

RESEARCH ARTICLE

An unusually high substitution rate in transplant-associated BK polyomavirus *in vivo* is further concentrated in HLA-C-bound viral peptides

Pilar Domingo-Calap^{1,2*}, Benjamin Schubert^{3,4}, Mélanie Joly^{1,2,5}, Morgane Solis^{1,2,6}, Meiggie Untrau^{1,2}, Raphael Carapito^{1,2,7}, Philippe Georgel^{1,2}, Sophie Caillard^{1,2,5}, Samira Fafi-Kremer^{1,2,6}, Nicodème Paul^{1,2}, Oliver Kohlbacher^{3,4,8,9,10,11}, Fernando González-Candelas¹², Seiamak Bahram^{1,2,7*}



OPEN ACCESS

Citation: Domingo-Calap P, Schubert B, Joly M, Solis M, Untrau M, Carapito R, et al. (2018) An unusually high substitution rate in transplant-associated BK polyomavirus *in vivo* is further concentrated in HLA-C-bound viral peptides. *PLoS Pathog* 14(10): e1007368. <https://doi.org/10.1371/journal.ppat.1007368>

Editor: Marco Vignuzzi, Institut Pasteur, FRANCE

Received: September 19, 2018

Accepted: September 28, 2018

Published: October 18, 2018

Copyright: © 2018 Domingo-Calap et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript and its Supporting Information files.

Funding: This work has been published under the framework of the Laboratoire d'Excellence (LABEX) TRANSPLANTEX [ANR-11-LABX-0070_TRANSPLANTEX] and benefits from a funding from the French government, managed by the French National Research Agency (ANR) as part of the « Investments for the future » program (SB)(<http://www.agence-nationale-recherche.fr/>)

1 Plateforme GENOMAX, Laboratoire d'ImmunoRhumatologie Moléculaire, INSERM UMR_S1109, LabEx Transplantex, Centre de Recherche d'Immunologie et d'Hématologie, Faculté de Médecine, Fédération de Médecine Translationnelle de Strasbourg (FMTS), Université de Strasbourg, Strasbourg, France, **2** Fédération Hospitalo-Universitaire, OMICARE, Centre de Recherche d'Immunologie et d'Hématologie, Strasbourg, France, **3** Center for Bioinformatics, University of Tübingen, Tübingen, Germany, **4** Applied Bioinformatics, Department of Computer Science, Tübingen, Germany, **5** Service de Néphrologie et Transplantation Rénale, Hôpitaux Universitaires de Strasbourg, France, **6** Laboratoire de Virologie, Plateau Technique de Microbiologie, Pôle de Biologie, Hôpitaux Universitaires de Strasbourg, France, **7** Laboratoire Central d'Immunologie, Plateau Technique de Biologie, Nouvel Hôpital Civil, France, **8** Quantitative Biology Center, Tübingen, Germany, **9** Faculty of Medicine, University of Tübingen, Tübingen, Germany, **10** Biomolecular Interactions, Max Planck Institute for Developmental Biology, Tübingen, Germany, **11** Institute for Translational Bioinformatics, University Hospital Tübingen, Tübingen, Germany, **12** Unidad Mixta Infección y Salud Pública FISABIO/Universitat de València, Institute for Integrative Systems Biology I2SysBio (CSIC-UV) and CIBER en Epidemiología y Salud Pública, Valencia, Spain

* Current address: Harvard Medical School, Department of Cell Biology, Boston, MA, United States of America

* domingocalap@gmail.com (PDC); siamak@unistra.fr (SB)

Abstract

Infection with human BK polyomavirus, a small double-stranded DNA virus, potentially results in severe complications in immunocompromised patients. Here, we describe the *in vivo* variability and evolution of the BK polyomavirus by deep sequencing. Our data reveal the highest genomic evolutionary rate described in double-stranded DNA viruses, i.e., 10^{-3} – 10^{-5} substitutions per nucleotide site per year. High mutation rates in viruses allow their escape from immune surveillance and adaptation to new hosts. By combining mutational landscapes across viral genomes with *in silico* prediction of viral peptides, we demonstrate the presence of significantly more coding substitutions within predicted cognate HLA-C-bound viral peptides than outside. This finding suggests a role for HLA-C in anti-viral immunity, perhaps through the action of killer cell immunoglobulin-like receptors. The present study provides a comprehensive view of viral evolution and immune escape in a DNA virus.

[investissementsdavenir/](#)). Additional support was received from the Strasbourg High Throughput Next Generation Sequencing facility (GENOMAX) (SB)(no URL available), the Institut National de la Santé et de la Recherche Médicale (INSERM)(SB) (www.inserm.fr), Initiative d'Excellence (IDEX) fund of the University of Strasbourg (UNISTRA)(SB) (www.unistra.fr), the Institut Universitaire de France (IUF)(SB)(<http://www.iufrance.fr>), projects BFU2014-58656R and BFU2017-89594R from Ministry of Economy and Competitiveness (MINECO; Spanish Government) (<http://www.idi.mineco.gob.es>)(FGC) and the project PROMETEO/2016/122 from the Generalitat Valenciana (FGC) (www.gva.es/). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Author summary

Little is known about the mechanisms of evolution and viral immune escape in double-stranded DNA (dsDNA) viruses. Here, we study the evolution of BK polyomavirus and observe the highest genomic evolutionary rate described so far for a dsDNA virus, in the range of RNA viruses, which usually evolve rapidly. Furthermore, the prediction of viral peptides to determine immune escape suggests a specific role of HLA-C in antiviral immunity. These findings are helpful for future advances in antiviral therapies and provide a step forward in our understanding of *in vivo* viral evolution in humans.

Introduction

Viral evolutionary rates can vary strongly depending on the method used to estimate them [1, 2]. Among Baltimore groups, the fastest evolving entities are single-stranded (ss) RNA and reverse-transcribing (RT) viruses, with rates ranging between 10^{-2} and 10^{-5} substitutions per site per year (s/s/y). The rates of double-stranded (ds) RNA and ssDNA viruses range between 10^{-3} and 10^{-6} s/s/y, whereas dsDNA viruses evolve more slowly (10^{-3} and 10^{-8} s/s/y) [3, 4]. It is important to note that only few estimates on dsDNA viruses are published. In fact, higher estimates are based on specific genes, as estimated for human papillomavirus 16 (E6 and E7), human adenovirus (hexon), or JC virus (VP1), which are in the order of 10^{-3} s/s/y [4, 5]. Regarding estimates based on dsDNA complete genomes, all of them range between 10^{-5} and 10^{-7} s/s/y [3, 5]. This finding confirms that viruses are fast evolving entities whereas humans have much lower evolutionary rates (10^{-8} – 10^{-9} s/s/y). However, the well-established co-divergence of viral populations with their hosts suggests the possibility of low evolutionary rates in viruses as well. For example, polyomaviruses were historically considered to be examples of human-virus co-divergence, and have been used as markers for human migration patterns, with proposed estimates ranging from 1.41×10^{-7} to 4×10^{-8} s/s/y [6, 7]. Detailed studies are needed to better understand dsDNA virus evolution *in vivo*, especially in viruses that can be considered as potential pathogens.

In vertebrates, the major driving force in anti-viral immunity is the high level of polymorphism in human leukocyte antigen (HLA) genes. Despite a few recent reports [8, 9], limited information is presently available on the extent of viral variability *in vivo*, especially at the whole viral genome level, and only a few studies have tackled this variability in conjunction with the HLA genotype of infected individuals. Consequently, viral escape mutants—i.e., viruses that produce mutated peptides that are no longer able to bind to cognate HLA molecules—have been mainly studied for limited model epitopes in *in vitro* systems and in highly relevant RNA viruses such as HIV, HCV, influenza or dengue (see the following historical references [10, 11]; for a recent review and full bibliography on the subject see [12]). It is not surprising that RNA viruses can adapt to circumvent the immune responses [4], but little is known about viral escape in DNA viruses.

A better understanding of the epitopes involved in viral escape from the immune system could be useful for the development of vaccines and specific treatments. Here, we initiate a dual approach using the BK virus (BKV) as a model. BKV, which was detected for the first time in 1971, is a 5.1 kb dsDNA virus of the *Polyomaviridae* family that harbors six genes (Agnogene, VP1 to VP3, large T antigen “LTA” and small t antigen “stA”) [13]. The primary infection occurs essentially in childhood and the virus infects up to 90% of the human population. The virus remains persistent throughout life, primarily in the urinary tract [14]. High-level replication mainly occurs in immunocompromised hosts and, more specifically in those receiving

modern immunosuppressive regimens, notably post-kidney transplantation. BKV-associated diseases, especially BKV-associated nephropathy, affect 1–10% of transplant recipients [15, 16] and may lead to loss of the allograft and even death [17]. There are no specific prophylactic or curative treatments, and early diagnosis, as well as quick restoration of immunity (through dampening of immunosuppression), remain the most effective strategies to control the disease.

Results and discussion

High level of variability in BKV as detected by NGS

Access to the virus in the bloodstream and/or urine within a transplant setting, where HLA alleles of both donors and recipients are known, provides a unique opportunity to study viral evolution *in vivo* in the context of the individual's (both recipient and donor) HLA class I genotype. A retrospective cohort of 96 patients—225 samples—that underwent solid organ (N = 83) or hematopoietic cell transplantations (N = 13), harboring a minimum of 10^4 viral copies/mL in whole blood or urine, was selected. Quantitative real-time PCR showed that the viral titers in blood ($8.98 \times 10^4 \pm 2.47 \times 10^4$ copies/mL) were significantly lower than those in urine ($2.16 \times 10^9 \pm 3.94 \times 10^8$ copies/mL) (Mann-Whitney U = 315.0, two-tailed, $P < 0.0001$). After complete deep genome sequencing of all 225 samples and alignment to the BKV Dunlop reference strain (GenBank accession number NC001538), an average of 110 ± 3 polymorphisms per sample was observed with an average median coverage of 3043 ± 78 reads/position (S1 Table, GenBank accession numbers KT896230–KT896454; see Methods). In total, 37.88% of all amino acid positions in the Agnoprotein, 12.43% in VP1, 9.97% in VP2, 11.21% in VP3, 8.20% in LTA and 8.72% in stA, were found to be polymorphic (S2 Table). *Agnogene* is the only gene that is not under apparent selective constraints (Nei-Gojobori test, $P = 0.8663$), while the others are under purifying selection (Nei-Gojobori test, $P < 0.0001$, Fig 1, see Methods). Only a few single nucleotide insertions or deletions were detected in the viral genes (S3 Table). Due to methodological limitations (short reads) the non-coding control region was not included in the analyses.

