RESEARCH ARTICLE

Identification of Two Novel Members of the Tentative Genus *Wukipolyomavirus* in Wild Rodents

Juozas Nainys[1], Albertas Timinskas[1], Julia Schneider[2], Rainer G. Ulrich[2], Alma Gedvilaite[1]*

1 Department of Eukaryote Genetic Engineering, Institute of Biotechnology, Vilnius University, Vilnius, Lithuania, 2 Institute for Novel and Emerging Infectious Diseases, Friedrich-Loeffler-Institut, Federal Research Institute for Animal Health, Greifswald-Insel Riems, Germany

* alma.gedvilaite@bti.vu.lt

Abstract

Two novel polyomaviruses (PyVs) were identified in kidney and chest-cavity fluid samples of wild bank voles (*Myodes glareolus*) and common voles (*Microtus arvalis*) collected in Germany. All cloned and sequenced genomes had the typical PyV genome organization, including putative open reading frames for early regulatory proteins large T antigen and small T antigen on one strand and for structural late proteins (VP1, VP2 and VP3) on the other strand. Virus-like particles (VLPs) were generated by yeast expression of the VP1 protein of both PyVs. VLP-based ELISA and large T-antigen sequence-targeted polymerase-chain reaction investigations demonstrated signs of infection of these novel PyVs in about 42% of bank voles and 18% of common voles. In most cases only viral DNA, but not VP1-specific antibodies were detected. In additional animals exclusively VP1-specific antibodies, but no viral DNA was detected, indicative for virus clearance. Phylogenetic and clustering analysis including all known PyV genomes placed novel bank vole and common vole PyVs amongst members of the tentative *Wukipolymavirus* genus. The other known four rodent PyVs, Murine PyV and Hamster PyV, and Murine pneumotropic virus and Mastomys PyV belong to different phylogenetic clades, tentatively named *Orthopolyomavirus* I and *Orthopolyomavirus* II, respectively. In conclusion, the finding of novel vole-borne PyVs may suggest an evolutionary origin of ancient wukipolyomaviruses in rodents and may offer the possibility to develop a vole-based animal model for human wukipolyomaviruses.

Introduction

The members of the *Polyomaviridae* family are small and non-enveloped DNA viruses. The circular dsDNA genome of polyomaviruses (PyVs) of approximately 5000 base-pairs (bp) is associated with host cell histone proteins H2A, H2B, H3 and H4 in a chromatin-like, super-coiled complex [1]. All PyVs possess a very similar genome organization. Five to nine proteins are encoded by early and late transcriptional regions separated by a non-coding control region

(NCCR) which controls the transcription of both early and late promoters and contain the origin of DNA replication (ORI) [2, 3]. Early non-structural tumor (T) proteins are involved in the interference with cell cycle regulation and, in some cases, induction of cellular transformation or tumor formation [4, 5]. Late region usually encodes three structural proteins VP1, VP2 and VP3 which are translated from partially overlapping open reading frames (ORFs) and form icosahedral capsids of approximately 40–45 nm in diameter [3]. The VP1 protein is the major capsid protein which interacts with the cellular receptor molecule and is responsible for cellular entry [1]. Minor capsid proteins VP2 and VP3 may be necessary for specific encapsidation of the polyomavirus genome after genome replication [1]. Some of the PyVs encode additional regulatory and structural proteins known as VP4, agnoproteins or alternative large T ORF protein (ALTO) [6, 7]. The unique protein ALTO is encoded by several PyVs including Murine PyV (MPyV) and Hamster PyV (HaPyV) [7]. These additional proteins may act in virus release and/or genome replication, regulation of the cell cycle and DNA repair and/or capsid assembly or may even be incorporated into the viral capsid [6, 8–11].

Improvements in molecular diagnostic techniques led to identification of new polyomaviruses in different mammal, bird and fish species [12–17]. Because of growing quantity and variety of PyVs, the *Polyomaviridae* study group of the International Committee on Taxonomy of Viruses (ICTV) has recently suggested the division of the polyomaviruses into three tentative genera: *Orthopolyomavirus*, *Wukipolyomavirus* and *Avipolyomavirus* [3]. Based on the phylogenetic tree generated from an alignment of concatenated VP1, VP2 and LTag amino acid sequences, other authors have divided PyVs into five lineages: *Orthopolyomavirus I* and *II*, *Wukipolyomavirus*, *Avipolyomavirus* and *Malawipolyomavirus* [12]. Until now four PyVs were found in rodents. Rodent-associated MPyV and HaPyV belong to the tentative genus *Orthopolyomavirus I* [3, 18–20], whereas MPtV and Mastomys PyV (MasPyV) are members of the tentative genus *Orthopolyomavirus II*, [12, 16, 21]. Similar to MPtV, MasPyV was found to encode only two T antigens, large T antigen (LTag) and small T antigen (STag), whereas MPyV and HaPyV encode a third middle T antigen (MTag). Multimammate mouse (*Mastomys* spp.*)* and house mouse (*Mus musculus*), reservoir of MasPyV, MPyV and MPtV, belong to the family Muridae within the order Rodentia. In contrast to the Syrian hamster (*Mesocricetus auratus*) which is host of HaPyV and belong to subfamily Cricetinae, voles, like the bank vole (*Myodes glareolus)* and the common vole (*Microtus arvalis)*, represent another subfamily Arvicolinae of the rodent family Cricetidae. Both vole species show fluctuations in their population densities that result in high economic losses during peak densities in agriculture and/or forestry industry [22]. In addition, both species harbor important zoonotic pathogens, such as two different hantaviruses, cowpox virus and leptospires [23–25].

In this study, we report the identification, characterization and regional prevalence of two new wild rodent PyVs: bank vole polyomavirus (BVPyV) and common vole polyomavirus (CVPyV). Cloning, sequencing, ORF identification, phylogenetic and clustering analyses of the entire new BVPyV and CVPyV genomes let us to attribute these two new viruses to the *Wukipolyomavirus* genus.

Methods

Sample Collection

The animals were handled according to the national and European legislation, namely the EU council directive 86/609/EEC for the protection of animals. Collection of samples was carried out in accordance with institutional guidelines and permission of the national authorities. Bank and common voles were collected within hantavirus monitoring projects during 2011–2013 in four federal states of Germany [26] (Reil, Rosenfeld, Drewes, Schmidt, Turni,

Wagner-Wiening, Jacob and Ulrich, unpublished data; see S1 and S2 Figs). Permits were issued by Regierungspräsidium Stuttgart; Landwirtschaft, Ländlicher Raum, Veterinär- und Lebensmittelwesen, Stuttgart (Baden-Wuerttemberg 35–9185.82/0261), by Landesamt für Natur, Umwelt und Verbraucherschutz Nordrhein-Westfalen, Recklinghausen (North Rhine-Westphalia 8.87–51.05.20.09.210), by Landesamt für Landwirtschaft, Lebensmittelsicherheit und Fischerei Mecklenburg-Vorpommern, Rostock (Mecklenburg-Western Pomerania 7221.3-030/09), and by Thüringer Landesamt für Lebensmittelsicherheit und Verbraucherschutz, Bad Langensalza (Thuringia 22-2684-04-15-107/09). Additional common voles were collected within the network "Rodent-borne pathogens" between 2004 and 2012 in seven federal states of Germany (Reil, Rosenfeld, Drewes, Schmidt, Turni, Wagner-Wiening, Schmidt and Ulrich, unpublished data) [27]. These animals were trapped and sacrificed by cooperation partners from forestry and consumer protection and food safety institutions during their official duties as part of their pest control measures. Therefore, specific animal ethics permits were not required.

After snap trapping or live trapping and subsequent cervical dislocation, animals were immediately frozen and stored at -20°C. Dissection and sample collection followed standard protocols as published recently [28]. Briefly, the frozen carcasses were thawed overnight and tissue samples taken during dissection in a defined order, starting with heart, then lung, liver, spleen, kidney, brain and intestine. For serological investigations, chest cavity fluid (CCF) was collected by washing the chest cavity with 1 ml phosphate-buffered saline. CCF and tissue samples were stored at -20°C until investigation.

Morphological species identification was verified by polymerase-chain reaction (PCR) amplification and sequencing of a partial cytochrome b gene, using DNA obtained from tail, heart or liver tissue following a previously published protocol [29].

DNA extraction and amplification

Viral DNA was extracted from 200 μL of wild rodent CCF or kidney samples using GeneJET Viral DNA and RNA Purification Kit according to manufacturer's recommendations (Thermo Fisher Scientific Baltics, Vilnius, Lithuania). Kidney samples were additionally treated for 1h with proteinase K (Thermo Fisher Scientific Baltics) before using GeneJET Viral DNA and RNA Purification Kit. Half of the sample of extracted viral DNA was used for PCR directly and other half of DNA was enriched and amplified by rolling circle amplification (RCA) as described previously [30, 31]. Briefly, 25 μL of viral DNA solution was treated with Plasmid-Safe ATP-Dependent DNase (Epicentre Biotechnologies, Madison, WI, USA). Then DNA was precipitated and amplified using phi29 DNA polymerase (Thermo Fisher Scientific Baltics) and exonuclease-resistant random primer (Thermo Fisher Scientific Baltics).

PCR

Diagnostic PCR was carried out using DreamTaq Green DNA Polymerase (Thermo Fisher Scientific Baltics). The initial search of PyVs was performed according to a previously described nested PCR protocol and PyV primers T-1f, T-1r, T-2f, T-2r (see S1 Table) [32]. For the determination of the prevalence of the new PyVs two sets of primers were used: BVPyV-T-D3, BVPyV-T-R3, CVPyV-T-D3 and CVPyV-T-R3 (see S1 Table).

New complete PyV genome and VP1-encoding sequences were amplified using Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific Baltics) and primers BVPyV-D1, BVPyV-R1, CVPyV-D1, CVPyV-R1, BVPyV-ATG, BVPyV-Stop, CVPyV-ATG and CVPyV-Stop (see S1 Table).

