similar but with a bit higher estimates and a new signal for L1 in the GII group (cf. Figures 1 and 2).

Given the estimated recombination values and the number of sequences at each group, the expected number of recombination events associated with each data set can be computed [26]. For example, for HPV-16 with a recombination value of 13 for the gene E7 (Figure 2) and $n = 8$ sequences, the expected total number of recombination events in the history of this sample is 34. The expected numbers of recombination events for the different data sets with detected significant recombination are shown in Table 2.

Simulation experiment
A set of simulations was performed using the parameter values corresponding to the evolutionary model, base composition and rate variation among sites for the E6 gene in the group GI (Table 1) to obtain some control samples of DNA sequences. This gene and group were selected because this combination had the highest estimated significant value (Figures 1 and 2 and Table 2). We set the recombination value to zero so that the simulated set of sequences was obtained without recombination. We then estimated from these sequences the population recombination value using the composite likelihood. The average recombination value for the simulated set of sequences was 0.08 ± 0.03 and the percentage of false positive recombination tests was 1%. The expected number of recombination events for this average number is 0.26.

Discussion
The estimation of the best-fit model of nucleotide substitution is relevant in phylogenetics [27]. However, model-based approaches for estimating recombination do not rely on a specific phylogeny and in consequence, they are expected to be robust to model misspecification. This seems to be the case when estimating recombination using the composite likelihood [21,22]. Nevertheless, we have provided the best-fit models of nucleotide substitution under the Akaike information criteria (AIC) and have used them to estimate the population recombination rate. As expected, the model complexity had not a major effect on the estimation.
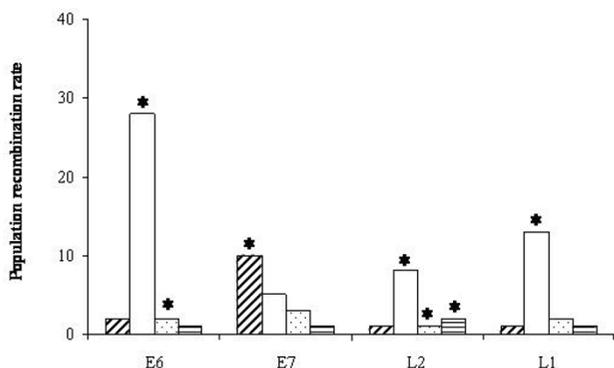
Figure 1
HPV population recombination estimates under a gene conversion model assuming a Jukes-Cantor evolution model with 2 alleles. For each gene and from left to right: Bars with upper right lines: HPV-16. Empty bars: GI. Light shaded bar: GII. Bars with horizontal lines: GIII. *: Recombination test was significant.
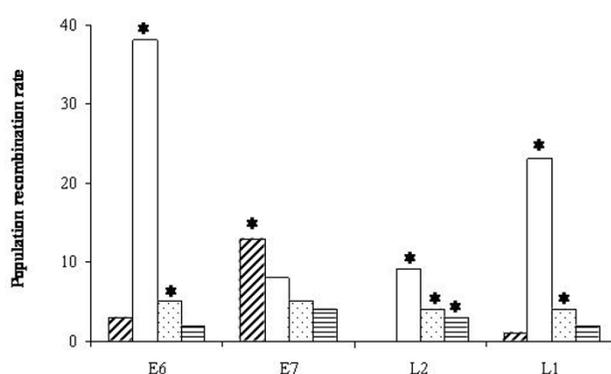

Figure 2
HPV population recombination estimates under a gene conversion model with complex substitution models and rate variation among sites. For each gene and from left to right: Bars with upper right lines: HPV-16. Empty bars: GI. Light shaded bar: GII. Bars with horizontal lines: GIII. *: Recombination test was significant.

The existence of the recombination signal in the DNA sequences of HPVs is important because the genes tested are commonly used to build HPV phylogenies [28] and it is known that recombination can mislead phylogenetic inferences [29]. Furthermore, the detected recombination is in agreement with recently reported phylogenetic incongruence at the putative high-risk ancestor node showing that one or more presumed old recombination events should explain a non monophyletic evolution of oncogenic HPVs [18,19]. The confirmation of ancient papillomavirus recombination has also been recently thoroughly argued by a statistical and phylogenetic recombination detection study [20].

In contrast to that previous study, we have used a model-based method (the coalescent composite likelihood estimator) to infer recombination from HPV DNA sequences.

Model-based methods are known to be preferred over substitution and phylogenetic ones [23,30-32]. Thus, the composite-likelihood estimator maximizes the chances of detecting recombination avoiding, however, the inference of recombination when it is absent (false positive detection).