The occurrence of mutations is the main process generating genetic variability, but other processes, such as genetic drift, gene flow, selection and recombination, are responsible for shaping the genetic structure and variation of viral populations. Here, we present evidence that BKV is under strong purifying selection even in the immunocompromised host. Several specific features of the *Polyomaviridae* (e.g., limited size of the genome, small number of genes and overlapping transcription units) likely account for this outcome. In addition, the prevalence of purifying selection in essential genes is anticipated in all viruses as there is a requirement to complete the viral cycle, even in immunocompromised hosts. Most mutations in coding regions must be deleterious, and a high substitution rate implies the accumulation of mutations with deleterious effects [18]. This phenomenon is well known in RNA viruses, which have high mutation rates and short replication times. Similar results have been shown comparing mutational fitness effects and evolution in ssRNA and ssDNA viruses [19, 20]. Our study supports the hypothesis, in concordance with other recent findings [21], that the evolutionary rate gap between small dsDNA and RNA viruses might not be as wide as previously thought. A recent study in lentiviruses has revealed that the combined effects of sequence saturation and purifying selection can explain the time-dependent pattern of rate variation. Purifying selection acts on the genetic diversity over long timeframes by removing a large number of transient deleterious mutations that are still present within short timeframes [4].

Phylogenetic analysis: Incongruent results between serology and genotyping

Phylogenetic analysis with all BKV complete genomes available from GenBank (Fig 2A) suggested the existence of three large groups or genotypes represented by serotypes I, II/III, and

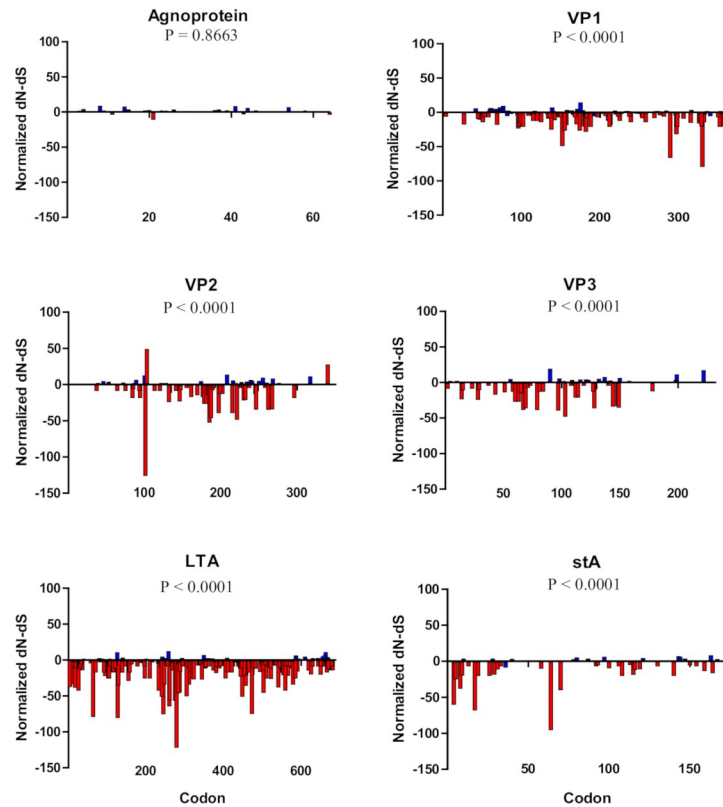


Fig 1. Distribution of the normalized dN-dS per codon among proteins. The six proteins are represented (Agnoprotein, VP1 to VP3, large T antigen “LTA” and small t antigen “stA”). Non-significant values are shown in blue, and significant values in red (positive values for positive selection and negative values for purifying selection, two-tailed binomial distribution). *P*-values correspond to the Nei-Gojobori test of neutrality for each gene.

<https://doi.org/10.1371/journal.ppat.1007368.g001>

IV, with subtypes within genotypes. Limited differences (short branch lengths) between the previously designated genotypes II and III suggested the existence of only one genotype II/III with two subtypes (in contrast to more pronounced differences between serotypes II and III). A similar phylogenetic classification was observed by analyzing only the VP1 gene (Fig 2B). Incidentally, this finding indicated that the current BKV classification should be revised due to inconsistencies between serotyping and genotyping. Next, to establish the genotype of our samples, one reference strain of each genotype and subtype was used for the phylogenetic analysis (Fig 2C). Most of our samples (80.88%) belonged to genotype I, whereas genotypes IV and II/III were less represented (13.78% and 5.3% respectively). The clustering was patient-dependent but independent of the sample origin (urine or blood) and suggested that some samples likely contained a mixture of genotypes. This mixture might be due to multiple lifelong infections or the replication of viruses from the recipient and/or the donor.

High intra- and inter-patient evolutionary rates in BKV

Intra- and inter-patient evolutionary rates were estimated. BKV sequences from samples with possible recombination or a mixture of genotypes according to the RDP output [22] were removed from the analysis (see Methods). We estimated an intra-patient substitution rate for BKV in transplanted patients in the range of 4.90×10^{-4} – 1.22×10^{-3} substitutions per nucleotide site per year (*s/s/y*). No differences between substitution rates in solid organ and hematopoietic cell transplant recipients were found (*t*-test, *P* = 0.2581). To estimate the inter-patient

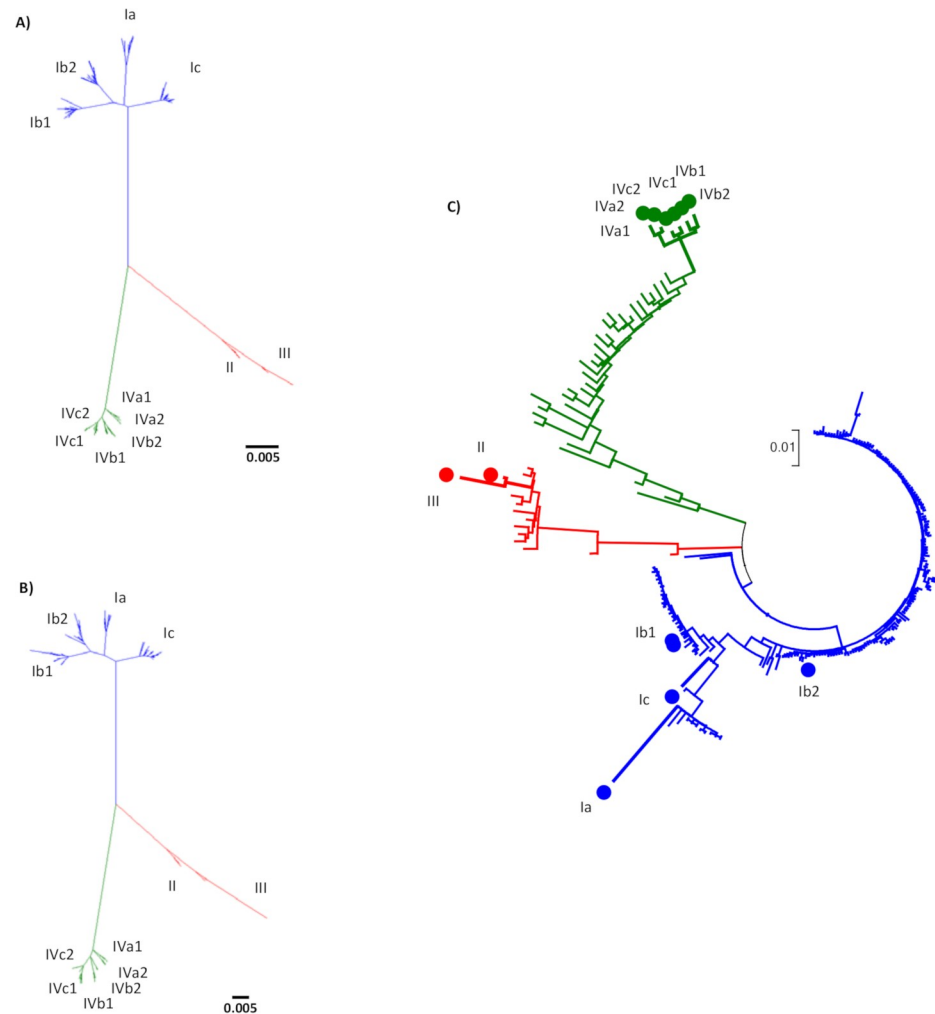


Fig 2. Maximum likelihood phylogenetic trees of BK polyomavirus. Three major groups are found: genotype I in blue, a single group including genotypes II/III in red, and genotype IV in green. (A) Unrooted ML phylogenetic tree with 309 complete genome published sequences retrieved from NCBI. (B) Unrooted ML phylogenetic tree with 309 VP1 gene sequences retrieved from NCBI. (C) Unrooted ML phylogenetic tree with 225 complete genome consensus sequences obtained in this study by next-generation sequencing and one reference strain of each genotype and subtype. Reference strains are marked with dots (Ia, Ib1, Ib2, Ic, II, III, IVa1, IVa2, IVb1, IVb2, IVc1, IVc2).

<https://doi.org/10.1371/journal.ppat.1007368.g002>

evolutionary rate, the best substitution (molecular clock) and demographic model according to marginal likelihood analyses was the relaxed log-normal uncorrelated clock with Bayesian skyline demographic prior. The estimated inter-patient evolutionary rate ranged from 1.00×10^{-5} – 2.15×10^{-4} (95% HDI) for a maximum sampling interval of 568 days. The estimate was quite robust to different demographic and molecular clock models (S4 Table).

