

# Retrospective Phylodynamic and Phylogeographic Analysis of the Human Papillomavirus 16 E6 Gene in the Mediterranean Region

Oussama Souiai<sup>1</sup> and Ameni Sallemi<sup>1,2</sup>

<sup>1</sup>Laboratory of Bioinformatics, Biomathematics and Biostatistics (BIMS), Institut Pasteur de Tunis, Tunis, Tunisia. <sup>2</sup>Institut Supérieur de Biotechnologie de Sidi Thabet, Ariana, Tunisia.

Bioinformatics and Biology Insights  
Volume 17: 1–5  
© The Author(s) 2023  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/11779322231178598



**ABSTRACT:** Human papillomavirus 16 (HPV16) is considered to be strongly correlated with the development of cervical cancer. Among the 8 HPV16 genes, the E6 constitutes a remarkable marker to follow the evolutionary history and spatial phylodynamics of HPV16 in the Mediterranean basin. Thus, this work aims to decipher the major evolutionary events and crosstalks in the Mediterranean basin with a focus on Tunisian strains regarding the E6 oncogene. In this study, we first extracted the available and annotated Mediterranean strains of HPV16 E6 gene sequences ( $n = 155$ ) from the NCBI nucleotide database. These sequences were aligned, edited, and used for the downstream phylogenetic analyses. Finally, a Bayesian Markov Chain Monte Carlo approach was applied to reconstruct the evolutionary history of HPV16 migration. Our results showed that the HPV circulating in Tunisia derived from a Croatian ancestor around the year 1987. This starting point spreads to most European countries to reach northern Africa through the Moroccan gateway in 2004.

**KEYWORDS:** HPV16, E6, phylogeography, evolution, substitution rate

**RECEIVED:** February 6, 2023. **ACCEPTED:** May 11, 2023.

**TYPE:** Short Report

**FUNDING:** The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was funded by the Laboratory of Bioinformatics, Biomathematics, and Biostatistics (LR16IPT09), Institut Pasteur de Tunis.

**DECLARATION OF CONFLICTING INTERESTS:** The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**CORRESPONDING AUTHOR:** Oussama Souiai, Laboratory of Bioinformatics, Biomathematics and Biostatistics (BIMS), Institut Pasteur de Tunis, 1002 Tunis, Tunisia. Email: souiai@gmail.com

## Introduction

Human papillomavirus (HPV) is the most common sexually transmitted virus. Around 10% of the infected women will undergo precancerous lesions in their cervical tissue (Yim and Park, 2005)<sup>1</sup>. Depending on their oncogenic potential, HPVs can cause either genital warts when associated to low-risk types (6, 11, 40, 42, 43, 44, and 54) or cervical cancer when it comes to high-risk alpha-HPV types (16, 18, 31, 33, 35, 39, 45, 51, 52, and 58)<sup>2</sup>. HPV16 and 18 are the most common high-risk HPV genotypes, accounting for nearly 80% of all cervical cancer cases.<sup>3</sup> This cancer is the fourth most common cancer affecting women worldwide,<sup>4</sup> with an estimation of 604 000 new cases in 2020, mainly affecting women aged from 45 to 59 years.<sup>5</sup> A scopus on the Tunisian prevalence for this cancer positioned it as the fifth cancer affecting women in the country representing 5.74% of total cancer cases in 2020, leading to 185 deaths per year.<sup>6</sup>

This double-stranded DNA (dsDNA) virus is organized into 3 genomic regions: the early (E), the long control region (LCR), and the late (L) region. Its genome has shown a slow evolution rate compared with other virus families. Indeed, HPV coding region substitution rates are estimated to be between  $2 \times 10^{-8}$  and  $5 \times 10^{-9}$  substitutions per site per year.<sup>7</sup> Typically occurring in dsDNA viruses, this low rate results from the absence of DNA polymerase, prompting the virus to hijack the host's cellular repair machinery during replication. The E6 oncogene, encoding 158 amino acids, is involved in cell cycle alteration and cell differentiation inhibition,<sup>8</sup> making it a suitable genetic marker for

tracking phylogenetic and phylogeographic evolution of papillomaviruses. Previous phylogenetic analyses focused on the prevalent HPV serotypes in Tunisia and North Africa characterized solely the lineages.<sup>3</sup> To our best knowledge, the transmission routes of this virus were not yet elucidated. This work represents an unprecedented phylogeny and phylogeography analysis of the virus, aiming to explore the evolutionary processes of the HPV16 E6 gene that shape the genetic diversity of the HPV strains occurring in Tunisia and its surrounding Mediterranean areas.