Generation of plasmids and sequence determination

All DNA manipulations were performed according to standard procedures [33]. Enzymes and kits for DNA manipulations were purchased from Thermo Fisher Scientific Baltics. Recombinants were screened in *E. coli* DH10B cells. The PCR amplified PyV genomes and DNA fragments were cloned into pJET1.2 vector using CloneJET PCR Cloning Kit (Thermo Fisher Scientific Baltics) and sequenced using a BigDye Terminator v3.0 Cycle Sequencing kit on an ABI Prism 3130 Genetic Analyzer (Applied Biosystems Hitachi, Tokyo, Japan). Sequencing primers are listed in S1 Table. The PCR amplified, cloned and determined BVPyV and CVPyV partial or complete genome sequences were compared to the GenBank nonredundant nucleotide and protein databases using BLASTn and BLASTx, respectively.

The novel genome sequences were submitted to GenBank: BVPyV KS/14/281(GenBank: KR612368), BVPyV KS/14/289 (GenBank: KR612369), BVPyV KS/14/328 (GenBank: KR612370), BVPyV KS/14/336 (GenBank: KR612371), BVPyV KS/13/999 (GenBank: KR612372), CVPyV KS/13/947 (GenBank: KR612373), CVPyV KS/13/980 (GenBank: KR612374).

PCR-amplified BVPyV and CVPyV VP1-encoding sequences were cloned into previously described yeast expression vector pFX7 for target protein expression [34].

Gene and protein annotation for the novel BVPyV and CVPyV genomes

Search for protein coding ORFs in PyV genomes was performed in two steps. First, all ORFs in the genome that begin at traditional initiation codons (AUG and GUG) were analyzed. Second, to investigate the presence of proteins that might be produced by alternative splicing of the early or late transcripts, putative exons of LTag, VP2 and VP1 genes were translated by all six open reading frames and full amino acid sequences were analyzed within non-redundant protein sequence database at NCBI by means of BLASTp [35]. ORF identification was based on splice donor/acceptor site analyses and alternatively similarity search with annotated proteins of other PyVs. First, possible splicing variants were analyzed by submitting LTag gene sequences to Human Splicing Finder database [36]. Then, splicing sites were evaluated in the context of protein sequences that they produce, and exons were identified based on homology to known PyV LTag proteins (S2 Table). ORI region was analyzed based on identified and defined SV40 and MCPyV ORI sequences [37–39].

Expression, purification and characterization of recombinant VP1 proteins

The plasmids pFX7-BVPyV-VP1 or pFX7-CVPyV-VP1 were transformed into *Saccharomyces cerevisiae* strain AH22-214 (*a, leu2-3,112, his4-519*). Yeast transformants harboring plasmids with VP1 encoding genes were grown as described previously [40]. Yeast biomass harboring recombinant proteins was harvested by centrifugation and was stored at -20°C until use. The recombinant proteins were purified using a previously described protocol of ultracentrifugation in sucrose and cesium chloride gradients [40].

To prove formation of VP1-derived VLPs the samples of purified recombinant PyV-VP1 proteins were placed on 200-mesh carbon coated gold grids, negatively stained with 2% aqueous uranyl acetate solution and examined using Morgagni 268 electron microscope (FEI Inc., Hillsboro, OR, USA). The generation and characterization of Washington University PyV (WUPyV)-derived VP1 VLPs has been described recently [41].

Indirect ELISA

The bank vole and common vole CCF samples were tested by an indirect ELISA. The microtiter plates with F-shape cavities (Nerbe Plus, Winsen/Luhe, Germany) were coated with 100 μL per well of recombinant BVPyV-, CVPyV- or WUPyV-derived VP1 VLPs and measles virus nucleoprotein [42] for negative control diluted in coating buffer (0.1 M sodium carbonate, pH 9.5) to a concentration of 2 μg mL^{-1} and incubated overnight at +37°C. The plates then were washed 4 times with washing buffer (10mM Tris, 30mM NaCl, 0.01% Tween-20, pH 7.2). After washing the coated plates were blocked by addition of 100 μL ROTI®-BLOCK (ROTH, Karlsruhe, Germany) per well and incubated for 1 h at room temperature (RT) on a plate shaker. Thereafter, the plates were rinsed four times with washing buffer. Wild rodent CCF samples were serially diluted (1:50, 1:100 and 1:200) with washing buffer containing 3% bovine serum albumin (BSA) and 100 μL of such diluted serum was added to each well. The polyclonal anti-WUPyV-VP1 mouse serum (raised against the WUPyV VP1 protein) was diluted 1:300 and used as positive control [41]. After incubation for 2 h at +37°C the plates were washed 8 times with washing buffer and incubated for 1 h with 100 μL/well of horseradish peroxidase (HRP)-conjugated goat anti-mouse IgG (BioRad, Hercules, CA, USA) diluted 1:600 in washing buffer. The plates were rinsed 5 times with washing buffer. HRP activity was detected by the addition of 100 μL of ready-to-use TMB Stabilized Chromogen substrate (Invitrogen, Camarillo, CA, USA) to each well. The plates were incubated for 30 min at RT in the dark, and the reaction was stopped by adding 150 μL of 10% sulphuric acid per well. The optical density (OD) was measured at 450 nm (reference filter 620 nm) in a microplate reader (Tecan, Groedig, Austria).

Polyomavirus entire genome sequence data

Two strategies were implemented to extract all entire PyV genomes known to date (2015/27/04). First, 77 PyV RefSeq representative genomes as well as so called 'PyV neighbors' genome data according to the information on the internet site [43] were downloaded from National Center for Biotechnology Information (NCBI) repositories. Secondly, all NCBI nucleotide databases were screened for complete PyV genomes using data mining techniques. DNA sequences shorter than 3500 or longer than 8000 nucleotides (partial non-integrated or completely or partially integrated genomes) were skipped and remaining sequences were manually reviewed to confirm their PyV origin. At least one of T antigens and at least one of capsid proteins typical for PyVs (commonly LTag and VP1 were strongest representatives) were required to be present. The latter test was carried out using BLAST search for homologous proteins in NCBI protein database for candidate protein while expecting high extent of "polyomavirus" keyword for top ranked homologs found. If no "polyomavirus" protein was found using this method, a TBLASTN search was performed through the tested genome using known PyV proteins (LTag, STag, VP1, VP2) as bait. Results are described in detail in the PyV protein sequence data section below and S3 Table. Finally, both sets of genome sequences compiled by use of these two strategies mentioned above were combined and after removing duplicates it yielded an initial set of 1316 complete PyV genomes (including the 7 newly sequenced vole PyV genomes). Genome sequences of 15 PyVs had one (10 PyVs) or more (5 PyVs) ambiguous nucleotides that are not tolerated by some of the used methods. In order not to lose potentially unique genomes (by removing them from the dataset) the ambiguous nucleotides were replaced by most common regular nucleotides found at corresponding regions (identified by searching with a 41 nucleotide sequence centered on the ambiguous nucleotide) in other PyV genomes (all replacements and justifications can be viewed in S3 Table). Three genomes (namely JCPyV strain SA50329_06, and two HaPyV genome sequences: X02449 and PPCCGAAA), that had long unassigned nucleotide (N) stretches with no strong suggestions for replacement, were removed. Thereafter, the final set contained 1313 PyV

genomes. Prior to viral DNA sequence clustering, a non-redundant genome sequence dataset was constructed as follows: genome-to-genome pairwise comparisons (described below) were used to build a set of corresponding DNA strands of PyV complete genomes and after that CD-HIT [44] was used to remove identical sequences. This resulted in a subset of 1029 PyV genomes (S3 Table). As a representative, but smaller PyV dataset was needed for phylogenetic analysis, CD-HIT was also used and sequences were filtered to 85% identity level. The used identity level of 85% is similar to genome sequence identity level (81–84%) which was proposed as criterion by the *Polyomaviridae* study group of ICTV to define separate PyV species [3]. NCBI RefSeq representatives were picked to represent CD-HIT clusters when available. After these manipulations a final set of 98 representative complete PyV genomes was obtained (S3 Table).

Polyomavirus protein sequence data

FASTA libraries of protein sequences for LTag, STag, VP1 and VP2 were compiled using the two final sets of genomes (complete PyV genome sets of 1029 and 98 representatives). The fitness of selected representative sequences was further evaluated, according to each of the four proteins. Protein sequences of each of the 98 clusters were aligned using MAFFT [45] (L-INS-i algorithm) and resulting alignments were examined using JALVIEW [46, 47] to confirm the presence of typical cluster features in the selected representative sequences. Seven representative protein sequences were manually revised: missing C-terminal parts of protein (coded on separate exons) were added in case of two LTag proteins (MCPyV isolate MKL-1 and MCPyV isolate MCC339); the VP1 protein of Sea otter polyomavirus 1 isolate 6831–13 has not been previously annotated and was added to the set after TBLASTN search using VP1 sequences from other PyVs; the putative VP1 protein of bat polyomavirus isolate BtMp-PyV/SX2013 was extended at the N-terminus to fit better to other VP1 proteins. Full details on protein representative sequence rearrangements are given in S3 Table. We failed to identify STag sequences in black sea bass and guitarfish PyV genomes, despite having used TBLASTN and PSI-TBLASTN methods to scan corresponding genomes with all other representative STags and constructed STag profiles as baits, respectively. STag profiles were built with PSI-BLAST [48] using non-redundant protein set (further filtered to 80% sequence identity with CD-HIT) obtained from NCBI. Obtained representative protein datasets were used for phylogenetic analysis.