The evolutionary rates based on the maximum likelihood and least-squares methods implemented in *treedater* were similar when applied to the whole data set (4.30×10^{-3} s/s/y) but with large parametric bootstrap confidence intervals (in the 10^{-20} to 10^{14} range), thus preventing their consideration as reasonable estimates. However, when the dataset was reduced to the sequences of genotype I ($n = 56$) the average evolutionary rate was estimated at 1.33×10^{-4} (95% CI = 3.13×10^{-6} – 5.59×10^{-3}). These values were close to those obtained with the Bayesian approach described previously.

It is usually assumed that RNA viruses evolve at a rate of 10^{-4} s/s/y, while dsDNA can be close to 10^{-8} s/s/y [23]. ssDNA viruses with small genomes evolving at rates similar to those of RNA viruses have been reported previously [24, 25], as illustrated by the canine parvovirus, with a substitution rate of 1.7×10^{-4} s/s/y [26]. In the case of dsDNA, many evolutionary rates have been calculated under the assumption of co-divergence between viral and human populations, as observed for polyomaviruses. Recently, the substitution rate for JC polyomavirus was evaluated at 1.7×10^{-5} s/s/y [27]. Based on this result, Bayesian analyses suggested the substitution rate of BKV to be on the order of 10^{-5} s/s/y [5, 28], while another study found only minor nucleotide substitutions in the genes encoding late proteins [29].

Here we estimated a substitution rate for BKV on the order of 10^{-3} – 10^{-5} s/s/y (Fig 3). Our experimental results show, for the first time using whole-genome sequencing of *in vivo* viral populations (in a large monocentric cohort), that the genomic evolutionary rate of a dsDNA virus can be as high as that of RNA viruses. It is important to note that the sampling window of sequences may affect the estimates of evolutionary rates, because very short timescales can inflate them. A recent study has shown that estimates of evolutionary rates were lower for broader sampling levels and longer timeframes for both, DNA and RNA viruses, suggesting that the time dependence of substitution rates is ubiquitous among all viruses [4]. For example, lentivirus evolutionary rates from serial samples over a few years within a single patient or host are in the order of 10^{-3} s/s/y [30], reflecting those observed in this study in a small dsDNA virus.

In addition, a previous study comparing the evolution of ssRNA and ssDNA viruses has shown that small genomes (< 5 kb) can evolve rapidly [24] regardless of their encoding material, and that the well-known correlation between genome size and mutation rate [70] can also hold for evolutionary rates. Here, we show that small dsDNA genomes can also evolve as fast as single-stranded ones. Although BKV uses the host DNA polymerase for its replication, the virally-encoded Agnoprotein inhibits dsDNA break repair activity, thereby potentially increasing the error rate during BKV DNA replication [71]. Interestingly, cell tropism of RNA viruses was recently suggested as a key factor in their capacity to evolve, since viruses replicating in epithelial cells (as BKV) are characterized by rapid replication and higher substitution rates [72].

To investigate the relationship between the evolutionary rate of the virus and the immunosuppressive drug regimen—hence the strength of the immune system—we analyzed such information in our kidney transplant recipient cohort (the largest subgroup in our cohort). Kidney transplant patients were given either anti-thymocyte globulin (ATG) (immunological high-risk patients) or anti-Interleukin-2 receptor (anti-IL-2R) (immunological low-risk patients) as induction treatments, and tacrolimus (immunological high-risk patients) or cyclosporine (immunological low-risk patients) as maintenance therapy. Mycophenolate mofetil and steroids were also part of both drug regimens (for high- and low-risk patients). Evolutionary analysis of the different subgroups showed no significant differences in the mutational load (full negative binomial mixed model regression with random effect intercept to account for repeated measures) nor in inter-patient substitution rates where ranges overlapped between treatments (ATG 6.12×10^{-4} – 1.03×10^{-5} s/s/y, Anti-IL-2R 8.60×10^{-4} – 1.36×10^{-5} s/s/y, tacrolimus 4.64×10^{-4} – 9.31×10^{-6} s/s/y, and cyclosporine 1.72×10^{-3} – 1.11×10^{-5} s/s/y).

Immune escape in BKV associated with HLA-C epitopes

To investigate the genetic immune escape mechanism of BKV, potential T-cell epitopes presented by HLA class I were predicted using both donor and recipient HLA alleles, combined with the viral substitutions found herein (S1 Fig, S2, S5 and S6 Tables, see Methods). In this

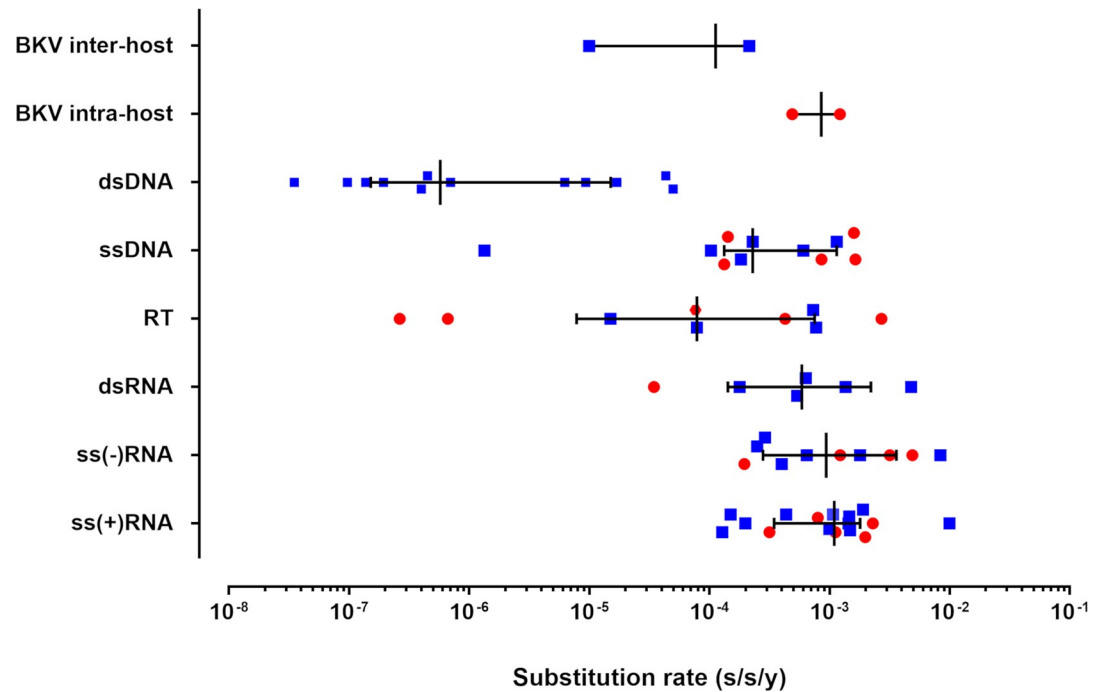


Fig 3. Genomic evolutionary rates for the major Baltimore groups and BKV. Substitution rates are given as substitutions per nucleotide site per year (*s/s/y*). For the major groups (*dsDNA*: double-stranded DNA viruses—BKV [5, 7, 28] (time span of sequences (TSS) of 29 years (y), 25 y, and 32 y, respectively), JC polyomavirus [27, 31] (TSS 33 y and 13 y, respectively), herpes simplex virus 1 [32, 33] (TSS not available and 21 y, respectively), human papillomavirus 18 [34] (TSS not available), monkeypox virus [35] (TSS 7 y), variola virus [5] (TSS 31 y), varicella zoster virus [5] (TSS 37 y); *ssDNA*: single-stranded DNA viruses—African cassava mosaic virus [25] (TSS 5 y), banana bunchy top virus [36] (TSS 2 months), human bocavirus [37] (TSS 1 y), human parvovirus B19 [38, 39] (TSS 14 y and 28 y, respectively), porcine circovirus 2 [40] (TSS 27 y), tomato yellow leaf curl virus [41] (TSS 29 y); *RT*: retroviruses—avian hepatitis B virus [42] (TSS 22 y), human hepatitis B virus [42–44] (TSS 22 y, 25 y and 35 y, respectively); human immunodeficiency virus 1 [45] (TSS 2 y), primate T-cell lymphotropic virus [45] (TSS 2 y); *dsRNA*: double-stranded RNA viruses—bluetongue virus [46] (TSS 48 y), human rotavirus [47] (TSS 16 y), homalodisca vitripennis virus [48] (TSS 2 y); *ss(-)RNA*: single-stranded RNA viruses with negative polarity—Ebola virus [49] (TSS 4 months), fever, thrombocytopenia and leukocytopenia syndrome virus [50] (TSS 4 y), influenza A virus [51, 52] (TSS 28 y and 1 y, respectively), hepatitis delta virus [53] (TSS 3 y), human respiratory syncytial virus [54] (TSS 10 y), rabies virus [55] (TSS 30 y), rift valley fever virus [56] (TSS 10 y); and *ss(+)RNA*: single-stranded RNA viruses with positive polarity—avian coronavirus [57] (TSS 41 y), barley yellow dwarf virus [58] (TSS 2 y), dengue virus [59] (TSS 29 y), foot-and-mouth disease virus [60] (TSS 75 y), hepatitis A virus [61] (TSS 13 y), hepatitis C virus [62] (TSS 20 y), Japanese encephalitis virus [63] (TSS 60 y), Middle East respiratory syndrome coronavirus [64] (TSS 4 months), porcine reproductive and respiratory syndrome virus [65] (TSS 3 y), rubella virus [66] (TSS not available), severe acute respiratory syndrome coronavirus [67] (TSS 4 months), St. Louis encephalitis virus [68] (TSS 46 y), Venezuelan equine encephalitis virus [69] (TSS 54 y)). Each point represents the value of a previously published genomic evolutionary rate (note that for some references, more than one substitution rate is represented in the caption). Red circles represent short time span estimates (< 5 years) and blue squares represent long-time span estimates (> 5 years). Medians with interquartile ranges are indicated. In the case of the inter- and intra-host genomic evolutionary rates of BKV, the values are represented as a range of values obtained in this study.