## Material and Methods

### Dataset

Human papillomavirus 16 (HPV16) E6 gene sequences were extracted from the GenBank database. The sequences used for the phylogenetic analysis belonged to the following countries: Tunisia ( $n = 15$ ), Croatia ( $n = 4$ ), Greece ( $n = 16$ ), Spain ( $n = 15$ ), Italy ( $n = 34$ ), Egypt ( $n = 4$ ), Algeria ( $n = 20$ ), and Morocco ( $n = 71$ ). Sequences from Egypt and Algeria were not retained in the phylodynamic analysis, as there is no indication of isolation or sampling dates.

### Phylogenetic analyses

A total of 179 sequences were aligned using MUSCLE (Edgar, 2004)<sup>9</sup>. Multiple sequence alignment was manually edited using BioEdit software (Hall, 1999)<sup>10</sup>.

The F81 substitution model was inferred as the most appropriate by applying the jModelTest v2.1.10 program (Posada,



2008)<sup>11</sup>. The phylogenetic analysis was performed using PhyML v3.3.3 (Guindon and Gascuel, 2003)<sup>12</sup>, and the resulting tree was visualized with FigTree v1.4.3. To examine the temporal signal in time-stamped sequences, the maximum likelihood (ML) tree was analyzed with TempEST v.1.5.3.<sup>13</sup>

The time of the most recent common ancestor (tMRCA) estimates were calculated as the height 95% highest posterior density (HPD) years before the most recent sampling dates.

### Phylogeographic analyses

The BEAST v1.8.4 package was used to construct Bayesian maximum clade credibility (MCC) phylogenetic trees for the E6 gene. To infer the geographic origins of Tunisian HPV16 viruses, the location trait was analyzed as a discrete trait diffusion model. To estimate the transition rates between different locations, a Bayesian stochastic search variable selection (BSSVS), with a symmetrical discrete trait substitution model with a strict clock assumption was used. The Markov chain Monte Carlo (MCMC) was run under an Extended Bayesian Skyline model for 200 million generations, sampling every 1000 states. An effective sample size (ESS > 200) was observed for the major parameters. The ESS values were evaluated with the TRACER v1.8.4 tool. TreeAnnotator v1.8.4 was used to generate the MCC tree.

Ultimately, the MCC tree was visualized using the FigTree v1.4.3 and Nextstrain programs, thus allowing the computing of the tMRCA and the HPD at 95% for Tunisian HPV16 viruses. A posterior probability threshold of (pp) > .95 was used to assess the tree nodes with SPREAD3.<sup>14</sup>

To study the spatiotemporal diffusion of HPV, a Keyhole Markup Language (KML) file was generated from the MCC tree with SPREAD3<sup>14</sup> and then visualized with Google Earth Pro.

SPREAD3 was also used to calculate the Bayes factors (BFs) for pairwise diffusion rate between locations. BFs greater than 3 were considered as statistically satisfying.

## Results and Discussion

### *Predominance of European lineage among the Mediterranean HPV16 isolates*

The generated phylogenetic tree showed that the isolates were separated into clades corresponding to HPV16 lineages (Figure 1). Most of the isolates formed a cluster with lineage A, except for the Egyptian ones. Among these isolates, Tunisian HPV16 variants belonged to the European German sublineage (A2) while Moroccan variants were represented ubiquitously in all A sublineages (European, German, and Asian). Overall, 12 sequences from Algeria, Egypt, Greece, Italy, and Spain were grouped with the American lineage (D), while 6 Algerian and Spanish sequences were derived from the African lineage 1 (B) sequences. Meanwhile, 3 Greek and Spanish strains were classified with African lineage 2 (C).