Multiple sequence alignments for phylogenetic analysis

Multiple sequence alignments for the four protein datasets were produced using MAFFT (L-INS-i algorithm). By means of PSI-BLAST [48], these alignments were used to find close relatives of the analyzed proteins within non-redundant protein sequence database at NCBI as well as homologous proteins with known 3-dimensional (3D) structure to be used for alignment refinement. For each of the four sets of proteins, obtained sequences were aligned using L-INS-i (two versions: with "-ep 0.123" parameter (which allow higher restrictions for gap extension as recommended at [49]) or without [default]) and PROMALS3D [50–51]. PDB structures used for LTag: 4gdfA and 3pf4 (represents SV40),4bnpA (JCPyV), 3qfqA(MCPyV); for STag: 2pf4E (SV40), 1fafA (MPyV); for VP1: 1sieA (MPyV), 4mbyA (B-lymphotropic PyV), 4fmh (MCPyV), 4posA (HPyV9,), 4mj0A (BKPyV), 3bwrA (SV40), 4x16 (JCPyV), 1cn3A (MPyV strain Crawford), 4pcgA (HPyV6), 4pchA (HPyV7), 3s7x (WUPyV), 3s7vA (KIPyV); for VP2: 1cn3F (structural data for 29 aa fragment VP2 of MPyV, small plaque variant of strain Crawford). Refined final alignments of reference proteins (n = 98) were obtained after removal of the extra sequences from alignments. Resulting alignments were concatenated into alignments of all four analyzed proteins. To avoid duplication of information, N-terminal part (~70 amino acid-long fragment) identical between LTag and STag was removed in

alignment for STag before concatenation. A filtered version of the alignment for phylogenetic analysis was obtained by removing weak alignment positions from each of the four alignments with GBLOCKS [52–53] using mildest filtering parameters on GBLOCKS server [54].

Phylogenetic analysis

Phylogenetic analysis was carried out using maximum likelihood (ML) algorithm. For each alignment, ProtTest [55] was used to select the best-fit evolutionary model to be used in phylogenetic analysis (excluding estimate of proportion of invariable sites (I) because of comparatively small test set). Le and Gascuel [56] amino acid substitution matrix (LG) with GAMMA (G) model of rate heterogeneity and optimization of substitution rates (F) was found as best model (LG+G+F) for all four proteins and, thus, this model was used in final analysis with concatenated alignments. For every analysis we produced 500 trees and 1000 bootstrap replicates using RAxML [57]. Convergence test as implemented in RAxML was used to confirm that obtained number of bootstrap replicates [58, 59] was sufficient (in all cases 50–150 replicates were sufficient). Phylogenetic trees were visualized using DENDROSCOPE [59]. A majority rule consensus tree of all 500 trees was made to evaluate their differences. The leave-one-out approach for assessing site-specific congruence/incongruence with the underlying tree based on the evolutionary placement algorithm [60] was used as implemented in RAxML. Several methods were used to evaluate the confidence of branch placements within the tree. In addition to the standard bootstrapping method, calculated Shimodaira–Hasegawa (SH) [61], Internode Certainty (IC) and Tree Certainty (TC) values [62, 63] were calculated. Uncertain taxa within the constructed phylogenetic tree were also identified alternatively by use of RogueNaRok [64].

Comparison of entire PyV genomes

To compare whole genome sequences, a method that does not require sequence alignments to get a final estimate of similarity was developed. The implemented sequence comparison methodology is very simple and some similarity might be found to procedures used by Fitch [65, 66]. Analysis was done by gathering statistics of best pairs of sequence spans of the length "w" taken from two genomes under comparison. Thus, all w-mers taken from one genome were compared with w-mers from a second genome by means of a sliding-window procedure and counted only pairs of w-mers which had higher count of identical nucleotides in their alignment than some predefined critical value (in this work w = 50 and critical value n = 30). The critical value "n" was determined by modeling statistics of comparison results for randomized genome sequences. The value was set to mean low level of random background (no higher than 10–15% of meaningless hits in comparison to lowest similarities between PyV genomes).

We suggest that the probability p of total count of pairs of w-mers of at least S to appear by chance during genome comparison could be calculated using the extreme value distribution [67–69]:

$$p(S \geq x) = 1 - \exp(-e^{-\lambda(x-u)}) \tag{1}$$

The equation has two parameters: the characteristic value u can be thought of as the mean of x, and λ is a scale parameter. Best fit values for u and λ for gathered statistics of randomized genomes were 18 and 0.1 respectively (for case of w = 50 and n = 30 used in this study).

Cluster analysis of genome sequences

To cluster polyomavirus genomes, precalculated probabilities p (Eq 1) for all 1029x1029 polyomavirus genome pairs were used as input for CLANS [70]. The genomes were clustered using

different levels of p to highlight different evolutionary relationships. Initially dynamic clustering was run using 3D space, but final representations of clusters were transformed into 2D space.

Results and Discussion

Identification and genome analysis of new PyVs

Application of the nested broad-spectrum PCR technique [32], resulted in the detection of approximately 250 bp-long PCR products for six out of 172 bank voles and nine out of 85 common voles (S4 Table). After DNA sequence determination, a BLASTn search in GenBank revealed two novel PyV-related nucleotide sequences with the highest similarity of 75% and 74% to the LTag-encoding region of WUPyV and MCPyV respectively. The sequence similarity of both 250 bp DNA fragments to other members of the tentative genus *Wukipolyomavirus* was not significant. The similarity of the bank vole-derived DNA sequence to other rodent PyVs was not significant, but the DNA sequence amplified from common voles was 69% identical to HaPyV LTag sequence. The entire PyV genomes from five bank voles and two common voles were amplified, cloned and sequenced (S4 Table).

Three of five BVPyV genomes amplified from CCF samples of bank voles (KS/14/281, KS/14/289, KS/14/328) captured at the same location (Crailsheim, Baden-Wuerttemberg; see S1 Fig, site #3) were 5032 bp long and differed by one to five single nucleotide polymorphisms (SNP) from rodent KS/14/336-derived, 5031 bp-long BVPyV genome (S2 Table). The fifth genome amplified from bank vole sample KS/13/999 captured at another location (Gotha, Thuringia, see S1 Fig, site #2) one year earlier was 5011-bp long and 98% identical to the genome of BVPyV-KS/14/281. The 81 SNPs in BVPyV KS/13/999 genome compared to BVPyV KS/14/281 strain were mainly located in intergenic regions, VP1 gene and C-terminal end of LTag encoding gene. The remaining part of early region encoding LTag and STag antigens (2300 bp genome fragment) contained no mutations at all. A variability of viral strains from the same host was demonstrated for most of human and non-human PyVs [71–74]. However, sequenced CPyV genomes from common voles collected in distant regions of Germany demonstrated much less divergence than genomes of BVPyV detected in bank voles. Thus, the CVPyV genome sequence generated from CCF sample of common vole KS/13/947 captured in Gotha, Thuringia (see S2 Fig, site #9) differed from CVPyV genome generated from kidney sample of rodent KS/13/980 (Jeeser, Mecklenburg-Western Pomerania, see S2 Fig, site #3) by one SNP only. Both CVPyV genome sequences were 5024 bp long (S2 Table).

All seven cloned and sequenced genomes had the typical PyV genome organization, including putative ORFs for early regulatory proteins (LTag and STag) on one strand and for structural late proteins (VP1, VP2 and VP3) on the other strand (Fig 1A and 1B). Both regions were separated by the NCCR. LTag and STag encoding sequences were identified based on homology to known PyV proteins, on genomic location and sequences resembling classical eukaryotic splice donor and acceptor consensus signals. Two splicing sites were detected for LTag early transcript (S2 Table). Such analysis showed that LTag mRNA contains three exons encoding 651 aa protein for four BVPyV strains (or 650 amino acids, aa, for KS/13/999) and 642 aa for both CVPyV strains. Further, the early genome regions of CVPyV and BVPyV do not encode MTag or ALTO proteins, which is in contrast to HaPyV or MPyV [7]. However, further experiments with mRNA isolated from virus infected cells are needed to fully investigate the presence of LTag splicing variants because LTag in other rodent PyVs are encoded by only two exons. An additional putative ORF (tentatively named X-ORF) was found in the NCCR region of BVPyV and CVPyV (Fig 1A and 1B, magenta arrow). ORF sizes and locations were not conserved between BVPyV strains KS/14/281 (162 bp) and KS/13/999 (189 bp) or
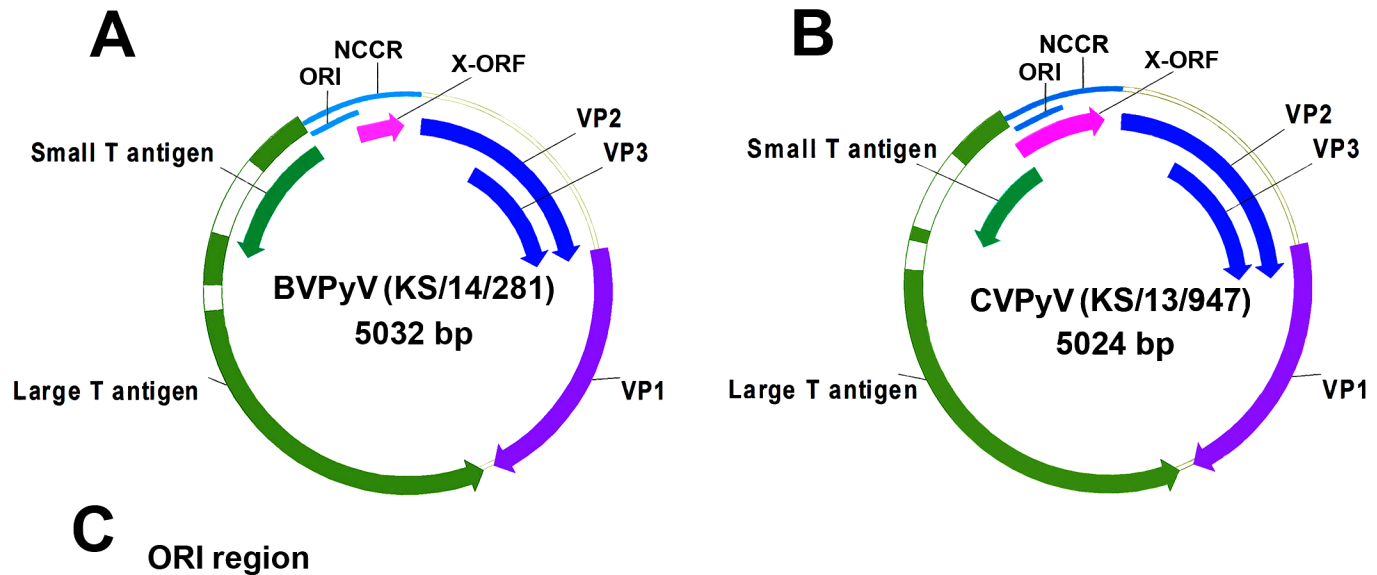
Fig 1. Schematic presentation of the genome organization of BVPyV (A) and CVPyV (B) and alignment of ORI region sequences of BVPyV strains (KS/14/281 and KS/13/999) and CVPyV strain KS/13/947 (C). The positions of LTag introns are shown uncolored. Putative LTag binding sites (GAGGC pentanucleotides) are shown in red; AT-rich region/TATA box is shown in green; palindromic repeats shown in magenta; early palindrome (EP) region is underlined; LTag ATG codon is shown in blue (according [38], [39]).