<https://doi.org/10.1371/journal.ppat.1007368.g003>

way, we determined the putative HLA ligandome of the virus as linked to the individual’s cognate HLA genotype. Interestingly, the two codons in VP2 that appeared to be under positive selection corresponded to codons within predicted epitopes. The VP2 103 codon, the one with the highest level of significant difference, was found in three predicted HLA-C epitopes (KFFDDWDHKVSTV, FFDDWDHKV and FFDDWDHKVSTV), and codon 340 was located within two HLA-A predicted epitopes (TTNKRRSR and TTNKRRSRSSR).

We also found a higher fraction of observed amino acid substitutions within HLA-C epitopes compared with the fraction of amino acid substitutions outside of HLA-C epitopes (one-sided Wilcoxon signed test, $P = 3.71 \times 10^{-10}$). The opposite behavior was observed for HLA-A

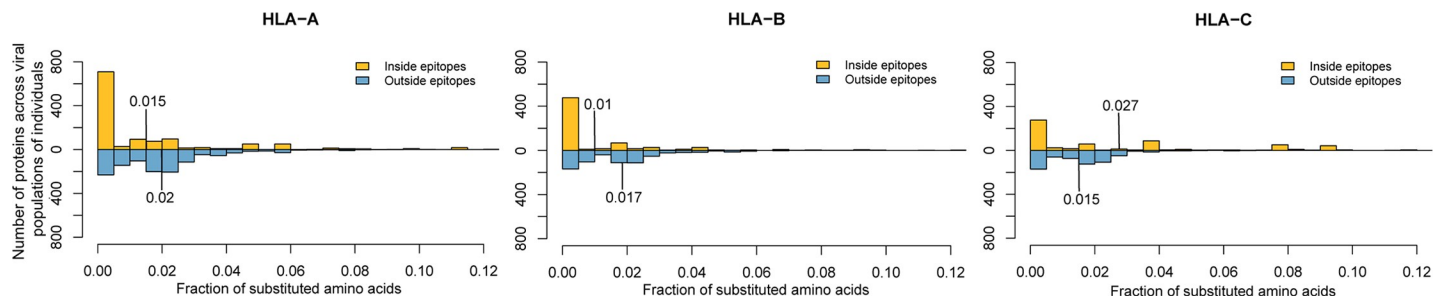


Fig 4. Fraction of amino acid substitutions within and outside of predicted epitopes presented by HLA-A, -B and -C molecules across individuals. The detected amino acid substitutions of a viral population were mapped onto reference proteins and the fraction of mutated amino acids within and outside of predicted epitopes of each viral protein and hosts HLA allele were calculated for each viral population found in patient and donor respectively. The fraction of substituted amino acids within HLA-A and -B presented epitopes (yellow) is significantly lower compared with the fraction outside (blue), while the fraction of amino acid substitutions in HLA-C binding epitopes is significantly higher compared with the fraction outside.

<https://doi.org/10.1371/journal.ppat.1007368.g004>

and -B presented epitopes (one-sided Wilcoxon signed, HLA-A: $P = 4.17 \times 10^{-29}$; HLA-B: $P = 1.35 \times 10^{-26}$) (Fig 4). This difference in contribution of HLA loci was independent of the transplantation type (solid organ or hematopoietic) or the origin of the HLA loci (whether from the donor or the recipient) as assessed by a three-way ANOVA ($P = 0.7947$). Therefore, our results suggest that HLA-C might be specifically involved in the immune response against BKV through its peptide selection capacity for viral peptides. A possible mechanistic explanation for this finding stems from the amply documented interaction of HLA-C with natural killer (NK) and T cells expressing the killer cell immunoglobulin-like receptors (KIR). Notably, the relevance of KIR and HLA-C interactions has been described for viral infections [73, 74], and the involvement of NK cells in the immune response against BKV has also been reported [75, 76], although further investigations should be done to confirm this hypothesis.

High evolutionary rates in RNA viruses allow them to escape immune pressures. Interactions between HLA epitopes and viruses have been described for a variety of RNA viruses, such as HIV, HCV, influenza or dengue, while little is known about immune escape in DNA viruses. A few studies in HPV-16 or herpes simplex virus have been done to improve vaccine design and drug development, but those studies have only examined a fraction of the proteins and not at whole-genome sequencing data [77–79]. This work, to our knowledge, is the first in which predicted epitopes from whole genome sequencing have been studied in an *in vivo* cohort, in conjunction with cognate HLA alleles, to understand the mechanism involved in immune escape in a DNA virus. Our results of viral escape combined with the high evolutionary rate described herein suggest that a combination of drugs should be used as potential treatment against BKV, as commonly used in highly variable viruses such as HIV and HCV, due to the variable viral populations present in a single patient as observed in our study.

Conclusions

The present work describes an unusually fast evolutionary rate for BKV *in vivo* and charts its interaction with the immune system—through the analysis of cognate HLA alleles—whilst considering the whole viral genome and not only candidate epitopes. It further offers a blueprint for similar analyses in other viruses and helps to better rationalize anti-viral therapy and candidate vaccine development. Our results suggest that small dsDNA viruses should be treated as RNA viruses due to their similarities in evolution and immune escape. Thus, a combination of drugs might be necessary for the treatment of BKV, as used for fast evolving RNA viruses. It is important to note that new analytic methods for the study of the evolutionary

rates are needed to better understand the effect of time spans and improve the comparison between estimates.

Materials and methods

Patients and samples

Ninety-six transplanted patients between 2012 and 2013 from the Strasbourg University Hospitals (France) with high levels of post-transplant BKV viremia—as detected by routine BKV testing at the hospital's clinical virology laboratory—were enrolled in this study. Sixty-eight patients underwent kidney transplantation, 12 were lung recipients, 3 received double (kidney-heart; heart-lung or kidney-pancreas) transplants and 13 hematopoietic stem cell transplantation. A total of 225 samples, including 197 urine (from 94 patients) and 28 whole blood (from 13 patients) were included. Urine samples were collected longitudinally for 36 patients.

Ethics statement

All patients were enrolled in the study following the Helsinki guidelines. Written informed consent for genetic testing was obtained from all patients and the study was approved by the Strasbourg University Hospitals institutional review board (RNI DC-2013-1990).

DNA isolation, quantitative BKV real-time PCR, PCR and sanger sequencing

Urine and whole blood samples were collected, and DNA was purified using the QIAextractor instrument (Qiagen, Hilden, Germany), following the DX protocol. Extracted DNA was stored at -80°C until analysis. Blood and urine specimens were assessed using the BK virus R-gene quantification kit (Biomérieux, Lyon, France) following the manufacturer's recommendations. DNA was amplified by Phusion Polymerase (New England Biolabs, MA, USA) using specific overlapping primers. Nested PCR was performed for samples with a low BKV DNA load (usually blood samples). PCR products were purified using the GeneJET DNA purification Kit (ThermoFisher Scientific, Waltham, MA, USA) and quantified with Qubit (ThermoFisher Scientific, Waltham, MA, USA). Twenty-one urine-blood paired samples were used for sequencing by the Sanger method using an ABI Prism 3130 Genetic Analyzer (ThermoFisher Scientific, Waltham, MA, USA). Bi-directional sequencing was performed with the Big Dye Terminator v3.1 kit (ThermoFisher Scientific, Waltham, MA, USA) following the manufacturer's recommendations. Chromatograms were analyzed with the Staden package (24) to obtain the consensus sequence for each sample. These consensus sequences were obtained to compare with the results after the next-generation sequencing assembly to validate our pipeline.

Next-generation sequencing (NGS) and sequence assembly

All 225 urine and blood samples were sequenced by NGS. PCR products from the same samples were pooled in equimolar amounts and library construction with barcodes was performed according to the Fragment Library Preparation protocol using the AB Library Builder System (ThermoFisher Scientific, Waltham, MA, USA). Libraries were quantified by Qubit (ThermoFisher Scientific, Waltham, MA, USA) and then pooled in equimolar amounts for Template beads preparation using the SOLiD EZ beads System (ThermoFisher Scientific, Waltham, MA, USA). Template beads were subjected to sequencing using SOLiD 5500 (ThermoFisher Scientific, Waltham, MA, USA) with the paired-end 75 bp / 35 bp workflow. Sequences were assembled against the Dunlop reference strain (GenBank accession number NC001538) using LifeScope software (ThermoFisher Scientific, Waltham, MA, USA). Comparison with Sanger

sequencing was performed to ascertain the correct assemblies. To quantify the variability per sample, mutations were analyzed with SeqMan software (DNASTAR, Madison, Wisconsin, USA). For each sample, we obtained a list of variants with their genomic location, coverage, and quality metrics, among others. To establish a cutoff for variant calling, we introduced internal controls including (a) a clone from the Dunlop reference strain, pBK (BKV34-2) plasmid (ATCC 45025) prepared by miniprep (ThermoFisher Scientific, Waltham, MA, USA); (b) PCR amplicons from the same clone; and (c) PCR amplicons in duplicate from three of the samples. These controls were processed using the same sequencing methodology to establish the rate of sequencing and PCR errors. The final list of variants was selected by means of a Fisher's exact one-sided test comparing evidence obtained from the data for every potential polymorphism to the estimated error rate using our internal controls. Based on this analysis, BKV sequence variants found in less than 0.5% of reads were removed from the analysis.

Mutation analysis

Sequences were aligned and assembled against the Dunlop strain by Muscle implemented in MEGA version 6 [80] with default parameters in order to compare and determine point mutations, insertions, deletions, and other sequence variations. For better analysis of coding regions, individual datasets per gene were obtained. Further analysis of synonymous and non-synonymous substitutions and the Nei-Gojobori test of neutrality were performed with MEGA version 6 [80].

Phylogenetic analyses

Phylogenetic analyses of the whole genome consensus sequences obtained from all samples, and for each gene separately, were performed using MEGA version 6 [80]. Maximum likelihood phylogenetic trees were constructed with the general time reversible model (GTR) of nucleotide substitution with gamma distribution to account for rate heterogeneity among sites, as this model achieved the lowest AIC score. Similar analyses were performed for 309 BKV complete genome sequences collected from GenBank (all items found by searching the NCBI nucleotide database for "BK polyomavirus complete genome").