These results suggest that the variants circulating in these areas (Algeria, Egypt, Italy, Spain, and Greece) could be recombinant strains. Such results might be explained by the heterogeneous mixture of these populations and by the migration patterns between countries, which led us to perform further phylogenetic analysis.

### *HPV16 evolves more rapidly than other subtypes*

Through the Bayesian analysis, we estimated the mean evolutionary rate to  $5.532 \times 10^{-4}$  substitutions/site/year (95% HPD interval  $2.4626 \times 10^{-4}$ ,  $8.8793 \times 10^{-4}$  substitutions/site/year). This is in line with the findings by Firth et al<sup>15</sup> reporting a value of  $4.94 \times 10^{-4}$  (95% HPD interval  $0.7490 \times 10^{-4}$ ,  $9.812 \times 10^{-4}$  substitutions/site/year) based on the L1 gene. In contrast, other studies have reported that HPV mutation rates vary in the range of  $10^{-8}$  to  $10^{-7}$  substitution/year,<sup>16</sup> thus suggesting that HPV16 evolves more rapidly. Our findings might be explained by the selective force exerted by vaccines on high-risk strains of HPV, thus leading to a competition between strains to be maintained in the niche (Paul A. Orlando et al 2011)<sup>17</sup>.

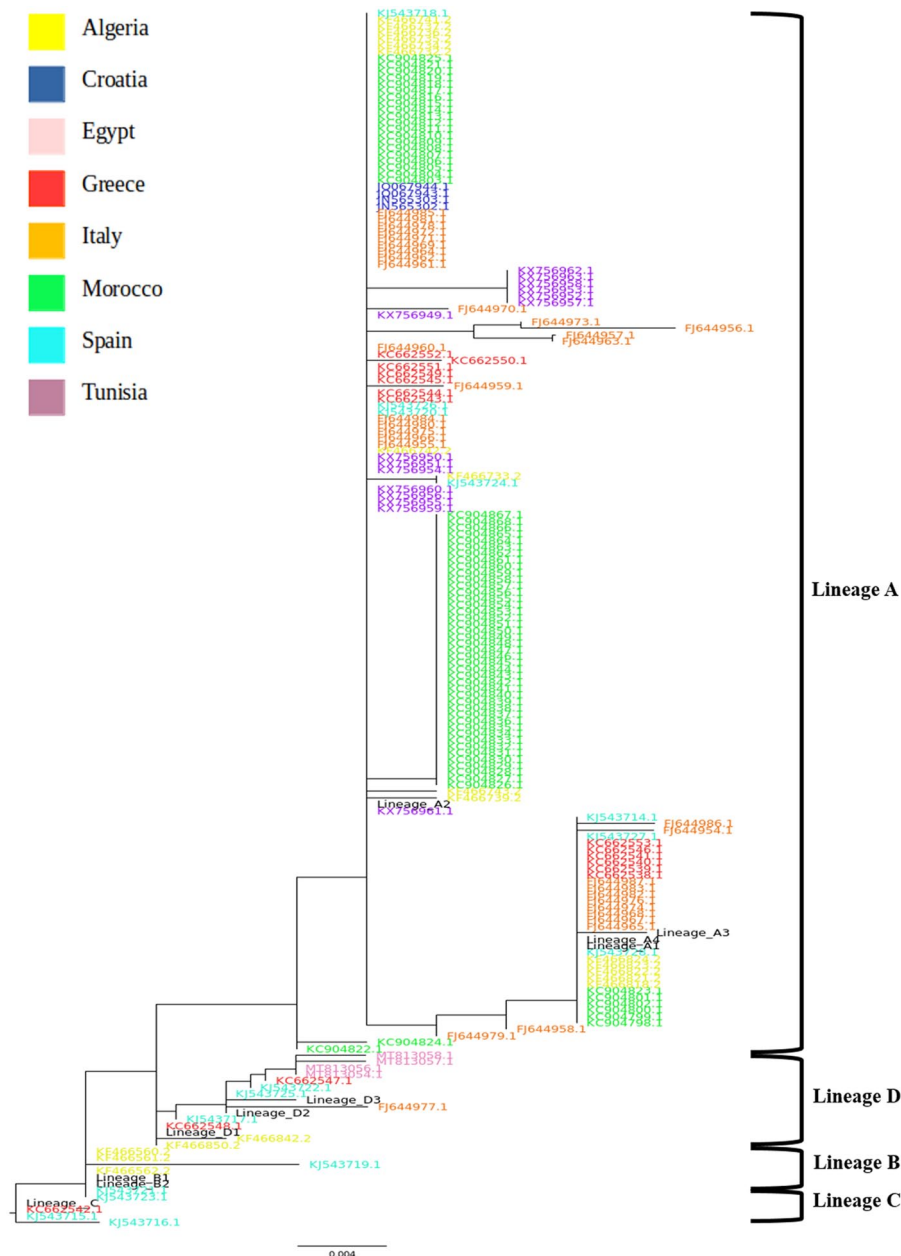
### *2005 as the pivotal year of population increase*