Regarding the existence of model complexity and rate variation among sites in the HPV samples, it is known that the amount of divergence and rate variation in the data could affect recombination estimation in some cases [32]. However, rate variation i.e. variation in the rate of nucleotide substitution along the sites in the DNA sequence should have no effect on the recombination estimates obtained using the composite-likelihood estimator as has been previously shown [22]. However, the *Pairwise* program assumes a simple Jukes-Cantor model with two alle-

Table 2: Expected number of recombination events

Gene	Group	Number of sequences	Recombination value	Recombination Events
L1	I	14	23	73
L1	II	8	4	10
L2	I	14	9	29
L2	II	8	4	10
L2	III	12	3	9
E6	I	14	38	121
E6	II	8	5	13
E7	HPV16	8	13	34

les [21]. Therefore, to account for possible effects of model complexity and rate variation we also estimated the recombination rate using the extended model possibilities of *Kpairwise* [22] and confirmed that model complexity and rate variation had no qualitative effect onto recombination estimates.

Moreover, we have designed an experiment to check the possibility of false positive detection due to recombination artifacts because of the model complexity and the high diversity underlying the data. As shown above, there was not significant recombination estimation under the parameters considered.

In addition, we were able to estimate the expected number of recombination events in those cases with significant recombination detection. As expected, the higher numbers were found for group GI, which incorporates a major number of branches from PV phylogeny [19]. Importantly, not all recombination events are detectable [32], thus those expectations just provide an upper-bound below which the real number of detectable events should relay [30].

Although we have detected significant recombination signal at all genes and groups in one combination or another, perhaps the most important result is that recombination was detected at intra-type level in HPV16. This may indicate that recombination is occurring at a relative high frequency in current carcinogenic HPV types and variants. HPV recombination should not be exceptional nowadays since the frequency of co-infection with more than one HPV type is not a rare event, and new recombinant types could be currently being generated. Provided that the oncogenicity of specific HPV intra-type variants appear to vary geographically and also with the ethnic origin of the population studied [33], more research is necessary to assess whether such a variation could relate to different recombinant forms. Moreover, the majority of the vaccines under investigation, both the therapeutic and the prophylactic ones, are based on the use of these genes or their products, to obtain the prevention and the treatment of the infections of the more prevalent high risk (types 16 and 18) and low risk types (6 and 11) [6/and references there in, 34/and references there in]. From the present work it is clear that a better knowledge of HPV evolutionary relationships will be important concerning the optimal number of types to include in vaccines as well as the possibility of cross-reactive immunity among HPV types [2,5]. Also, to obtain consensus and ancestor HPV sequences, this could be used in vaccine design to minimize genetic differences between vaccine strains [35]. The existence of recombination should also be of interest to pharmacogenomic studies, i.e. to learn how genetic variation influence response to drugs.

Conclusion

We provide new support to the recent evidence of recombination in HPV. In addition, we perform an evolutionary characterization, estimating best-fit models of nucleotide substitution and rate variation among sites, of some important HPV DNA sequence sets. Using simulations, we have shown that the detected recombination signals should not be artifacts. Thus, we found that the gene with recombination in most of the groups is L2 but the highest values were detected in L1 and E6. Gene E7 was recombinant only within the HPV16 type. The topic deserves further study because recombination is an important evolutionary mechanism that could have high impact in both pharmacogenomics (i.e. the effect of genetic variation on response to drugs) and vaccine development.

Methods
Sequences, groups and models of nucleotide substitution
HPV sequences for the genes E6, E7, L1 and L2 were obtained from the public database at Los Alamos National Laboratory [36] (GenBank accession numbers are given in Table 3). We classified these sequences according to phylogenetic criteria [19] but also by epidemiological criteria and clinical outcome [37]. The groups set as follows: Group I (GI) included the 14 most common high-risk types (16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 73, 82; *n* = 14 sequences including just one variant of type 16). Group II (GII) included 6 low risk types (6, 11, 40, 42, 43, 44; *n* = 8 sequences including the 3 variants of type 6, see Table 3). Group III (GIII) included 3 low risk types plus 5 undetermined risk types which cluster together [19] (61, 72, 81, 62, 71, 83, 84, 89; *n* = 12 or 11 sequences including 5 or 4 variants of type 71, see Table 3. For this group, L1 has 13 sequences because of 4 variants from type 71 plus 2 additional variants from 72 and 81 types). Finally, we consider *per se* the group *HPV-16*, which included HPV16 variants (*n* = 8 sequences, Table 3). All selected sequences pertain to the genus Alpha [38]. Sequences were aligned with ClustalX [39] and then corrected by hand. The best-fit model of nucleotide substitution was selected under the Akaike information criteria (AIC) with Modeltest v3.6 [40], using maximum likelihood (ML) estimates from PAUP* [41].