CVPyV (426 bp) genomes (Fig 1 and S2 Table). The putative proteins encoded by these X-ORFs in BVPyV strains KS/14/281 (53 aa) and KS/13/999 (62 aa) or both CVPyV strains (141 aa) had different length, were less than 35% identical in their amino acid sequences and had no sequence similarity to any protein from protein data bases, so they are unlikely to be synthesized and/or of functional relevance.

The genome alignments to the nucleotide sequences in GenBank showed that the closest known relative was WUPyV which in alignable regions on a whole genome basis shared 79% identity with BVPyV and 67% identity with CVPyV. BVPyV and CVPyV DNA sequence identity over the whole genome (alignable region did not include part of NCCR) was 83%, so according to the criteria of the *Polyomaviridae* study group of ICTV [3] these two viruses may represent two new separate PyV species.

The difference of the genome length of BVPyV strains KS/13/999 (5011 bp) and KS/14/281 (5032 bp) are due to indels within the NCCR sequence and intergenic region between LTag and VP1 genes, except deletion of one codon encoding N649 in LTag sequence and one nucleotide in the first intron of LTag. The insertion of four nucleotides in noncoding region between LTag and VP1 genes in BVPyV (KS/13/999) genome was also determined. This is in line with previous studies indicating that the highest level of variation is common for noncoding regions of PyVs [73, 75].

NCCR sequences of PyVs are suggested to be important for the host adaptation of certain PyVs [75]. The core region of both novel vole-derived PyVs demonstrated a high level of sequence conservation. Five putative LTag binding sites (GAGGC), palindromic repeats and early palindrome (EP) region [38, 39] were detected in NCCR ORI region of both viruses (Fig 1C). Interestingly, a 12 bp deletion in the NCCR region of BVPyV variant KS/13/999 did not change the arrangement of main ORI elements found in KS/14/281 and other three BVPyV
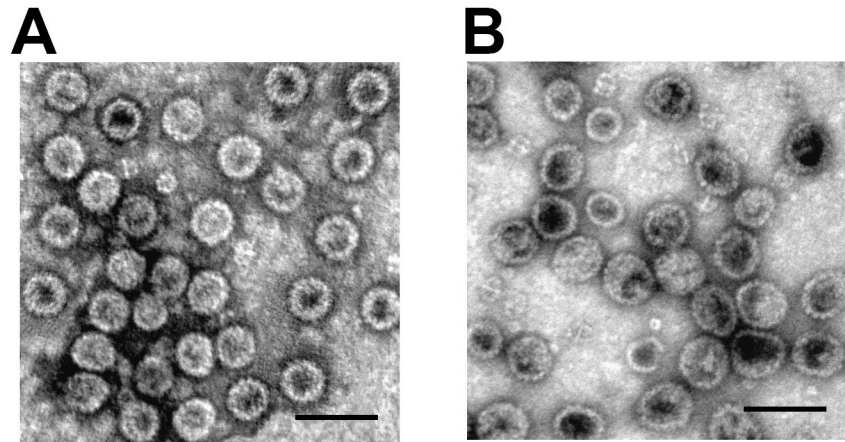
A

B

Fig 2. Detection of BVPyV (A) and CVPyV (B) VP1 protein VLP formation by electron microscopy.
Bars, 100 nm.

doi:10.1371/journal.pone.0140916.g002

variants. This deletion had changed only the distance between ORI and LTAg ATG codon, placing it closer to ATG (Fig 1C). Analogously, ORI region sequences in CVPyV and BVPyV genomes were very similar with the same arrangement and sequences of ORI elements but CVPyV ORI was positioned closer to LTag ATG codon alike to BVPyV variant KS/13/999 (Fig 1C). In contrast, the second part of NCCR sequences, where a variety of cellular transcription factors bind and regulate the NCCR, were more diverse and only 65% identical between BVPyV and CVPyV or 95% between BVPyV variants KS/14/281 and KS/13/999.

Generation of BVPyV- and CVPyV-derived VP1 VLP

Full-length comparison of the VP1 gene sequences revealed that KS/13/999 strain sequence differed from other BVPyV VP1 sequences by 31 SNPs, but only one SNP was non-synonymous (A327T). For this reason one gene was amplified from strain KS/14/281 and used for the expression of BVPyV-derived VP1 in yeast. CVPyV-derived VP1 gene was amplified from strain KS/13/947.

Recombinant VP1 proteins of BVPyV and CVPyV were efficiently generated in yeast. The ability of the VP1proteins to self-assemble into VLPs was tested after purification by electron microscopy (EM) analysis. It was demonstrated that VP1 proteins derived from both new viruses were assembling into PyV-like VLPs (Fig 2A and 2B). The size of BVPyV- derived VLPs was 45–50 nm and that of CVPyV-derived VLPs 50–55 nm in diameter. The VP1 VLPs were used for serological screening of vole CCF samples by indirect ELISA. A parallel analysis of WUPyV-VP1-derived VLPs should allow conclusions on the potential cross-reactivity of antibodies with VLPs of this PyV and the two novel vole PyVs.

Prevalence and host specificity of vole-associated PyVs

Bank vole and common vole kidney- and CCF-derived and RCA-enriched DNA samples were examined using specific diagnostic primers for both viruses (S1 Table). The novel assay detected BVPyV in 23 CCF and 41 kidney samples of bank voles (S4 Table). Samples of CCF and kidney tissues were both positive for BVPyV only in two animals. All other samples were PCR-positive only in one of two examined tissues. BVPyV VP1-VLP-based indirect ELISA detected 17 CCF samples with anti-BVPyV VP1 antibodies, with only 6 of them being

identified by PCR as positive. In most cases, BVPyV VP1 VLP-specific antibodies were also recognized by WUPyV VP1 VLPs (16 from 18), but less frequently by CVPyV VP1 VLPs (7 from 18). Taken together, 73 of 172 (42.4%) bank voles showed signs of previous or acute BVPyV infection (S4 Table). The positive animals originated from nine out of ten trapping sites from the three investigated federal states Baden-Wuerttemberg, Thuringia and Mecklen-burg-Western Pomerania, indicating a broad geographical distribution of the virus. Only at one site in Baden-Wuerttemberg (Steinheim) none of the two bank vole samples was positive (S4 Table).

Diagnostic CVPyV-specific PCR identified 11 of 85 animals to be CVPyV positive in one of the tissues used for PCR (S4 Table). ELISA analysis of available CCF samples of common voles showed that only five of 72 rodents had antibodies against CVPyV VP1 VLPs (S4 Table). These antibodies efficiently recognized WUPyV VP1 (in all five animals) and less efficiently BVPyV VP1 VLPs (two of five animals). Antibodies were mainly found in PCR-negative animals. Taken together, 15 of 85 (17.6%) common voles indicated signs of a previous or ongoing CVPyV infection. Positive voles originated from seven states of Germany (S4 Table). Infected animals were found in all eight federal states investigated.

Our results demonstrated that about 42% of bank voles and 18% of common voles were infected with new PyVs. Both viruses were found in CCF and kidney samples. Interestingly, despite a 90% amino acid sequence similarity of BVPyV and CVPyV VP1 proteins antibodies from rodents positive with VLPs of these new viruses were cross-reacting with VP1 VLPs of each other less efficiently, than with WUPyV VP1 which is only 48% identical to VP1 proteins of both new PyVs (S4 Table). The exclusive detection of viral DNA or of anti-VP1 antibodies in bank and common voles might be in contrast to the usually observed establishment of a chronic PyV infection in their hosts [1]. PyV persistence might be established in the kidney with virus release or alternatively PyVs might stay in a latent state [1]. In line, antibodies against virus proteins, especially VP1, are detected soon after infection and remain for a long time, while virus DNA might be not detected anymore [1, 76]. Future studies would have to prove a latent state of the infection with the novel PyVs resulting in a failure of virus DNA detection in anti-VP1-positive animals.

Phylogenetic analysis

LTag, STag, VP1 and VP2 proteins known to be common for all PyVs were selected for phylogenetic analysis. As the quality of phylogenetic analysis depends highly on the quality of provided sequence alignments, in this study four separate alignments for each LTag, STag, VP1 and VP2 proteins as well as a concatenated one of 98 representative PyVs were produced using three different methods. A maximum likelihood tree with highest bootstrap support values and lowest uncertainty of branch placement was obtained for MAFFT alignment with "—ep 0.123" when the alignment was not reduced by use of GBLOCKS (Fig 3).

Most uncertainly placed taxa pointed out by RogueNaRok server (reviewing taxa support values below 80%) were HPyV12 (0.910), African elephant PyV (AePyV; 0.099), bovine PyV (BoPyV; 0.091) and dolphin PyV (DoPyV; 0.035). Lower uncertainty values were shown for three PyVs, bat isolate BtRp (0.013), Tadarida brasiliensis 1 (0.004) and JCV135-40 (0.002).