Herein, we estimated that the population size followed 3 distinct phases (Figure 2A). The initial phase showed that the population size remained constant from the date of our sampling to mid-2004. The second phase showed a log-equivalent increase in population size in 2005. The population size then remained stationary prior to a phase of decrease that occurred in 2012.

### *The HPV evolutionary trajectory from Europe to North Africa*

As shown in both Figures 2B and C, HPV16 emerged in Croatia in 1987 (Table 1), then spread to Italy in 1994 prior to an appearance in Spain 5 years later in 1999. These transitions were strongly supported with BF=37.8; pp=.95 and BF=1419.8; pp=.99, respectively. Interestingly, a transition was also observed between Greece and Spain with BF=2371.4 and pp=1. Our analysis also suggests that the observed strains arose in Greece in the early 2000s (BF=2371.3, pp=.99), potentially explaining the previously observed prevalence increase in Greece in 2008<sup>18</sup> with a peak reached in 2011.<sup>19</sup> Interestingly, the period between 2010 and 2011 coincided with an increase in HIV infections, thus leading to higher prevalence of HPV (Lui, 2018)<sup>20</sup>.

In our study, the introduction to North Africa was initiated through Morocco in 2004 from Greece (BF=5.1; pp=.76) and Spain (BF=0.7; pp=.31). Despite a low BF, the transition between Spain and Morocco remains plausible, given the



**Figure 1.** Phylogenetic analysis of the E6 gene with selected HPV16 sequences. The phylogenetic tree was constructed by the ML method using PhyML software.

fluctuations and the very active Spanish–Moroccan cross-border population exchanges.

Since the introduction of the virus in Morocco in 2004 (Figure 2B), the size of the viral population has continued to grow (Figure 2A). Indeed, the prevalence of HPV16 and 18 in Moroccan women with cervical cancer increased from 51.7% in 1999 to 2000 to 89.6% in 2009 to 2012, according to the studies conducted in Rabat and Casablanca.<sup>21</sup> Finally, HPV16 reached Tunisia in 2011 from Morocco. The latter transition is statistically significant (BF = 34.6; pp = .95).