To genotype the populations in the different samples, two approaches were performed. First, phylogenetic trees with all our samples and one of the reference strains for each genotype and subtype were obtained following the methodology explained previously. We determined the genotype as the shortest branch distance to one reference. The second approach was based on the methodology proposed by Luo and colleagues, in which point mutations specifically reported in particular genotypes are described [81].

Estimation of substitution rates

To estimate the evolutionary rates of BKV, intra- and inter-patient analyses were performed. Upon multiple alignment, consensus sequences were tested using RDP software [22] for potential recombination, and those with positive results using at least two different methods implemented in the RDP package were removed from the ensuing analyses. Samples showing mixtures of genotypes were also excluded since they could interfere with the calculation of the substitution rate. To estimate the intra-patient substitution rate, we used urine samples from twenty-five patients collected at different times (the first positive samples and after 6 months). To calculate substitutions per site per year, we considered all the different genomic positions between two different times that were fixed in the populations. All the substitutions that reverted to the reference base were not included since the possibility of them already being present in the ancestral population at a low frequency could not be ruled out. Thereby only

substitutions appearing *de novo* and exhibiting a high proportion in the population (fixed substitutions, more than 80% of the reads) were included in this approach. With this methodology, we obtained conservative estimates.

To estimate the inter-patient substitution rate, the consensus sequence for the first available urine sample of each patient with a known date of sampling was selected. After being tested by RDP, a dataset of 79 BKV sequences was used to estimate the inter-patient evolutionary rate (sequences from 15 patients were potential recombinants). A maximum likelihood phylogenetic tree was obtained using Phyml [82] with the GTR model with gamma distribution and invariant sites to account for heterogeneity among sites. This model was determined to be the most appropriate for this dataset with jModeltest [83]. TempEst analysis was conducted to detect a correlation between genetic divergence and sampling time, and it assured a temporal signal in our inter-patient dataset (S2 Fig) [84]. We used Bayesian estimates of the evolutionary rate with dated tips as implemented in BEAST [85]. Based on previous results by Firth et al. [5], we considered three molecular clock models (strict, relaxed log-normal uncorrelated, and relaxed exponential uncorrelated) and two demographic models (constant population size and Bayesian skyline). The GTR model with a gamma distribution and invariant sites was used as the nucleotide substitution model in all combinations. Model selection was performed through computation of the marginal likelihood using path sampling and stepping stone sampling analyses [86]. A lognormalPrior with a mean of 1×10^{-6} and a standard deviation of 1.0 was used for the substitution rate. Two independent runs of 30 million steps with 10% burn-in were used to obtain the median and 95% high probability density intervals for the relevant parameters in each model. In all cases, the effective sample size was > 200 , as checked with Tracer v. 1.5 (available from <http://beast.bio.ed.ac.uk>).

In addition, we used the recently developed method of Volz and Frost which uses maximum likelihood and least squares to estimate evolutionary rates and dates based on relaxed molecular clocks. The method is implemented in the R package *treedater* [87].

Prediction of BKV epitopes

To predict BKV-encoded T-cell epitopes that can be presented by HLA alleles, HLA high-resolution typing (2 fields) was done at the Etablissement Français du Sang Grand Est (Strasbourg) using a sequence-specific oligonucleotide technology. High-resolution typing data of *HLA-A*, *-B* and *-C* of 75 available donor / recipient pairs were used in each analysis, using the recipient's viral populations in each case (S5 Table).

NetMHC 3.4 [88] was used to predict the peptide binding affinities of potential HLA class I epitopes occurring in BKV Dunlop reference proteins to HLA class I alleles of the patients and donors. Peptides eliciting a predicted IC_{50} of less than 50 nM were considered epitopes. IC_{50} values represent the concentration of the peptide that will displace 50% of a standard peptide from the HLA molecule. The lower the IC_{50} value, the stronger is the affinity of the peptide for the tested HLA molecule. According to the NetMHC parameters, peptides with $IC_{50} < 50$ nM were considered high-affinity binders. IC_{50} values of 5 nM and 500 nM were also tested, but a cutoff of 50 nM was chosen as the best indicator (at a 5 nM threshold not enough peptides were predicted to bind; at 500 nM all possible peptides within a given proteins were predicted to bind). Furthermore, all predicted epitopes were tested with NetChop 3.1 [89] to predict whether the epitopes could have been produced by the human proteasome using default parameters. All strong binding peptides with a high likelihood of being correctly cleaved (score prediction higher than the default threshold of 0.5) were included in further analyses.

To calculate the fraction of substituted amino acids within and outside of HLA epitopes, the substitutions detected in the specific viral populations of each patient were mapped onto

viral reference proteins, and the number of substitutions that occurred within and outside of the predicted epitopes were calculated for each protein and HLA allele of each patient and donor respectively. The counts were normalized to the number of potentially mutable amino acids per category (i.e., within or outside of epitopes), to make them comparable across proteins of varying length.

Statistical comparison of the internal and external fractions was performed with a one-sided Wilcoxon signed test for each HLA allele to identify the direction of the difference. The *P*-values were Bonferroni corrected to account for multiple testing.

Supporting information

S1 Fig. BK polyomavirus proteins and location of predicted epitopes. (A) Agnoprotein, (B) VP1, (C) VP2, (D) VP3, (E) large T antigen “LTA” and (F) small t antigen “stA”. Variable amino acids are shown in yellow. Location of predicted epitopes for each protein presented by HLA-A, -B and -C are presented in grey.

(TIF)

S2 Fig. Root-to-tip regression analysis for whole-genome BK polyomavirus sequences. The root-to-tip genetic distance against sampling time is shown for the BK polyomavirus phylogeny with a maximum sampling time of 568 days. The sampling time is given in days ($R^2 = 0.086$, $P < 0.05$).

(TIF)

S1 Table. Patients and samples enrolled in this study. The transplant organ, whether the patient developed the associated nephropathy BKVAN, source (urine/blood), viral load, total number of polymorphisms found in each sample and median coverage are represented.

(PDF)

S2 Table. Single nucleotide polymorphisms in the coding regions found in the 225 samples (from 96 patients). The genomic position as well the reference and substitution nucleotides and amino acids are shown. The percentage of samples in which the position was found is indicated. The genomic position and the reference base and amino acid correspond to the BKV Dunlop reference strain.

(PDF)

S3 Table. Insertions and deletions detected in the viral genomes of the 225 samples (from 96 patients). The positions, locus, reference and polymorphism, and percentage of samples with the polymorphism are shown. The genomic position and the reference base according to the BKV Dunlop reference strain.

(PDF)

S4 Table. Inter-patient substitution rates. Summary of interpatient evolutionary rate estimates (substitutions/site/year, *s/s/y*) of BKV using different molecular clock (strict, relaxed log-normal uncorrelated and relaxed exponential uncorrelated) and demography (constant size and Bayesian skyline) models. Median and 95% high-density interval (HDI) intervals are shown. Estimates were obtained after two independent runs of 30 million generations each with a 10% burn-in. Convergence of the runs ($ESS > 200$) was checked with Tracer.

(PDF)

S5 Table. Allele frequencies of MHC class I in our cohort. Alleles are shown for HLA-A, -B and -C at the 2nd field of resolution for donors and recipients.

(PDF)

S6 Table. BK polyomavirus predicted epitopes presented by HLA-A, -B and -C by protein. Agnoprotein, VP1-3, large T antigen “LTA” and small t antigen “stA” predicted peptides presented by HLA-A, -B and -C from the BK polyomavirus Dunlop reference strain are listed. The starting and ending amino acid of the protein, length of the peptide, peptide sequence, and HLA allele that can present peptide are shown. The IC₅₀ for each peptide and specific HLA allele are also included.
(PDF)

Acknowledgments

We thank Drs M. Javad Aman (Integrated BioTherapeutics, Gaithersburg, MD), Marco Colonna and David Wang (both at Washington University, St. Louis, MO), Rafael Sanjuán (University of Valencia, Spain), and Cathal Seoighe (National University of Ireland, Galway) for critical reading of this manuscript. We thank Dr Darren Martin for his help with the RDP package analysis and Dr Santiago F. Elena for earlier discussions. We thank Marion Le Gentil, Sandra Michel and Dr Clotilde Muller for technical assistance and/or help with patient recruitment.

Author Contributions

Conceptualization: Pilar Domingo-Calap, Seiamak Bahram.

Data curation: Pilar Domingo-Calap, Benjamin Schubert, Raphael Carapito, Samira Fafi-Kremer, Nicodème Paul, Oliver Kohlbacher.

Formal analysis: Pilar Domingo-Calap, Raphael Carapito, Nicodème Paul, Oliver Kohlbacher, Fernando González-Candelas, Seiamak Bahram.

Funding acquisition: Raphael Carapito, Fernando González-Candelas, Seiamak Bahram.

Investigation: Pilar Domingo-Calap, Benjamin Schubert, Mélanie Joly, Meiggie Untrau, Raphael Carapito, Samira Fafi-Kremer, Oliver Kohlbacher, Fernando González-Candelas, Seiamak Bahram.

Methodology: Pilar Domingo-Calap, Benjamin Schubert, Nicodème Paul, Oliver Kohlbacher, Fernando González-Candelas, Seiamak Bahram.

Project administration: Philippe Georgel, Oliver Kohlbacher, Seiamak Bahram.

Resources: Pilar Domingo-Calap, Benjamin Schubert, Mélanie Joly, Morgane Solis, Sophie Caillard, Samira Fafi-Kremer, Oliver Kohlbacher, Seiamak Bahram.

Software: Benjamin Schubert, Nicodème Paul, Oliver Kohlbacher.

Supervision: Pilar Domingo-Calap, Raphael Carapito, Oliver Kohlbacher, Fernando González-Candelas, Seiamak Bahram.

Validation: Pilar Domingo-Calap, Oliver Kohlbacher, Fernando González-Candelas, Seiamak Bahram.

Visualization: Pilar Domingo-Calap.

Writing – original draft: Pilar Domingo-Calap, Philippe Georgel, Fernando González-Candelas, Seiamak Bahram.