Recombination estimation
To study HPV recombination we used the composite likelihood estimator [42] and its permutation test [21], which is one of the most powerful techniques to detect the recombination signal from DNA sequences [22]. This method is a model-based population genetic approach, which allows for both a linear recombination and a gene conversion model. The composite likelihood estimator is implemented by the program *Pairwise* from the package Ldhat, freely available at [43], and also by the extension *Kpairwise*[22] which allows for complex nucleotide mod-

Table 3: Accession numbers for HPV sequences including genes E6, E7, L2 and L1

Group I	Accession number	Group II	Accession number
HPV 16	[GenBank:NC_001526, GenBank:AF125673, GenBank:AF402678, GenBank:AF472508, GenBank:AF472509, GenBank:AF534061, GenBank:AF536179, GenBank:AF536180]	HPV 6	[GenBank:NC_000904, GenBank:NC_001355, GenBank:NC_001668]
HPV 18	[GenBank:NC_001357]	HPV 11	[GenBank:NC_001525]
HPV 31	[GenBank:NC_001527]	HPV 40	[GenBank:NC_001589]
HPV 33	[GenBank:NC_001528]	HPV 42	[GenBank:NC_001534]
HPV 35	[GenBank:NC_001529]	HPV 43	[GenBank:NC_005349]
HPV 39	[GenBank:NC_001535]	HPV 44	[GenBank:NC_001689]
HPV 45	[GenBank:NC_001590]	**Group III**	
HPV 51	[GenBank:NC_001533]	HPV 61	[GenBank:NC_001694]
HPV 52	[GenBank:NC_001592]	HPV 72	[GenBank:NC_006164]
HPV 56	[GenBank:NC_001594]	HPV 81	[GenBank:NC_005351]
HPV 58	[GenBank:NC_001443]	HPV 62	[GenBank:NC_006357]
HPV 59	[GenBank:NC_001635]	HPV 71	[GenBank:AY330620, GenBank:AY330621, GenBank:AY330622, GenBank:AY330623, GenBank:NC_002644]
HPV 73	[GenBank:NC_006165]	HPV 83	[GenBank:NC_000856]
HPV 82	[GenBank:NC_002172]	HPV 84	[GenBank:NC_002676]
		HPV 89	[GenBank:NC_004103]

els and rate variation among sites and is freely available at [44]. We considered using a gene conversion model as more adequate since the PV genome is circular. Recombination was estimated as the population recombination rate, which is *4Nr* where *N* is the effective population size and *r* the recombination rate per gene.

Given the recombination values estimated and the number of sequences at each group the expected number of recombination events *E*(*R*) associated to each data set can be computed by using formulae 5 in Hudson and Kaplan [26],

$$E(R) = (4Nr)\sum_{i=1}^{n-1}\frac{1}{i}$$, where *n* is the number of sequences in

the sample.

Simulations
To check that HPV sequences do not generate recombination false positives under the composite likelihood estimator due to their particular combination of model complexity and rate variation, we simulated 100 DNA samples of 15 sequences each and longitude 500 bp using a coalescent model [45]. We used the software *recoal1.7* from David Posada and available upon request from him [46]. Specifically, we set the simulations with the parame-

ter values corresponding to the evolutionary model, base composition and rate variation for the gene and group which had the highest estimated significant value. We also set the recombination value to zero. We then used the obtained sequences to estimate the average recombination value and the percentage of false positive recombination tests.

Competing interests
The author(s) declare that they have no competing interests.

Authors' contributions
MA carried out the acquisition of data, sequence alignment, participated in recombination analyses and helped to draft the manuscript. AC-R conceived and designed the study, participated in the recombination analyses, performed data analysis and simulations and drafted the manuscript. All authors read and approved the final manuscript.