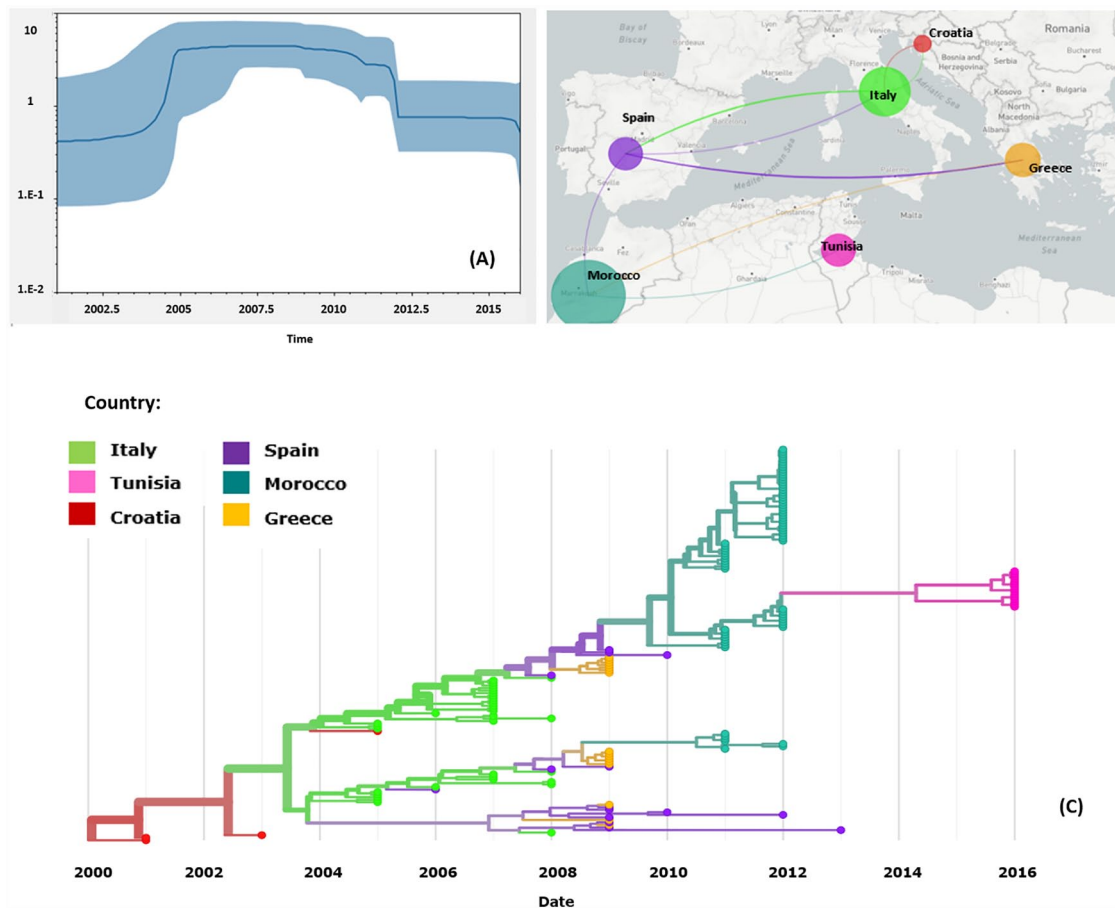
The latter transition should not overshadow other potential transitions. Indeed, an analysis of the phylodynamic patterns of HPV58 showed that the introduction of HPV was initiated

from West Africa.<sup>22</sup> A synchronous transition emanating from this part of the continent for the HPV16 remains possible, especially since the migratory flows toward Tunisia have been very important during the last decade.

### Sampling bias

Sensitivity analysis was performed using a “randomized tip swap” approach to evaluate the influence of sampling bias for the location trait diffusion model.<sup>23</sup>

The main analysis predicted that Croatia was the most likely ancestral root (root state probability). As shown by the comparison of the root state posterior probability (Table 2)



**Figure 2.** Estimated evolution of HPV16 in the Mediterranean countries. (A) Bayesian skyline plot of HPV16. The Bayesian Skyline plot reconstruction of the E6 gene shows the evolution in population size. Median (dark line) and upper and lower 95% HPD (blue region) estimates of effective population size (y-axis) through time in years (x-axis) are shown. (B) Spatiotemporal dynamics of HPV16 circulating in the Mediterranean region between 1987 and 2016. (C) The MCC tree of E6's gene HPV16 sequences summarizing all of the trees obtained during the MCMC search.

**Table 1.** Estimated tMRCA of HPV16 based on E6 gene for all the countries.

COUNTRY	MEAN TMRCA (95% HPD)
Croatia	1987 (1985-1990)
Greece	2000 (1999-2001)
Italy	1994 (1993-1995)
Morocco	2004 (2004-2005)
Spain	1999 (1998-2001)
Tunisia	1999 (1998-2001)

Abbreviations: HPD, highest posterior density; HPV16, human papillomavirus 16; tMRCA, the most recent common ancestor.

between the main analysis and the random “tip swap” analysis,<sup>23</sup> this result was not influenced by sample selection bias (main analysis,  $pp = .93$ ; tip swap analysis,  $pp = .02$ ).

Only 2 identified transitions in spatiotemporal dynamics appeared to be affected by sampling, as they were characterized by a large variation in transition rate between the main and tip

swap analysis. These biased transitions included the virus transition event between Italy and Spain and the one between Spain and Greece.