Writing – review & editing: Pilar Domingo-Calap, Benjamin Schubert, Mélanie Joly, Morgane Solis, Meiggie Untrau, Raphael Carapito, Philippe Georgel, Sophie Caillard, Samira

Fafi-Kremer, Nicodème Paul, Oliver Kohlbacher, Fernando González-Candelas, Seiamak Bahram.

References

1. Holmes EC. Molecular clocks and the puzzle of RNA virus origins. *Journal of virology*. 2003; 77(7):3893–7. Epub 2003/03/14. <https://doi.org/10.1128/JVI.77.7.3893-3897.2003> PMID: 12634349; PubMed Central PMCID: PMC150674.
2. Ho SY, Lanfear R, Bromham L, Phillips MJ, Soubrier J, Rodrigo AG, et al. Time-dependent rates of molecular evolution. *Molecular ecology*. 2011; 20(15):3087–101. Epub 2011/07/12. <https://doi.org/10.1111/j.1365-294X.2011.05178.x> PMID: 21740474.
3. Sanjuan R. From molecular genetics to phylodynamics: evolutionary relevance of mutation rates across viruses. *PLoS pathogens*. 2012; 8(5):e1002685. Epub 2012/05/10. <https://doi.org/10.1371/journal.ppat.1002685> PMID: 22570614; PubMed Central PMCID: PMC3342999.
4. Duchene S, Holmes EC, Ho SY. Analyses of evolutionary dynamics in viruses are hindered by a time-dependent bias in rate estimates. *Proceedings Biological sciences*. 2014; 281(1786). Epub 2014/05/23. <https://doi.org/10.1098/rspb.2014.0732> PMID: 24850916; PubMed Central PMCID: PMC4046420.
5. 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. *Molecular biology and evolution*. 2010; 27(9):2038–51. Epub 2010/04/07. <https://doi.org/10.1093/molbev/msq088> PMID: 20363828; PubMed Central PMCID: PMC3107591.
6. Yasunaga T, Miyata T. Evolutionary changes of nucleotide sequences of papova viruses BKV and SV40: they are possibly hybrids. *Journal of molecular evolution*. 1982; 19(1):72–9. Epub 1982/01/01. PMID: 6298432.
7. Krumbholz A, Bininda-Emonds OR, Wutzler P, Zell R. Evolution of four BK virus subtypes. *Infection, genetics and evolution: journal of molecular epidemiology and evolutionary genetics in infectious diseases*. 2008; 8(5):632–43. Epub 2008/06/28. <https://doi.org/10.1016/j.meegid.2008.05.006> PMID: 18582602.
8. Faria NR, Azevedo Rdo S, Kraemer MU, Souza R, Cunha MS, Hill SC, et al. Zika virus in the Americas: Early epidemiological and genetic findings. *Science*. 2016; 352(6283):345–9. Epub 2016/03/26. <https://doi.org/10.1126/science.aaf5036> PMID: 27013429; PubMed Central PMCID: PMC4918795.
9. Carroll MW, Matthews DA, Hiscox JA, Elmore MJ, Pollakis G, Rambaut A, et al. Temporal and spatial analysis of the 2014–2015 Ebola virus outbreak in West Africa. *Nature*. 2015; 524(7563):97–101. Epub 2015/06/18. <https://doi.org/10.1038/nature14594> PMID: 26083749.
10. Phillips RE, Rowland-Jones S, Nixon DF, Gotch FM, Edwards JP, Ogunlesi AO, et al. Human immunodeficiency virus genetic variation that can escape cytotoxic T cell recognition. *Nature*. 1991; 354(6353):453–9. Epub 1991/12/12. <https://doi.org/10.1038/354453a0> PMID: 1721107.
11. Oldstone MB. How viruses escape from cytotoxic T lymphocytes: molecular parameters and players. *Virology*. 1997; 234(2):179–85. Epub 1997/08/04. <https://doi.org/10.1006/viro.1997.8674> PMID: 9268148.
12. Kloverpris HN, Leslie A, Goulder P. Role of HLA Adaptation in HIV Evolution. *Frontiers in immunology*. 2015; 6:665. Epub 2016/02/03. <https://doi.org/10.3389/fimmu.2015.00665> PMID: 26834742; PubMed Central PMCID: PMC4716577.
13. Gardner SD, Field AM, Coleman DV, Hulme B. New human papovavirus (B.K.) isolated from urine after renal transplantation. *Lancet*. 1971; 1(7712):1253–7. Epub 1971/06/19. PMID: 4104714.
14. Hirsch HH, Steiger J. Polyomavirus BK. *The Lancet Infectious diseases*. 2003; 3(10):611–23. Epub 2003/10/03. PMID: 14522260.
15. Hirsch HH, Knowles W, Dickenmann M, Passweg J, Klimkait T, Mihatsch MJ, et al. Prospective study of polyomavirus type BK replication and nephropathy in renal-transplant recipients. *The New England journal of medicine*. 2002; 347(7):488–96. Epub 2002/08/16. <https://doi.org/10.1056/NEJMoa020439> PMID: 12181403.
16. Rice SJ, Bishop JA, Apperley J, Gardner SD. BK virus as cause of haemorrhagic cystitis after bone marrow transplantation. *Lancet*. 1985; 2(8459):844–5. Epub 1985/10/12. PMID: 2864573.
17. Nickenleit V, Hirsch HH, Binet IF, Gudat F, Prince O, Dalquen P, et al. Polyomavirus infection of renal allograft recipients: from latent infection to manifest disease. *Journal of the American Society of Nephrology: JASN*. 1999; 10(5):1080–9. Epub 1999/05/08. PMID: 10232695.
18. Butcher D. Muller's ratchet, epistasis and mutation effects. *Genetics*. 1995; 141(1):431–7. Epub 1995/09/01. PMID: 8536988; PubMed Central PMCID: PMC1206738.

19. Cuevas JM, Domingo-Calap P, Sanjuan R. The fitness effects of synonymous mutations in DNA and RNA viruses. *Molecular biology and evolution*. 2012; 29(1):17–20. Epub 2011/07/21. <https://doi.org/10.1093/molbev/msr179> PMID: 21771719.
20. Domingo-Calap P, Sanjuan R. Experimental evolution of RNA versus DNA viruses. *Evolution; international journal of organic evolution*. 2011; 65(10):2987–94. Epub 2011/10/05. <https://doi.org/10.1111/j.1558-5646.2011.01339.x> PMID: 21967437.
21. Aiewsakun P, Katzourakis A. Time-dependent rate phenomenon in viruses. *Journal of virology*. 2016. Epub 2016/06/03. <https://doi.org/10.1128/JVI.00593-16> PMID: 27252529.
22. Martin DP, Lemey P, Lott M, Moulton V, Posada D, Lefevre P. RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics*. 2010; 26(19):2462–3. Epub 2010/08/28. <https://doi.org/10.1093/bioinformatics/btq467> PMID: 20798170; PubMed Central PMCID: PMC2944210.
23. Duffy S, Shackelton LA, Holmes EC. Rates of evolutionary change in viruses: patterns and determinants. *Nature reviews Genetics*. 2008; 9(4):267–76. Epub 2008/03/06. <https://doi.org/10.1038/nrg2323> PMID: 18319742.
24. Domingo-Calap P, Cuevas JM, Sanjuan R. The fitness effects of random mutations in single-stranded DNA and RNA bacteriophages. *PLoS genetics*. 2009; 5(11):e1000742. Epub 2009/12/04. <https://doi.org/10.1371/journal.pgen.1000742> PMID: 19956760; PubMed Central PMCID: PMC2776273.
25. Duffy S, Holmes EC. Validation of high rates of nucleotide substitution in geminiviruses: phylogenetic evidence from East African cassava mosaic viruses. *The Journal of general virology*. 2009; 90(Pt 6):1539–47. Epub 2009/03/07. <https://doi.org/10.1099/vir.0.009266-0> PMID: 19264617; PubMed Central PMCID: PMC4091138.
26. Shackelton LA, Parrish CR, Truyen U, Holmes EC. High rate of viral evolution associated with the emergence of carnivore parvovirus. *Proceedings of the National Academy of Sciences of the United States of America*. 2005; 102(2):379–84. Epub 2005/01/01. <https://doi.org/10.1073/pnas.0406765102> PMID: 15626758; PubMed Central PMCID: PMC544290.
27. Shackelton LA, Rambaut A, Pybus OG, Holmes EC. JC virus evolution and its association with human populations. *Journal of virology*. 2006; 80(20):9928–33. Epub 2006/09/29. <https://doi.org/10.1128/JVI.00441-06> PMID: 17005670; PubMed Central PMCID: PMC1617318.
28. Chen Y, Sharp PM, Fowkes M, Kocher O, Joseph JT, Koralnik IJ. Analysis of 15 novel full-length BK virus sequences from three individuals: evidence of a high intra-strain genetic diversity. *The Journal of general virology*. 2004; 85(Pt 9):2651–63. Epub 2004/08/11. <https://doi.org/10.1099/vir.0.79920-0> PMID: 15302959.
29. Takasaka T, Goya N, Ishida H, Tanabe K, Toma H, Fujioka T, et al. Stability of the BK polyomavirus genome in renal-transplant patients without nephropathy. *The Journal of general virology*. 2006; 87(Pt 2):303–6. Epub 2006/01/25. <https://doi.org/10.1099/vir.0.81368-0> PMID: 16432015.
30. Jahnke M, Holmes EC, Kerr PJ, Wright JD, Strive T. Evolution and phylogeography of the nonpathogenic calicivirus RCV-A1 in wild rabbits in Australia. *Journal of virology*. 2010; 84(23):12397–404. Epub 2010/09/24. <https://doi.org/10.1128/JVI.00777-10> PMID: 20861266; PubMed Central PMCID: PMC2976393.
31. Hatwell JN, Sharp PM. Evolution of human polyomavirus JC. *The Journal of general virology*. 2000; 81(Pt 5):1191–200. Epub 2000/04/18. <https://doi.org/10.1099/0022-1317-81-5-1191> PMID: 10769060.
32. Sakaoka H, Kurita K, Iida Y, Takada S, Umene K, Kim YT, et al. Quantitative analysis of genomic polymorphism of herpes simplex virus type 1 strains from six countries: studies of molecular evolution and molecular epidemiology of the virus. *The Journal of general virology*. 1994; 75 (Pt 3):513–27. Epub 1994/03/01. <https://doi.org/10.1099/0022-1317-75-3-513> PMID: 8126449.
33. Kolb AW, Ane C, Brandt CR. Using HSV-1 genome phylogenetics to track past human migrations. *PLoS one*. 2013; 8(10):e76267. Epub 2013/10/23. <https://doi.org/10.1371/journal.pone.0076267> PMID: 24146849; PubMed Central PMCID: PMC3797750.
34. Ong CK, Chan SY, Campo MS, Fujinaga K, Mavromara-Nazos P, Labropoulou V, et al. Evolution of human papillomavirus type 18: an ancient phylogenetic root in Africa and intratype diversity reflect coevolution with human ethnic groups. *Journal of virology*. 1993; 67(11):6424–31. Epub 1993/11/01. PMID: 8411344; PubMed Central PMCID: PMC238077.
35. Babkin IV, Babkina IN. Molecular dating in the evolution of vertebrate poxviruses. *Intervirology*. 2011; 54(5):253–60. Epub 2011/01/14. <https://doi.org/10.1159/000320964> PMID: 21228539.
36. Almeida RP, Bennett GM, Anhalt MD, Tsai CW, O'Grady P. Spread of an introduced vector-borne banana virus in Hawaii. *Molecular ecology*. 2009; 18(1):136–46. Epub 2008/11/29. <https://doi.org/10.1111/j.1365-294X.2008.04009.x> PMID: 19037897.
37. Babkin IV, Tyumentsev AI, Tikunov AY, Kurilshikov AM, Ryabchikova EI, Zhirakovskaya EV, et al. Evolutionary time-scale of primate bocaviruses. *Infection, genetics and evolution: journal of molecular*