References

1. Monsonego J, Bosch FX, Coursaget P, Cox JT, Franco E, Frazer I, Sankaranarayanan R, Schiller J, Singer A, Wright TC Jr., Kinney W, Meijer CJ, Linder J, McGoogan E, Meijer C: **Cervical cancer control, priorities and new directions.** *Int J Cancer* 2004, **108(3):**329-333.
2. Munoz N, Bosch FX, Castellsague X, Diaz M, de Sanjose S, Hammouda D, Shah KV, Meijer CJ: **Against which human papillomavirus types shall we vaccinate and screen? The international perspective.** *Int J Cancer* 2004, **111(2):**278-285.
3. Parkin DM, Bray F, Ferlay J, Pisani P: **Global cancer statistics, 2002.** *CA Cancer J Clin* 2005, **55(2):**74-108.
4. zur Hausen H: **Papillomaviruses and cancer: from basic studies to clinical application.** *Nat Rev Cancer* 2002, **2(5):**342-350.
5. Arvin AM, Greenberg HB: **New viral vaccines.** *Virology* 2006, **344(1):**240-249.
6. Harper DM, Franco EL, Wheeler CM, Moscicki AB, Romanowski B, Roteli-Martins CM, Jenkins D, Schuind A, Costa Clemens SA, Dubin G: **Sustained efficacy up to 4.5 years of a bivalent L1 virus-like particle vaccine against human papillomavirus types 16 and 18: follow-up from a randomised control trial.** *Lancet* 2006, **367(9518):**1247-1255.
7. Halpern AL: **Comparison of papillomavirus and immunodeficiency virus evolutionary patterns in the context of a papillomavirus vaccine.** *J Clin Virol* 2000, **19(1-2):**43-56.
8. Garcia-Vallve S, Alonso A, Bravo IG: **Papillomaviruses: different genes have different histories.** *Trends Microbiol* 2005, **13(11):**514-521.
9. Heck DV, Yee CL, Howley PM, Munger K: **Efficiency of binding the retinoblastoma protein correlates with the transforming capacity of the E7 oncoproteins of the human papillomaviruses.** *Proc Natl Acad Sci U S A* 1992, **89(10):**4442-4446.
10. Pushko P, Sasagawa T, Cuzick J, Crawford L: **Sequence variation in the capsid protein genes of human papillomavirus type 16.** *J Gen Virol* 1994, **75 (Pt 4):**911-916.
11. Shamanin V, Glover M, Rausch C, Proby C, Leigh IM, zur Hausen H, de Villiers EM: **Specific types of human papillomavirus found in benign proliferations and carcinomas of the skin in immunosuppressed patients.** *Cancer Res* 1994, **54(17):**4610-4613.
12. Smits HL, Traanberg KF, Krul MR, Prussia PR, Kuiken CL, Jebbink MF, Kleyne JA, van den Berg RH, Capone B, de Bruyn A, *et al.*: **Identification of a unique group of human papillomavirus type 16 sequence variants among clinical isolates from Barbados.** *J Gen Virol* 1994, **75 (Pt 9):**2457-2462.
13. Wieland U, Gross GE, Hofmann A, Sohendra N, Berlien HP, Pfister H: **Novel human papillomavirus (HPV) DNA sequences from recurrent cutaneous and mucosal lesions of a stoma-carrier.** *J Invest Dermatol* 1998, **111(1):**164-168.
14. Hameed M, Fernandes H, Skurnick J, Moore D, Kloser P, Heller D: **Human papillomavirus typing in HIV-positive women.** *Infect Dis Obstet Gynecol* 2001, **9(2):**89-93.
15. Haas S, Park TW, Voigt E, Buttner R, Merkelbach-Bruse S: **Detection of HPV 52, 58 and 87 in cervicovaginal intraepithelial lesions of HIV infected women.** *Int J Mol Med* 2005, **16(5):**815-819.
16. Chaturvedi AK, Goedert JJ: **Human papillomavirus genotypes among women with HIV: implications for research and prevention.** *AIDS* 2006, **20(18):**2381-2383.
17. Chaturvedi AK, Myers L, Hammons AF, Clark RA, Dunlap K, Kissinger PJ, Hagensee ME: **Prevalence and clustering patterns of human papillomavirus genotypes in multiple infections.** *Cancer Epidemiol Biomarkers Prev* 2005, **14(10):**2439-2445.
18. Bravo IG, Alonso A: **Mucosal human papillomaviruses encode four different E5 proteins whose chemistry and phylogeny correlate with malignant or benign growth.** *J Virol* 2004, **78(24):**13613-13626.
19. Narechania A, Chen Z, DeSalle R, Burk RD: **Phylogenetic incongruence among oncogenic genital alpha human papillomaviruses.** *J Virol* 2005, **79(24):**15503-15510.
20. Varsani A, van der Walt E, Heath L, Rybicki EP, Williamson AL, Martin DP: **Evidence of ancient papillomavirus recombination.** *J Gen Virol* 2006, **87:**2527-2531.