## Conclusions

Despite a recent renewed interest, there is an obvious gap in the analysis of sequences originating from North Africa.<sup>24</sup> This is particularly true when analyzing the viral phylodynamics of DNA viruses. Indeed, only a marginal number of analyses of this type have been performed on DNA viruses and most focus on RNA viruses given their higher mutation rates. However, phylodynamic analyses applied to slowly evolving pathogens can help elucidate longer-term evolutionary processes, such as speciation or host-pathogen co-divergence, as in the case of HPV.<sup>25</sup> This neglected aspect in the phylogenetic analyses was solely treated by Li et al<sup>22</sup> who characterized the major routes of the HPV58.

This study reported preliminary results of a retrospective spatiotemporal evolution of HPV16 in the Mediterranean basin countries. Our analysis suggests that HPV16 circulating in North Africa had Croatian ancestral strains and reached North Africa by the Moroccan gateway via various transitions. Nevertheless,

**Table 2.** Mean transition rates estimation between main and tip swap analysis.

FROM	TO	MEAN TRANSITION RATES (MAIN ANALYSIS)	BAYES FACTOR	POSTERIOR PROBABILITY	MEAN TRANSITION RATES (TIP SWAP ANALYSIS)
Croatia	Italy	0.692	37.8044729917232	0.958550230276498	0.956
Italy	Spain	1.422	1419.89190628866	0.998850006388853	0.988
Spain	Greece	2.094	2371.39764973359	0.99931111493825	0.956
Greece	Morocco	0.831	5.1911138579994	0.760506886072855	1.058
Spain	Morocco	0.864	0.754432050693237	0.315770467941845	1.012
Morocco	Tunisia	0.604	34.6528372443735	0.954950250276387	1.044

due to the lack of information and annotations on HPV16 strains from countries sharing a historical border, the estimation of some transitions might be biased. Thus, incorporating E6 sequence strains from Algeria, France, Turkey, and Egypt would certainly increase the accuracy of the viral spread pathway results.

Despite these limitations, this is the first study describing the phylogeny and phylogeography of HPV16 in the Mediterranean countries, which have long witnessed active exchange, whether through colonization or immigration.

### Acknowledgements

The authors thank Marwa Arbi, Mariem Hannachi, and Alia Benkahla for their contribution and help.

### Author Contributions

Conceptualization O.S.; data curation, A.S and O.S ; formal analysis and investigation : A.S and O.S ; methodology A.S and O.S ; supervision OS.; writing—review and editing, AS ans OS All authors have read and agreed to the published version of the manuscript.

### Information on Patient Consent

Not applicable.