The three tentative genera (*Orthopolyomavirus*, *Wukipolyomavirus*, and *Avipolyomavirus*) [3] can be easily recognized as distinct clades in our tree. In line with a previous study [12], the position of *Avipolyomavirus* group separates the *Orthopolyomavirus* group into two subgroups: *Orthopolyomavirus I* and *Orthopolyomavirus II*. Additionally besides the four above mentioned groups, Feltkamp et al. [12] recommended the distinction of one more lineage–*Malawipolyo-mavirus* (occupied by MWPyV, STLPyV, MXPyV, and HPyV10) which is also grouped
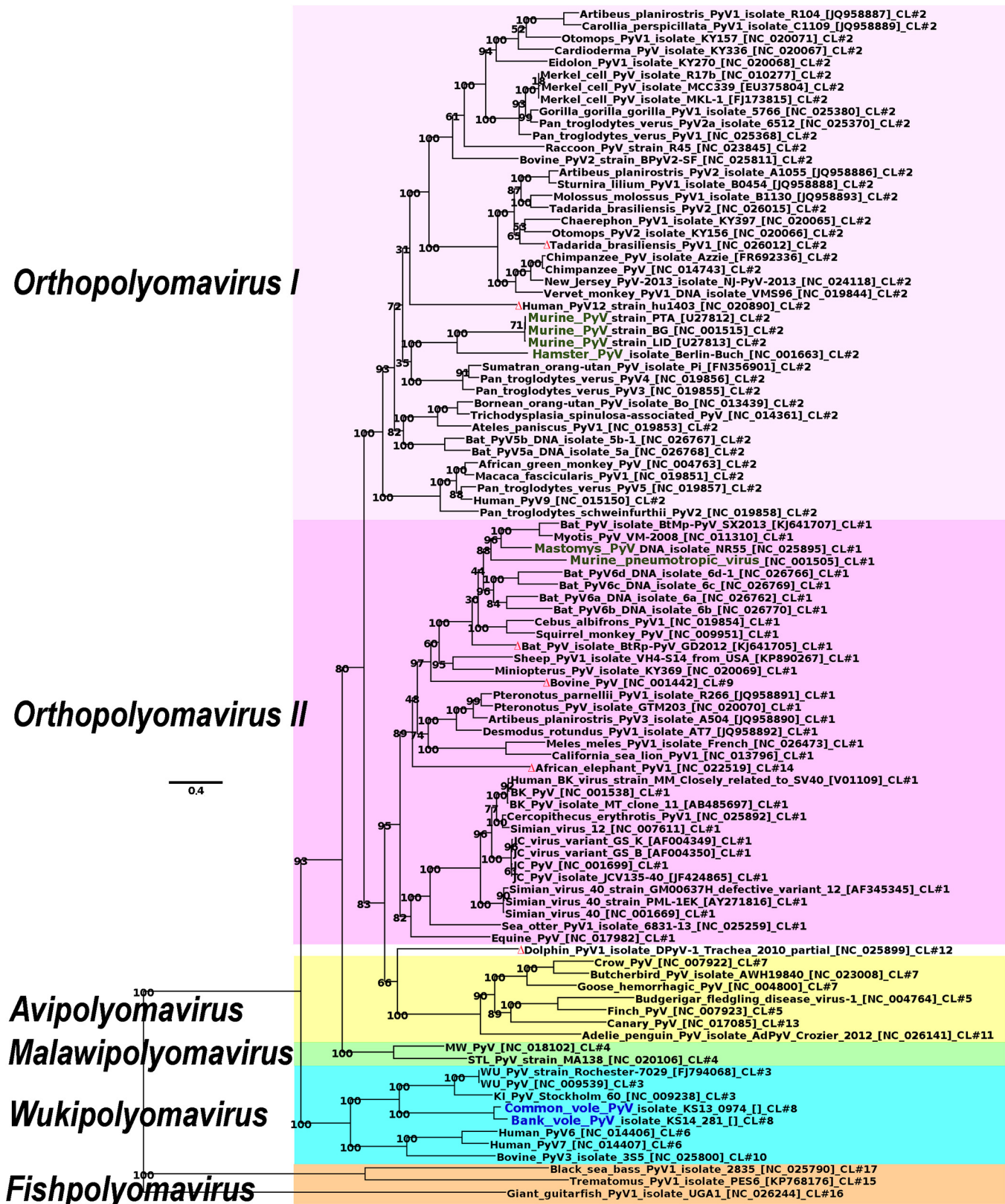
Fig 3. Phylogenetic tree of 98 representatives of *Polyomaviridae* family obtained using ML methods for concatenated LTag, STag, VP1 and VP2 protein sequence alignment. The tree is rooted at the most distant PyVs, namely fish polyomaviruses. Numbers at the nodes indicate statistical support

(bootstrap values) for branch placements. Red triangles mark taxons with least confidence in their placement, as determined using RogueNaRok. The five lineages: *Orthopolyomavirus I*, *Orthopolyomavirus II*, *Avipolyomavirus*, *Malawipolyomavirus* and *Wukipolyomavirus* are highlighted in different colors. Rodent polyomaviruses Murine PyV (MPyV), Hamster PyV (HaPyV), Murine pneumotropic virus (MPtV) and Mastomys PyV (MasPyV) are labeled in green. The novel viruses BVPyV and CVPyV are labeled in blue.

together in our phylogenetic analysis ([Fig 3](#)). Recently identified three fish PyVs (black sea bass [77], giant guitarfish (RefSeq NC_026244; unpublished) and *Trematomus* (RefSeq NC_026944; unpublished) form a sixth group.

New bank and common vole PyVs are confidently incorporated into the *Wukipolyomavirus* group of PyVs (bootstrap support values of 100) forming a vole/WUPyV/KIPyV sister clade to the HPyV6/HPyV7/BPyV3 clade. Interestingly, in discussing of possible evolutionary pathways of WU/KI PyVs, Gaynor and colleagues proposed that these PyVs might have originated from rodent (murine) PyVs [78]. Their proposal was based on the lack of a C-terminal "host range" domain in LTag and agnoprotein. Thus, our findings of new vole PyVs within *Wukipolyomavirus* group support their proposal. The other rodent PyVs were found at different positions within the tree showing stronger similarities for MPyV and HaPyV on one hand and MPtV and MasPyV on the other ([Fig 3](#)).

Whole genome comparisons

To evaluate the relatedness of PyVs, an alternative algorithm was developed in this study. Novel (global) measure for similarity of genome sequences was introduced by counting pairs of short fragments (of length w) that share higher sequence identity than some predefined value (n identical nucleotides per w) when aligned. This simple measure empowered comparison of any potentially meaningful w-mer pairs, including genomic stretches coding conservative functional domains of proteins, specific regulatory regions of genome, microRNAs or any other important signals within PyVs DNA sequence, without the use of any additional information. Sequence dynamic clustering software CLANS, with the calculated pairwise similarity scores (p) of PyV genomes as input, was used to investigate relatedness between all analyzed genomes ([Fig 4](#)).

At $p = 1 \times 10^{-69}$ all orthopolyomaviruses formed a homogeneous cluster, while other groups of PyVs were spread into smaller clusters or had orphans not showing relatedness at this level of similarity ([Fig 4D](#)). Thus, wukipolyomaviruses were separated into two clusters: one cluster consisted of WUPyVs, KIPyVs and new PyVs from bank and common voles and other was composed of HPyV6, HPyV7 and BoPyV3 ([Fig 4D](#)). The *Avipolyomavirus* group was separated into four subgroups while fish PyVs did not indicate any connections even at $p = 1 \times 10^{-56}$ similarity level ([Fig 4C](#)). Similarity level $p = 1 \times 10^{-86}$ would fit sufficiently with scheme of classification of PyVs where *Orthopolyomavirus* group is split into at least three lineages suggested by others: *Orthopolyomavirus I*, *Orthopolyomavirus II* and *Malawipolyomavirus* [12]. Low similarity of the three PyVs AePyV, DoPyV and BoPyV1, to other orthopolyomaviruses could indicate loss of some group specific features in genome sequence. It is interesting to note that during the phylogenetic analysis these three PyVs were also pointed out by RogueNaRok as having uncertain placement.