21. McVean GAT, Awadalla P, Fearnhead P: **A coalescent based-method for detecting and estimating recombination from gene sequences.** *Genetics* 2002, **160:**1231-1241.
22. Carvajal-Rodriguez A, Crandall KA, Posada D: **Recombination Estimation under Complex Evolutionary Models with the Coalescent Composite Likelihood Method.** *Mol Biol Evol* 2006, **23(4):**817-827.
23. Stumpf MPH, McVean GAT: **Estimating recombination rates from population-genetic data.** *Nature Reviews Genetics* 2003, **4:**959-968.
24. Yang Z: **Maximum likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites.** *Mol Biol Evol* 1993, **10(6):**1396-1401.
25. Jukes TH, Cantor CR: **Evolution of protein molecules.** In *Mammalian Protein Metabolism* Edited by: Munro HM. New York, NY , Academic Press; 1969:21-132.
26. Hudson RR, Kaplan NL: **Statistical properties of the number of recombination events in the history of a sample of DNA sequences.** *Genetics* 1985, **111:**147-164.
27. Posada D, Buckley TR: **Model selection and model averaging in phylogenetics: advantages of akaike information criterion and bayesian approaches over likelihood ratio tests.** *Syst Biol* 2004, **53(5):**793-808.
28. Myers G BHU Delius H, Favre M, Icenogle J, van Ranst M, and Wheeler C (eds): **Human papillomaviruses 1994. A compilation and analysis of nucleic and amino acid sequences.** Edited by: Myers G BHUDHFMIJRMWC. Los Alamos, NM , Los Alamos National Laboratory; 1994.
29. Posada D, Crandall KA: **The effect of recombination on the accuracy of phylogeny estimation.** *J Mol Evol* 2002, **54(3):**396-402.
30. Wiuf C, Christensen T, Hein J: **A simulation study of the reliability of recombination detection methods.** *Mol Biol Evol* 2001, **18:**1929-1939.
31. Brown CJ, Garner EC, Dunker KA, Joyce P: **The power to detect recombination using the coalescent.** *Mol Biol Evol* 2001, **18(7):**1421-1424.
32. Posada D, Crandall KA: **Evaluation of methods for detecting recombination from DNA sequences: computer simulations.** *Proc Natl Acad Sci U S A* 2001, **98(24):**13757-13762.
33. Burd EM: **Human papillomavirus and cervical cancer.** *Clin Microbiol Rev* 2003, **16(1):**1-17.
34. Leighton D: **Human papilloma virus vaccines. A review of advances in the development of HPV vaccines.** In *HIV Treatment Bulletin Volume 6*. Issue 7 I-Base; 2005.
35. Gaschen B, Taylor J, Yusim K, Foley B, Gao F, Lang D, Novitsky V, Haynes B, Hahn BH, Bhattacharya T, Korber B: **Diversity considerations in HIV-1 vaccine selection.** *Science* 2002, **296(5577):**2354-2360.
36. Alamos L: **HPV Sequence Database.** [http://hpv-web.lanl.gov/].
37. Munoz N, Bosch FX, de Sanjose S, Herrero R, Castellsague X, Shah KV, Snijders PJ, Meijer CJ: **Epidemiologic classification of human papillomavirus types associated with cervical cancer.** *N Engl J Med* 2003, **348(6):**518-527.
38. de Villiers EM, Fauquet C, Broker TR, Bernard HU, zur Hausen H: **Classification of papillomaviruses.** *Virology* 2004, **324(1):**17-27.
39. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG: **The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools.** *Nucleic Acids Res* 1997, **24:**4876-4882.
40. Posada D, Crandall KA: **Modeltest: testing the model of DNA substitution.** *Bioinformatics* 1998, **14(9):**817-818.
41. Swofford DL: **PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods).** 4th edition. Sunderland, Massachusetts , Sinauer Associates; 2002.
42. Hudson RR: **Two-locus sampling distributions and their application.** *Genetics* 2001, **159(4):**1805-1817.
43. McVean G: **LDhat2.0. A package for the analysis of recombination rates from population genetic data.** [http://www.stats.ox.ac.uk/~mcvean/LDhat/].
44. Carvajal-Rodríguez A: **Kpairwise: Estimating recombination rates under complex substitution models.** [http://darwin.uvigo.es/software/kpairwise.html].
45. Hudson RR: **Gene genealogies and the coalescent process.** *Oxford Surveys in Evolutionary Biology* 1990, **7:**1-44.
46. Posada D: **Computational evolutionary Biology lab.** [http://darwin.uvigo.es/].