### REFERENCES

1. Yim, E-K, Park JS. The role of HPV E6 and E7 oncoproteins in HPV-associated cervical carcinogenesis. *Cancer Res Treat.* 2005; 37: 319-324. doi: 10.4143/crt.2005.37.6.319.
2. Mirbahari SG, Sadeghi M. The prevalence of genus alpha human papillomavirus in women with uterine cervical infection and/or inflammation in Western Iran. *Mater Sociomed.* 2018;30:113-117. doi:10.5455/msm.2018.30.113-117
3. Jendoubi-Ferchichi M, Satouri L, Ghoul F, et al. Phylogeny and classification of human papillomavirus (HPV)16 and HPV18 variants based on E6 and L1 genes in Tunisian women with cervical lesions. *Asian Pac J Cancer Prev.* 2018;19:3361-3366. doi:10.31557/APJCP.2018.19.12.3361
4. Vu M, Yu J, Awolude OA, Chuang L. Cervical cancer worldwide. *Curr Probl Cancer.* 2018;42:457-465. doi:10.1016/j.cupr.2018.06.003
5. Gobrane HL-B, Hajjem S, Aounallah-Skhiri H, et al. Summary. *Sante Publique.* 2011;23:31-40.
6. Bruni L, Albero G, Serrano B, et al. *Human Papillomavirus and Related Diseases in Tunisia. Summary Report.* ICO/IARC Information Centre on HPV and Cancer (HPV Information Centre); 2021.
7. Bravo IG, Féllez-Sánchez M. Papillomaviruses: viral evolution, cancer and evolutionary medicine. *Evol Med Public Health.* 2015;2015:32-51. doi:10.1093/emph/cov003
8. De Filippis VR, Ayala FJ, Villarreal LP. Evidence of diversifying selection in human papillomavirus type 16 E6 but not E7 oncogenes. *J Mol Evol.* 2002;55:491-499. doi:10.1007/s00239-002-2344-y
9. Edgar RC. MUSCLE: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 2004; 113. doi: 10.1186/1471-2105-5-113.
10. Hall TA. Bioedit: A user-friendly biological sequence alignment editor and analysis program for windows 95/98/ NT; 1999. [https://doi.org/10.14601/PHYTOPATHOL\\_MEDITERR-14998U1.29](https://doi.org/10.14601/PHYTOPATHOL_MEDITERR-14998U1.29).
11. Posada D. JModelTest: Phylogenetic model averaging. *Mol Biol Evol.* 2008;25:1253-1256. doi: 10.1093/molbev/msn083
12. Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol.* 2003;52(5): 696-704. doi: 10.1080/10635150390235520.
13. Rambaut A, Lam TT, Max Carvalho L, Pybus OG. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol.* 2016;2:vew007. doi:10.1093/ve/vew007
14. Bielejec F, Rambaut A, Suchard MA, Philippe Lemey . SPREAD: spatial phylogenetic reconstruction of evolutionary dynamics. *Bioinformatics (Oxford, England).* 2011;27:2910-2912. doi:10.1093/bioinformatics/btr481
15. Firth C, Kitchen A, Shapiro B, Suchard MA, Holmes EC, Rambaut A. Using time-structured data to estimate evolutionary rates of double-stranded DNA viruses. *Mol Biol Evol.* 2010;27:2038-2051. doi:10.1093/molbev/msq088
16. Harari A, Chen Z, Burk RD. Human papillomavirus genomics: past, present and future. *Curr Probl Dermatol.* 2014;45:1-18. doi:10.1159/000355952
17. Orlando PA, Gatenby RA, Giuliano AR, et al. Evolutionary ecology of human papillomavirus: Trade-offs, coexistence, and origins of high-risk and low-risk types. *J Infect Dis.* 2012;205:272-279. doi: 10.1093/infdis/jir717
18. Stamatakis P, Papazafirovoulou A, Elefsiniotis I, et al. Prevalence of HPV infection among Greek women attending a gynecological outpatient clinic. *BMC Infect Dis.* 2010;10:27. doi:10.1186/1471-2334-10-27
19. Argyri E, Tsimplaki E, Papatheodorou D, Daskalopoulou D, Panotopoulou E. Recent trends in HPV infection and type distribution in Greece. *Anticancer Res.* 2018;38:3079-3084. doi:10.21873/anticancer.12565
20. Liu G, Sharma M, Tan N, et al. HIV-positive women have higher risk of hpv infection, precancerous lesions, and cervical cancer: A systematic review and meta-analysis. *AIDS* 2018;32(6):795-808. doi: 10.1097/QAD.00000000000001765.
21. Belglaiia E, Mougín C. Cervical cancer: current situation and management in Morocco. *Bull Cancer.* 2019;106:1008-1022. doi:10.1016/j.bulcan.2019.08.020
22. Li Y, Li Z, He Y, et al. Phylogeographic analysis of human papillomavirus 58. *Sci China C Life Sci.* 2009;52:1164-1172. doi:10.1007/s11427-009-0149-6
23. Hicks JT, Dimitrov KM, Afonso CL, et al. Global phylodynamic analysis of avian paramyxovirus-1 provides evidence of inter-host transmission and inter-continental spatial diffusion. *BMC Evol Biol.* 2019;19:108. doi:10.1186/s12862-019-1431-2
24. Hamdi Y, Zass L, Orthman H, et al. Human OMICs and computational biology research in Africa: current challenges and prospects. *OMICS.* 2021;25:213-233. doi:10.1089/omi.2021.0004
25. Pimenoff VN, de Oliveira CM, Bravo IG. Transmission between archaic and modern human ancestors during the evolution of the oncogenic human papillomavirus 16. *Mol Biol Evol.* 2017;34:4-19. doi:10.1093/molbev/msw214