Inside *Wukipolyomavirus* cluster at $p = 1 \times 10^{-86}$ genome similarity level, human PyVs and non-human PyVs separate into distinct clusters (BVPyV/CVPyVs–from WU/KI PyVs separates at $p = 1 \times 10^{-83}$, Bovine PyV3 from HPyV6/HPyV7 separates at $p = 1 \times 10^{-73}$). Despite this, it should be mentioned that most close representatives of HPyVs 6 and 7 are WU/KI PyVs (at $p = 1 \times 10^{-45}$ multiple connections) and so these HPyVs could be assigned into one cluster. A more detailed picture of the relatedness levels of PyVs is given in a colored matrix form in [S5 Table](#).
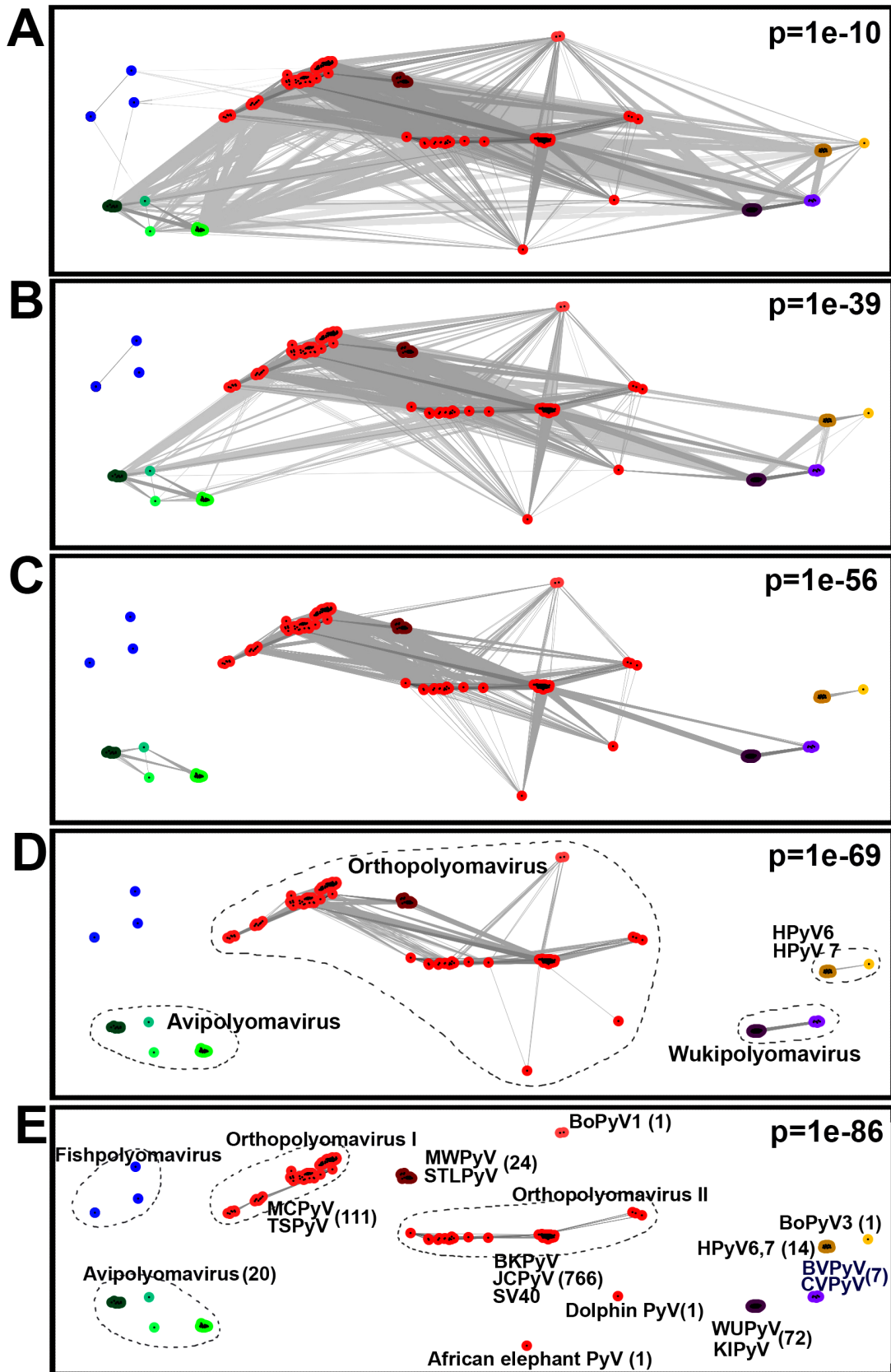
Fig 4. Relationships between PyV genomes, as shown at various cutoff similarity values. (A) cutoff p = 1x10^{-10}: all PyVs are nested indicating a common nature of all PyVs (not excluding fishpolyomaviruses); **(B)** cutoff p = 1x10^{-39}: all PyVs are nested, except fishpolyomaviruses; **(C)** cutoff p = 1x10^{-56}: Avipolyomaviruses are separated; **(D)** cutoff p = 1x10^{-69}: Wukipolyomaviruses are separated; **(E)** cutoff p = 1x10^{-86}: Orthopolyomaviruses are divided into 3 main lineages—*Orthopolyomavirus I*, *Orthopolyomavirus II* and *Malawipolyomavirus*. To note, 1x10^{-86} level indicates closest relationships, while 1x10^{-10}—most distant. Grey lines connect pairs of similar PyV genomes at selected cutoff level of p.

doi:10.1371/journal.pone.0140916.g004

Additional investigation of whole genome comparisons were also performed by analyzing PyV genomes for the DNA sequence features conserved between all PyVs and highlighting sequence features that are specific for CVPyV and BVPyV (Fig 5). The VP2 gene sequence carries little similarity among all PyVs (including WUPyV) except for a short region encoding the N-terminal end where a myristoylation target sequence is located. However, this gene of both vole PyVs is highly similar (88% identical). There could be several explanations: a recent branching event in the evolutionary history and/or presence of conserved specific VP2 function of vole PyVs. VP1 gene contains a few conserved sequence features (overall identity of BVPyV and CVPyV VP1 gene sequences is 82%). Notably the highest similarity between the VP1-coding sequences of all PyVs is seen in a region in the middle of the gene that encodes part of the core β-barrel structure. While both of the late region genes show little similarity among all PyVs, the LTag gene sequence is much more conserved. Conserved sequence features can be linked with known domain locations. All three main domains: ATPase, DnaJ domains and origin binding domain (OBD) retain similarity between all PyVs and the new ones as well (Fig 5) [79]. The largest conserved 350 bp length region encode Walker A and B boxes as well as the Surface Loop motif of the AAA+ ATPase domain [79]. Thus, this additional analysis on whole genome comparison algorithm fully confirms the current knowledge of the PyV genome regions and conservation of highly similar domains and motifs and strengthens the results of the CLANS analyses.

Conclusions

The growing number of newly discovered PyV uncovered the polyomavirus diversity which is reflected in phylogenetic tree as well as in CLANS analysis. The true genomic variety of PyVs is still unknown as well as the true host range or evolutionary rates. The numerous distinct PyV species found in humans or non-human primates seems in contrast with only four PyVs species known in rodents (MPyV, HaPyV, MPtV, and MasPyV). Rodents represent the largest order of mammalian species (contain 40% of all mammal species) and are a reservoir of more than 60 known zoonotic human pathogens [80, 81]. All four until now known rodent PyV genomes consistently grouped in *Orthopolyomavirus* I and II genera. Phylogenetic analysis and whole genome comparison analyses were used to classify the new bank vole and common vole PyVs as members of the tentative *Wukipolymavirus* genus. The finding of these novel PyVs suggest that more PyVs belonging to other clades might be present in other rodents and awaits their future discovery. The identification of the vole PyVs might help better understand the true genetic diversity and evolutionary history of the wukipolyomaviruses. The sequence divergence and phylogenetic positions of HaPyV and the vole-associated novel viruses may indicate a murine origin of HaPyV, instead a cricetide origin. In addition, the currently known diversity of PyVs cannot be explained by a long-term co-evolution of PyVs and mammalian hosts, but rather a split of an ancestral PyV into different lineages that thereafter might follow a co-evolution mechanism. Further, the finding of these vole PyVs may suggest that the wukipolyomavirus ancestor might have existed in a rodent-like animal. In addition, the vole wukipolyomaviruses may allow the development of novel rodent models for detailed studies on transmission and pathogenicity of the corresponding human PyVs.
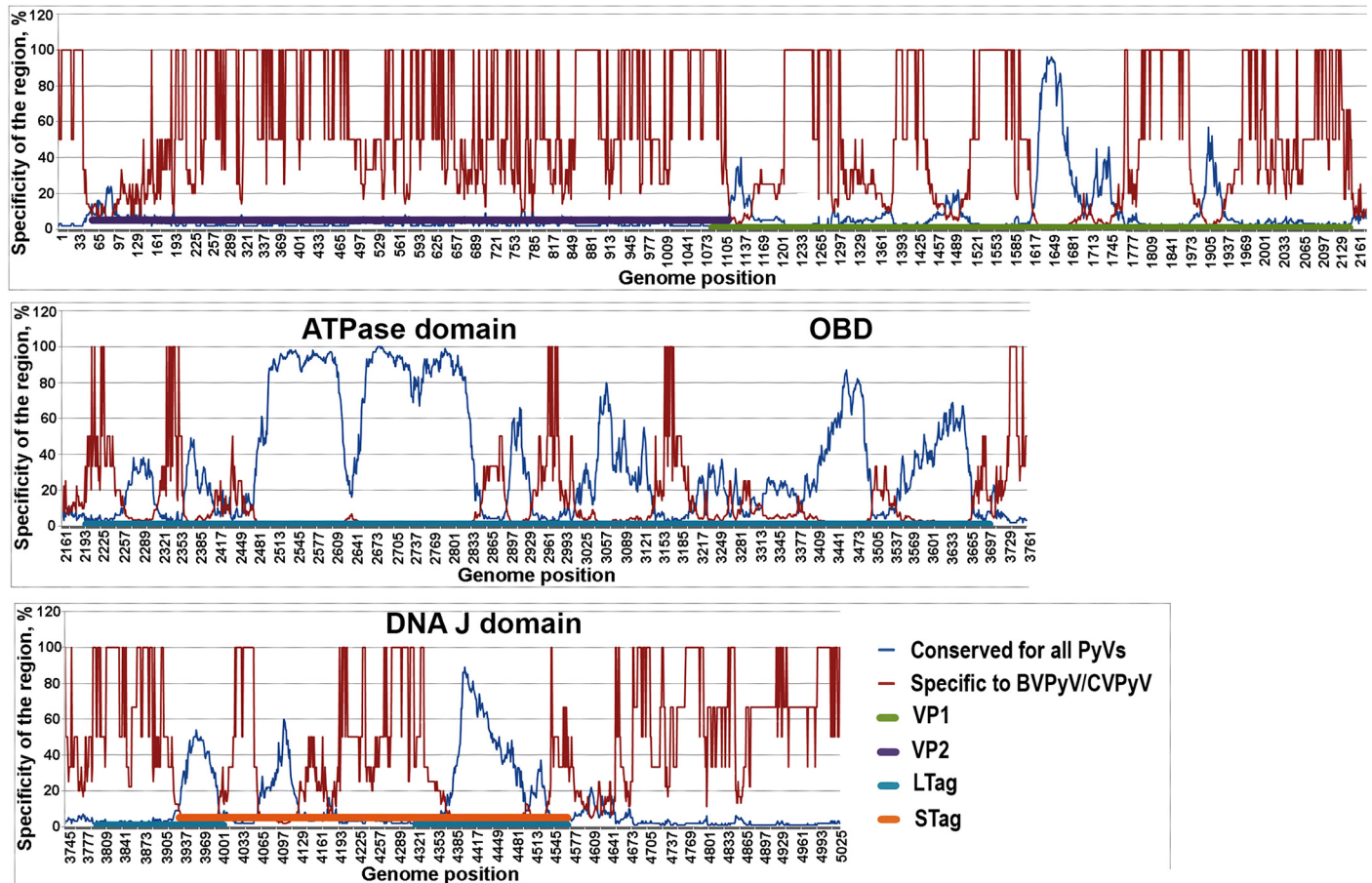
Fig 5. Sliding window analysis of genome sequences specific to CVPyV and BVPyV compared to other PyVs (%). The short fragments of a length w = 50 nt of BVPyV and CVPyV genomes were compared with w-mers of all PyV genomes by means of a sliding-window procedure. Only pairs of w-mers which had higher count of identical nucleotides in their alignment than some predefined critical value n = 30 when aligned were determined as similar. VP1, VP2, VP3, LTag, and STag genes are labeled with different colors. The nucleotide positions in genome are indicated at the bottom.