- epidemiology and evolutionary genetics in infectious diseases. 2013; 14:265–74. Epub 2013/01/15. <https://doi.org/10.1016/j.meegid.2012.12.023> PMID: 23313830.
38. Parsyan A, Szmaragd C, Allain JP, Candotti D. Identification and genetic diversity of two human parvovirus B19 genotype 3 subtypes. *The Journal of general virology*. 2007; 88(Pt 2):428–31. Epub 2007/01/26. <https://doi.org/10.1099/vir.0.82496-0> PMID: 17251559.
 39. Shackelton LA, Holmes EC. Phylogenetic evidence for the rapid evolution of human B19 erythrovirus. *Journal of virology*. 2006; 80(7):3666–9. Epub 2006/03/16. <https://doi.org/10.1128/JVI.80.7.3666-3669.2006> PMID: 16537636; PubMed Central PMCID: PMC1440363.
 40. Firth C, Charleston MA, Duffy S, Shapiro B, Holmes EC. Insights into the evolutionary history of an emerging livestock pathogen: porcine circovirus 2. *Journal of virology*. 2009; 83(24):12813–21. Epub 2009/10/09. <https://doi.org/10.1128/JVI.01719-09> PMID: 19812157; PubMed Central PMCID: PMC2786836.
 41. Lefevre P, Harkins GW, Lett JM, Briddon RW, Chase MW, Moury B, et al. Evolutionary time-scale of the begomoviruses: evidence from integrated sequences in the *Nicotiana* genome. *PloS one*. 2011; 6(5):e19193. Epub 2011/05/24. <https://doi.org/10.1371/journal.pone.0019193> PMID: 21603653; PubMed Central PMCID: PMC3095596.
 42. Zhou Y, Holmes EC. Bayesian estimates of the evolutionary rate and age of hepatitis B virus. *Journal of molecular evolution*. 2007; 65(2):197–205. Epub 2007/08/09. <https://doi.org/10.1007/s00239-007-0054-1> PMID: 17684696.
 43. Osiowy C, Giles E, Tanaka Y, Mizokami M, Minuk GY. Molecular evolution of hepatitis B virus over 25 years. *Journal of virology*. 2006; 80(21):10307–14. Epub 2006/10/17. <https://doi.org/10.1128/JVI.00996-06> PMID: 17041211; PubMed Central PMCID: PMC1641782.
 44. Hannoun C, Horal P, Lindh M. Long-term mutation rates in the hepatitis B virus genome. *The Journal of general virology*. 2000; 81(Pt 1):75–83. Epub 2000/01/21. <https://doi.org/10.1099/0022-1317-81-1-75> PMID: 10640544.
 45. Lemey P, Van Dooren S, Vandamme AM. Evolutionary dynamics of human retroviruses investigated through full-genome scanning. *Molecular biology and evolution*. 2005; 22(4):942–51. Epub 2005/01/07. <https://doi.org/10.1093/molbev/msi078> PMID: 15635055.
 46. Carpi G, Holmes EC, Kitchen A. The evolutionary dynamics of bluetongue virus. *Journal of molecular evolution*. 2010; 70(6):583–92. Epub 2010/06/08. <https://doi.org/10.1007/s00239-010-9354-y> PMID: 20526713.
 47. Lahon A, Walimbe AM, Chitambar SD. Full genome analysis of group B rotaviruses from western India: genetic relatedness and evolution. *The Journal of general virology*. 2012; 93(Pt 10):2252–66. Epub 2012/07/21. <https://doi.org/10.1099/vir.0.043497-0> PMID: 22815276.
 48. Stenger DC, Sisterson MS, French R. Population genetics of *Homalodisca vitripennis* reovirus validates timing and limited introduction to California of its invasive insect host, the glassy-winged sharpshooter. *Virology*. 2010; 407(1):53–9. Epub 2010/08/27. <https://doi.org/10.1016/j.virol.2010.08.002> PMID: 20739043.
 49. Tong YG, Shi WF, Liu D, Qian J, Liang L, Bo XC, et al. Genetic diversity and evolutionary dynamics of Ebola virus in Sierra Leone. *Nature*. 2015; 524(7563):93–6. Epub 2015/05/15. <https://doi.org/10.1038/nature14490> PMID: 25970247.
 50. Liu L, Chen W, Yang Y, Jiang Y. Molecular evolution of fever, thrombocytopenia and leukocytopenia virus (FTLSV) based on whole-genome sequences. *Infection, genetics and evolution: journal of molecular epidemiology and evolutionary genetics in infectious diseases*. 2016; 39:55–63. Epub 2016/01/10. <https://doi.org/10.1016/j.meegid.2015.12.022> PMID: 26748010.
 51. Chen R, Holmes EC. Avian influenza virus exhibits rapid evolutionary dynamics. *Molecular biology and evolution*. 2006; 23(12):2336–41. Epub 2006/09/02. <https://doi.org/10.1093/molbev/msl102> PMID: 16945980.
 52. Gachara G, Symekher S, Otieno M, Magana J, Opot B, Bulimo W. Whole genome characterization of human influenza A(H1N1)pdm09 viruses isolated from Kenya during the 2009 pandemic. *Infection, genetics and evolution: journal of molecular epidemiology and evolutionary genetics in infectious diseases*. 2016; 40:98–103. Epub 2016/02/28. <https://doi.org/10.1016/j.meegid.2016.02.029> PMID: 26921801.
 53. Chao YC, Tang HS, Hsu CT. Evolution rate of hepatitis delta virus RNA isolated in Taiwan. *Journal of medical virology*. 1994; 43(4):397–403. Epub 1994/08/01. PMID: 7964650.
 54. Tan L, Lemey P, Houspie L, Viveen MC, Jansen NJ, van Loon AM, et al. Genetic variability among complete human respiratory syncytial virus subgroup A genomes: bridging molecular evolutionary dynamics and epidemiology. *PloS one*. 2012; 7(12):e51439. Epub 2012/12/14. <https://doi.org/10.1371/journal.pone.0051439> PMID: 23236501; PubMed Central PMCID: PMC3517519.