doi:10.1371/journal.pone.0140916.g005

Supporting Information

S1 Fig. Trapping sites of bank voles (*Myodes glareolus*) in Germany. The map was generated using the ArcGis program package. Source: Geobasis-DE/BKG/GeoNutzV.
(TIF)

S2 Fig. Trapping sites of common voles (*Microtus arvalis*) in Germany. The map was generated using the ArcGis program package. Source: Geobasis-DE/BKG/GeoNutzV.
(TIF)

S1 Table. Primers used to detect, amplify and characterize BVPyV and CVPyV genomes.
(PDF)

S2 Table. Vole trapping site information and genome characterization of BVPyV and CVPyV.
(PDF)

S3 Table. PyV genome and protein sequence datasets used in this study.
(XLSX)

S4 Table. Results of PCR and ELISA analysis of bank vole and common vole CCF and kidney samples.
(XLSX)

S5 Table. Colored matrix form of levels of PyVs relatedness generated by counting pairs of short fragments of a length w = 50 nt that share higher sequence identity than predefined value n = 30 nt when aligned.
(XLSX)

Acknowledgments

Author Contributions

Conceived and designed the experiments: JN AG. Performed the experiments: JN AT AG. Analyzed the data: JN AT JS RGU AG. Contributed reagents/materials/analysis tools: JS RGU. Wrote the paper: AT RGU AG.

References

1. Imperiale MJ, Major EO. Polyomaviruses. In: Knipe DM, Howley PM, Griffen DE, Lamb RA, Martin RA, Roizman B, Straus SE, editors. Fields virology, 4th edn. Lippincott Williams and Wilkins, Philadelphia; 2007. pp. 2263–2298.

2. Meinke G, Bullock PA, Bohm A. Crystal structure of the simian virus 40 large T-antigen origin-binding domain. J Virol. 2006; 80(9):4304–12 PMID: 16611889

3. Johne R, Buck CB, Allander T, Atwood WJ, Garcea RL, Imperiale MJ, et al. Taxonomical developments in the family Polyomaviridae. Arch Virol. 2011; 156:1627–1634. doi: 10.1007/s00705-011-1008-x PMID: 21562881

4. Ahuja D, Saenz-Robles MT, Pipas JM. SV40 large T antigen targets multiple cellular pathways to elicit cellular transformation. Oncogene. 2005; 24:7729–7745. PMID: 16299533

5. Cheng J, DeCaprio JA, Fluck MM, Schaffhausen BS. Cellular transformation by Simian Virus 40 and Murine Polyoma Virus T antigens. Semin Cancer Biol. 2009; 19(4):218–228. doi: 10.1016/j.semcancer.2009.03.002 PMID: 19505649

6. Johne R, Müller H. Polyomaviruses of birds: etiologic agents of inflammatory diseases in a tumor virus family. J Virol. 2007; 81:11554–11559. PMID: 17715213

7. Carter JJ, Daugherty MD, Qi X, Bheda-Malge A, Wipf GC, Robinson K,et al. Identification of an over-printing gene in Merkel cell polyomavirus provides evolutionary insight into the birth of viral genes. Proc Natl Acad Sci USA. 2013; 110:12744–12749. doi: 10.1073/pnas.1303526110 PMID: 23847207

8. Khalili K, White MK, Sawa H, Nagashima K, Safak M. The agnoprotein of polyomaviruses: A multifunctional auxiliary protein. J. Cell Physiol. 2005; 204:1–7. PMID: 15573377

9. Shen PS, Enderlein D, Nelson CD, Carter WS, Kawano M, Xing L, et al. The structure of avian polyomavirus reveals variably sized capsids, nonconserved inter-capsomere interactions, and a possible location of the minor capsid protein VP4. Virology. 2011; 411:142–152. doi: 10.1016/j.virol.2010.12.005 PMID: 21239031

10. Gerits N, Moens U. Agnoprotein of mammalian polyomaviruses. Virology. 2012: 432:316–326. doi: 10.1016/j.virol.2012.05.024 PMID: 22726243

11. Giorda KM, Raghava S, Hebert DN. The Simian virus 40 late viral protein VP4 disrupts the nuclear envelope for viral release. J Virol. 2012; 86:3180–3192. doi: 10.1128/JVI.07047-11 PMID: 22238309

12. Feltkamp MC, Kazem S, van der Meijden E, Lauber C, Gorbalenya AE. From Stockholm to Malawi: recent developments in studying human polyomaviruses. J Gen Virol. 2013; 94(Pt 3):482–96. doi: 10.1099/vir.0.048462–0 PMID: 23255626

13. Ehlers B, Moens U. Genome analysis of non-human primate polyomaviruses. Infect Genet Evol. 2014; 26:283–94. doi: 10.1016/j.meegid.2014.05.030 PMID: 24933462

14. Tao Y, Shi M, Conrardy C, Kuzmin IV, Recuenco S, Agwanda B, et al. Discovery of diverse polyomaviruses in bats and the evolutionary history of the *Polyomaviridae*. J Gen Virol. 2013; 94(Pt 4):738–48. doi: 10.1099/vir.0.047928–0 PMID: 23239573

15. Kobayashi S, Sasaki M, Nakao R, Setiyono A, Handharyani E, Orba Y, et al. Detection of novel polyomaviruses in fruit bats in Indonesia. Arch Virol. 2015; 160(4):1075–1082. doi: 10.1007/s00705-015-2349-7 PMID: 25670407

16. Orba Y, Kobayashi S, Nakamura I, Ishii A, Hang'ombe BM, Mweene AS, et al. Detection and characterization of a novel polyomavirus in wild rodents. J Gen Virol. 2011; 92(Pt 4):789–95. doi: 10.1099/vir.0.027854–0 PMID: 21177925

17. Johne R, Wittig W, Fernandez-de-Luco D, Hofle U, Muller H. Characterization of two novel polyomaviruses of birds by using multiply primed rolling-circle amplification of their genomes. J Virol. 2006; 80: 3523–3531. PMID: 16537620

18. Gross L. A filterable agent, recovered from Ak leukemic extracts, causing salivary gland carcinomas in C3H mice. Proc Soc Exp Biol Med. 1953; 83(2):414–21. PMID: 13064287

19. Stewart SE, Eddy BE, Borgese NG. Neoplasms in mice inoculated with a tumor agent carried in tissue culture. J Natl Cancer Inst. 1958; 20:1223–1243. PMID: 13549981

20. Scherneck S, Ulrich R, Feunteun J. The hamster polyomavirus–a brief review of recent knowledge. Virus Genes. 2001; 22: 93–101. PMID: 11210944

21. Mayer M, Dorries K. Nucleotide sequence and genome organization of the murine polyomavirus, Kilham strain. Virology. 1991; 181(2):469–80. PMID: 1849675

22. Cornulier T, Yoccoz NG, Bretagnolle V, Brommer JE, Butet A, Ecke F, et al. Europe-wide dampening of population cycles in keystone herbivores. Science. 2013; 340 (6128):63–66. doi: 10.1126/science.1228992 PMID: 23559246

23. Schlegel M, Jacob J, Krüger DH, Rang A, Ulrich RG. Hantavirus emergence in rodents, insectivores and bats: What comes next? In: Johnson N, editor. The Role of Animals in Emerging Viral Diseases. Elsevier, Academic Press, Amsterdam, Boston, Heidelberg, London, New York, Oxford, Paris, San Diego, San Francisco, Singapore, Sydney, Tokyo; 2014. pp. 235–292.

24. Kinnunen PM, Henttonen H, Hoffmann B, Kallio ER, Korthase C, Laakkonen J, et al. Orthopox virus infections in Eurasian wild rodents. Vector borne Zoonotic Dis. 2011; 11(8):1133–1140. doi: 10.1089/vbz.2010.0170 PMID: 21453121

25. Mayer-Scholl A, Hammer J, Schmidt S, Ulrich R, Pfeffer M, Woll D, et al. *Leptospira spp*. In Rodents and Shrews in Germany. Int J Environ Res Public Health. 2014; 11(8): 7562–7574. doi: 10.3390/ijerph110807562 PMID: 25062275

26. Jacob J, Ulrich RG, Freise J, Schmolz E. Monitoring von gesundheitsgefährdenden Nagetieren: Projekte, Ziele und Ergebnisse. Bundesgesundheitsbl.—Gesundheitsforsch.–Gesundheitsschutz. 2014; 57: 511–518.