55. Davis PL, Bourhy H, Holmes EC. The evolutionary history and dynamics of bat rabies virus. *Infection, genetics and evolution: journal of molecular epidemiology and evolutionary genetics in infectious diseases*. 2006; 6(6):464–73. Epub 2006/04/20. <https://doi.org/10.1016/j.meegid.2006.02.007> PMID: 16621724.
56. Bird BH, Khristova ML, Rollin PE, Ksiazek TG, Nichol ST. Complete genome analysis of 33 ecologically and biologically diverse Rift Valley fever virus strains reveals widespread virus movement and low genetic diversity due to recent common ancestry. *Journal of virology*. 2007; 81(6):2805–16. Epub 2006/12/29. <https://doi.org/10.1128/JVI.02095-06> PMID: 17192303; PubMed Central PMCID: PMC1865992.
57. McKinley ET, Jackwood MW, Hilt DA, Kissinger JC, Robertson JS, Lemke C, et al. Attenuated live vaccine usage affects accurate measures of virus diversity and mutation rates in avian coronavirus infectious bronchitis virus. *Virus research*. 2011; 158(1–2):225–34. Epub 2011/05/05. <https://doi.org/10.1016/j.virusres.2011.04.006> PMID: 21539870.
58. Wu B, Blanchard-Letort A, Liu Y, Zhou G, Wang X, Elena SF. Dynamics of molecular evolution and phylogeography of Barley yellow dwarf virus-PAV. *PLoS one*. 2011; 6(2):e16896. Epub 2011/02/18. <https://doi.org/10.1371/journal.pone.0016896> PMID: 21326861; PubMed Central PMCID: PMC3033904.
59. Klungthong C, Zhang C, Mammen MP Jr., Ubol S, Holmes EC. The molecular epidemiology of dengue virus serotype 4 in Bangkok, Thailand. *Virology*. 2004; 329(1):168–79. Epub 2004/10/13. <https://doi.org/10.1016/j.virol.2004.08.003> PMID: 15476884.
60. Yoon SH, Park W, King DP, Kim H. Phylogenomics and molecular evolution of foot-and-mouth disease virus. *Molecules and cells*. 2011; 31(5):413–21. Epub 2011/03/31. <https://doi.org/10.1007/s10059-011-0249-6> PMID: 21448588; PubMed Central PMCID: PMC3887601.
61. Kulkarni MA, Walimbe AM, Cherian S, Arankalle VA. Full length genomes of genotype IIIA Hepatitis A Virus strains (1995–2008) from India and estimates of the evolutionary rates and ages. *Infection, genetics and evolution: journal of molecular epidemiology and evolutionary genetics in infectious diseases*. 2009; 9(6):1287–94. Epub 2009/09/03. <https://doi.org/10.1016/j.meegid.2009.08.009> PMID: 19723592.
62. Gray RR, Parker J, Lemey P, Salemi M, Katzourakis A, Pybus OG. The mode and tempo of hepatitis C virus evolution within and among hosts. *BMC evolutionary biology*. 2011; 11:131. Epub 2011/05/21. <https://doi.org/10.1186/1471-2148-11-131> PMID: 21595904; PubMed Central PMCID: PMC3112090.
63. Mohammed MA, Galbraith SE, Radford AD, Dove W, Takasaki T, Kurane I, et al. Molecular phylogenetic and evolutionary analyses of Muar strain of Japanese encephalitis virus reveal it is the missing fifth genotype. *Infection, genetics and evolution: journal of molecular epidemiology and evolutionary genetics in infectious diseases*. 2011; 11(5):855–62. Epub 2011/03/01. <https://doi.org/10.1016/j.meegid.2011.01.020> PMID: 21352956.
64. Cotten M, Watson SJ, Zumla AI, Makhdoom HQ, Palser AL, Ong SH, et al. Spread, circulation, and evolution of the Middle East respiratory syndrome coronavirus. *mBio*. 2014; 5(1). Epub 2014/02/20. <https://doi.org/10.1128/mBio.01062-13> PMID: 24549846; PubMed Central PMCID: PMC3944817.
65. Yoon SH, Kim H, Kim J, Lee HK, Park B. Complete genome sequences of porcine reproductive and respiratory syndrome viruses: perspectives on their temporal and spatial dynamics. *Molecular biology reports*. 2013. Epub 2013/10/15. <https://doi.org/10.1007/s11033-013-2802-1> PMID: 24122560.
66. Padhi A, Ma L. Molecular evolutionary and epidemiological dynamics of genotypes 1G and 2B of rubella virus. *PLoS one*. 2014; 9(10):e110082. Epub 2014/10/21. <https://doi.org/10.1371/journal.pone.0110082> PMID: 25329480; PubMed Central PMCID: PMC4201520.
67. Zhao Z, Li H, Wu X, Zhong Y, Zhang K, Zhang YP, et al. Moderate mutation rate in the SARS coronavirus genome and its implications. *BMC evolutionary biology*. 2004; 4:21. Epub 2004/06/30. <https://doi.org/10.1186/1471-2148-4-21> PMID: 15222897; PubMed Central PMCID: PMC446188.
68. Baillie GJ, Kolokotronis SO, Waltari E, Maffei JG, Kramer LD, Perkins SL. Phylogenetic and evolutionary analyses of St. Louis encephalitis virus genomes. *Molecular phylogenetics and evolution*. 2008; 47(2):717–28. Epub 2008/04/01. <https://doi.org/10.1016/j.ympev.2008.02.015> PMID: 18374605.
69. Auguste AJ, Volk SM, Arrigo NC, Martinez R, Ramkissoon V, Adams AP, et al. Isolation and phylogenetic analysis of Mucambo virus (Venezuelan equine encephalitis complex subtype IIIA) in Trinidad. *Virology*. 2009; 392(1):123–30. Epub 2009/07/28. <https://doi.org/10.1016/j.virol.2009.06.038> PMID: 19631956; PubMed Central PMCID: PMC2804100.
70. Drake JW. A constant rate of spontaneous mutation in DNA-based microbes. *Proceedings of the National Academy of Sciences of the United States of America*. 1991; 88(16):7160–4. Epub 1991/08/15. PMID: 1831267; PubMed Central PMCID: PMC52253.
71. Darbinyan A, Siddiqui KM, Slonina D, Darbinian N, Amini S, White MK, et al. Role of JC virus agnoprotein in DNA repair. *Journal of virology*. 2004; 78(16):8593–600. Epub 2004/07/29. <https://doi.org/10.1128/JVI.78.16.8593-8600.2004> PMID: 15280468; PubMed Central PMCID: PMC479055.
72. Hicks AL, Duffy S. Cell tropism predicts long-term nucleotide substitution rates of mammalian RNA viruses. *PLoS pathogens*. 2014; 10(1):e1003838. Epub 2014/01/15. <https://doi.org/10.1371/journal.ppat.1003838> PMID: 24415935; PubMed Central PMCID: PMC3887100.

73. Hadaya K, de Rham C, Bandelier C, Ferrari-Lacraz S, Jendly S, Berney T, et al. Natural killer cell receptor repertoire and their ligands, and the risk of CMV infection after kidney transplantation. *American journal of transplantation: official journal of the American Society of Transplantation and the American Society of Transplant Surgeons*. 2008; 8(12):2674–83. Epub 2008/11/27. <https://doi.org/10.1111/j.1600-6143.2008.02431.x> PMID: 19032228.
74. Khakoo SI, Thio CL, Martin MP, Brooks CR, Gao X, Astemborski J, et al. HLA and NK cell inhibitory receptor genes in resolving hepatitis C virus infection. *Science*. 2004; 305(5685):872–4. Epub 2004/08/07. <https://doi.org/10.1126/science.1097670> PMID: 15297676.
75. Babel N, Volk HD, Reinke P. BK polyomavirus infection and nephropathy: the virus-immune system interplay. *Nature reviews Nephrology*. 2011; 7(7):399–406. Epub 2011/05/26. <https://doi.org/10.1038/nrneph.2011.59> PMID: 21610680.
76. Trydzenskaya H, Juerchott K, Lachmann N, Kotsch K, Kunert K, Weist B, et al. The genetic predisposition of natural killer cell to BK virus-associated nephropathy in renal transplant patients. *Kidney international*. 2013; 84(2):359–65. Epub 2013/03/15. <https://doi.org/10.1038/ki.2013.59> PMID: 23486513.
77. Kumar A, Yadav IS, Hussain S, Das BC, Bharadwaj M. Identification of immunotherapeutic epitope of E5 protein of human papillomavirus-16: An in silico approach. *Biologicals: journal of the International Association of Biological Standardization*. 2015; 43(5):344–8. Epub 2015/07/28. <https://doi.org/10.1016/j.biologicals.2015.07.002> PMID: 26212000.
78. Kumar A, Hussain S, Yadav IS, Gissmann L, Natarajan K, Das BC, et al. Identification of human papillomavirus-16 E6 variation in cervical cancer and their impact on T and B cell epitopes. *Journal of virological methods*. 2015; 218:51–8. Epub 2015/03/25. <https://doi.org/10.1016/j.jviromet.2015.03.008> PMID: 25800725.
79. Bi J, Yang H, Yan H, Song R, Fan J. Knowledge-based virtual screening of HLA-A*0201-restricted CD8 + T-cell epitope peptides from herpes simplex virus genome. *Journal of theoretical biology*. 2011; 281(1):133–9. Epub 2011/05/03. <https://doi.org/10.1016/j.jtbi.2011.04.018> PMID: 21530544.
80. Tamura K, Stecher G, Peterson D, Filipinski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular biology and evolution*. 2013; 30(12):2725–9. Epub 2013/10/18. <https://doi.org/10.1093/molbev/mst197> PMID: 24132122; PubMed Central PMCID: PMC3840312.
81. Luo C, Bueno M, Kant J, Martinson J, Randhawa P. Genotyping schemes for polyomavirus BK, using gene-specific phylogenetic trees and single nucleotide polymorphism analysis. *Journal of virology*. 2009; 83(5):2285–97. Epub 2008/12/26. <https://doi.org/10.1128/JVI.02180-08> PMID: 19109389; PubMed Central PMCID: PMC2643714.
82. Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic biology*. 2003; 52(5):696–704. Epub 2003/10/08. PMID: 14530136.
83. Darriba D, Taboada GL, Doallo R, Posada D. jModelTest 2: more models, new heuristics and parallel computing. *Nature methods*. 2012; 9(8):772. Epub 2012/08/01. <https://doi.org/10.1038/nmeth.2109> PMID: 22847109; PubMed Central PMCID: PMC4594756.
84. Rambaut A, Lam TT, Max Carvalho L, Pybus OG. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus evolution*. 2016; 2(1):vew007. Epub 2016/10/25. <https://doi.org/10.1093/ve/vew007> PMID: 27774300; PubMed Central PMCID: PMC4989882.
85. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular biology and evolution*. 2012; 29(8):1969–73. Epub 2012/03/01. <https://doi.org/10.1093/molbev/mss075> PMID: 22367748; PubMed Central PMCID: PMC3408070.
86. Baele G, Li WL, Drummond AJ, Suchard MA, Lemey P. Accurate model selection of relaxed molecular clocks in bayesian phylogenetics. *Molecular biology and evolution*. 2013; 30(2):239–43. Epub 2012/10/24. <https://doi.org/10.1093/molbev/mss243> PMID: 23090976; PubMed Central PMCID: PMC3548314.
87. Volz EM, Frost SD. Scalable relaxed clock phylogenetic dating. *Virus evolution*. 2017; 3(2):vex025.
88. Nielsen M, Lundegaard C, Worning P, Lauemoller SL, Lamberth K, Buus S, et al. Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. *Protein science: a publication of the Protein Society*. 2003; 12(5):1007–17. Epub 2003/04/30. <https://doi.org/10.1110/ps.0239403> PMID: 12717023; PubMed Central PMCID: PMC2323871.
89. Nielsen M, Lundegaard C, Lund O, Kesmir C. The role of the proteasome in generating cytotoxic T-cell epitopes: insights obtained from improved predictions of proteasomal cleavage. *Immunogenetics*. 2005; 57(1–2):33–41. Epub 2005/03/04. <https://doi.org/10.1007/s00251-005-0781-7> PMID: 15744535.