27. Ulrich RG, Schmidt-Chanasit J, Schlegel M, Jacob J, Pelz HJ, Mertens et al. Network "Rodent-borne pathogens" in Germany: longitudinal studies on the geographical distribution and prevalence of hantavirus infections. Parasitol Res. 2008; 103;Suppl 1:S121–9. doi: 10.1007/s00436-008-1054-9 PMID: 19030894

28. Ulrich RG., Schmidt S, Rosenfeld UM, Groschup MH, Schlegel M. Hantavirus-Diagnostik in Reservoir-wirten. LabLoeffler. 2011; 4:13–17.

29. Schlegel M, Ali HS, Stieger N, Groschup MH, Wolf R, Ulrich RG. Molecular identification of small mammal species using novel cytochrome B gene-derived degenerated primers. Biochem Genet. 2012; 50: 440–447. doi: 10.1007/s10528-011-9487-8 PMID: 22193288

30. Rector A, Tachezy R, Van Ranst M. A sequence-independent strategy for detection and cloning of circular DNA virus genomes by using multiply primed rolling-circle amplification. J Virol. 2004; 78:4993–4998. doi: 10.1128/JVI.78.10.4993–4998.2004 PMID: 15113879

31. Johne R, Müller H, Rector A, van Ranst M, Stevens H. Rollingcircle amplification of viral DNA genomes using phi29 polymerase. Trends Microbiol. 2009; 17:205–211. doi: 10.1016/j.tim.2009.02.004 PMID: 19375325

32. Johne R, Enderlein D, Nieper H, Müller H. Novel polyomavirus detected in the feces of a chimpanzee by nested broad-spectrum PCR. J virol. 2005; 79(6): 3883–3887. PMID: 15731285

33. Sambrook J, Russell DW. Molecular cloning, A Laboratory Manual. Cold Spring Harbor: Cold Spring Harbor Press; 2001.

34. Sasnauskas K, Buzaite O, Vogel F, Jandrig B, Razanskas R, Staniulis J, et al. Yeast cells allow high-level expression and formation of polyomavirus-like particles. Biol Chem. 1999; 380(3):381–386. PMID: 10223341

35. Gish W, States DJ. Identification of protein coding regions by database similarity search. Nat Genet. 1993; 3(3):266–272. PMID: 8485583

36. Desmet FO, Hamroun D, Lalande M, Collod-Béroud G, Claustres M, Béroud C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. Nucleic Acids Res. 2009; 37(9):e67. doi: 10.1093/nar/gkp215 PMID: 19339519

37. Peng YC, Acheson NH. Polyomavirus large T antigen binds cooperatively to its multiple binding sites in the viral origin of DNA replication. J Virol. 1998; 72: 7330–7340. PMID: 9696829

38. Kwun HJ, Guastafierro A, Shuda M, Meinke G, Bohm A, Moore PS, et al. The Minimum Replication Origin of Merkel Cell Polyomavirus Has a Unique Large T-Antigen Loading Architecture and Requires Small T-Antigen Expression for Optimal Replication. J Virol. 2009; 83(23):12118–12128. doi: 10.1128/JVI.01336-09 PMID: 19759150

39. An P, Sáenz Robles MT, Pipas JM, Large T. Antigens of Polyomaviruses: Amazing Molecular Machines. Annu Rev Microbiol. 2012; 66:213–236.

40. Pleckaityte M, Zvirbliene A, Sezaite I, Gedvilaite A. Production in yeast of pseudotype virus-like particles harboring functionally active antibody fragments neutralizing the cytolytic activity of vaginolysin. Microb Cell Fact. 2011; 10.

41. Norkiene M Stonyte J, Ziogiene D, Mazeike E, Sasnauskas K, Gedvilaite A. Production of recombinant VP1-derived virus-like particles from novel human polyomaviruses in yeast. BMC Biotechnol. 2015; 15:68. doi: 10.1186/s12896-015-0187-z PMID: 26239840

42. Slibinskas R, Samuel D, Gedvilaite A, Staniulis J. Sasnauskas K. Synthesis of the measles virus nucleoprotein in yeast *Pichia pastoris* and *Saccharomyces cerevisiae*. J Biotechnol. 2004; 107:115–124. PMID: 14711495

43. Internet site: http://www.ncbi.nlm.nih.gov/genomes/GenomesGroup.cgi?taxid=151341.

44. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics. 2006; 22:1658–1659. PMID: 16731699

45. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 2002; 30:3059–3066. PMID: 12136088

46. Clamp M, Cuff J, Searle SM, Barton GJ. The jalview java alignment editor. Bioinformatics. 2004; 20:426–427. PMID: 14960472

47. Waterhouse A, Procter J, Martin D, Clamp M, Barton G. Jalview version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics. 2009; 25:1189–1191. doi: 10.1093/bioinformatics/btp033 PMID: 19151095

48. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSIBLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997; 25:3389–3402. PMID: 9254694

49. Internet site: http://mafft.cbrc.jp/alignment/software/epvalue.html

50. Pei J, Kim BH, Tang M, Grishin NV. PROMALS web server for accurate multiple protein sequence alignments. Nucleic Acids Res. 2007; 35:649–652.

51. Pei J, Grishin NV. PROMALS3D: multiple protein sequence alignment enhanced with evolutionary and three-dimensional structural information. Met Mol Biol. 2014; 1079:263–271.

52. Talavera G, Castresana J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst Biol. 2007; 56:564–577. PMID: 17654362

53. Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol. 2000; 17:540–552. PMID: 10742046

54. Internet site: http://molevol.cmima.csic.es/castresana/Gblocks_server.html

55. Darriba D, Taboada GL, Doallo R, Posada D. ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics. 2011; 27:164–1165.

56. Le SQ, Gascuel O. An improved general amino acid replacement matrix. Mol Biol Evol. 2008; 25:1307–1320. doi: 10.1093/molbev/msn067 PMID: 18367465

57. Stamatakis A. RAxML-VI-HPC: maximum likelihoodbased phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics. 2006; 22:2688–2690. PMID: 16928733

58. Pattengale ND, Alipour M, Bininda-Emonds ORP, Moret BME, Stamatakis A. How many bootstrap replicates are necessary? J Comput Biol. 2010; 17:337–354. doi: 10.1089/cmb.2009.0179 PMID: 20377449

59. Huson DH, Richter DC, Rausch C, Dezulian T, Franz M. Rupp R. Dendroscope: an interactive viewer for large phylogenetic trees. BMC Bioinformatics. 2007; 8:460. PMID: 18034891

60. Berger SA, Krompass D, Stamatakis A. Performance, accuracy and web server evolutionary placement of short sequence reads under maximum likelihood. Syst Biol. 2011; 60:291–302. doi: 10.1093/sysbio/syr010 PMID: 21436105

61. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol. 2010; 59:307–321. doi: 10.1093/sysbio/syq010 PMID: 20525638

62. Salichos L, Rokas A. Inferring ancient divergences requires genes with strong phylogenetic signals. Nature. 2013; 497:327–331. doi: 10.1038/nature12130 PMID: 23657258

63. Salichos L, Stamatakis A, Rokas A. Novel information theory-based measures for quantifying incongruence among phylogenetic trees. Mol Biol Evol. 2014; 31:1261–1271. doi: 10.1093/molbev/msu061 PMID: 24509691

64. Aberer AJ, Krompass D, Stamatakis A. Pruning Rogue Taxa Improves Phylogenetic Accuracy: An Efficient Algorithm and Webservice, Syst Biol. 2015; in press; server: http://rnr.h-its.org

65. Fitch WM. Locating Gaps in Amino Acid Sequences to Optimize the Homology Between Two Proteins. Biochem Genet. 1969; 3:99–108. PMID: 5364927

66. Fitch WM. An improved method of testing for evolutionary homology. J Mol Biol. 1966; 16:9–16. PMID: 5917736

67. Gumbel E.J. Statistics of extremes. Columbia Univ. Press, New York; 1958.

68. Altschul SF, Gish W. Local alignment statistics. Meth Enzymol. 1996; 266:460–480. doi: 10.1016/S0076-6879(96)66029-7 PMID: 8743700

69. Dembo A, Karlin S. Zeitouni O. Limit distribution of maximal non-aligned two-sequence segmental score. Ann Prob. 1994; 22:2022–2039.

70. Frickey T, Lupas A. CLANS: a Java application for visualizing protein families based on pairwise similarity. Bioinformatics. 2004; 20:3702–3704. PMID: 15284097

71. Moens U, Van Ghelue M. Polymorphism in the genome of non-passaged human polyomavirus BK: implication for cell tropism and the pathological role of the virus. Virology 2005; 331:209–231. doi: 10.1016/j.virol.2004.10.021 PMID: 15629766

72. Van Ghelue M, Khan MT, Ehlers B, Moens U. Genome analysis of the new human polyomaviruses. Rev Med Virol. 2012; 22:354–377. doi: 10.1002/rmv.1711 PMID: 22461085

73. Munoz LJ, Ludena D, Gedvilaite A, Zvirbliene A, Jandrig B, Voronkova T, et al. Lymphoma outbreak in a GASH:Sal hamster colony. Arch Virol. 2013; 158(11):2255–65. doi: 10.1007/s00705-013-1737-0 PMID: 23719671

74. Hill S, Murphy A, Cotten M, Palser A, Benson P, Lesellier S, et al. Discovery of a polyomavirus in European badgers (Meles meles) and the evolution of host range in the Polyomaviridae. J Gen Virol. 2015; 96(Pt 6):1411–22. doi: 10.1099/vir.0.000071 PMID: 25626684

75. White MK, Safak M, Khalili K. Regulation of gene expression in primate polyomaviruses. J Virol. 2009; 83(21):10846–10856. doi: 10.1128/JVI.00542-09 PMID: 19640999

76. Moens U, Van Ghelue M, Song X, Ehlers B. Serological cross-reactivity between human polyomaviruses. Rev Med Virol. 2013; 23:250–64. doi: 10.1002/rmv.1747 PMID: 23650080

77. Peretti A, FitzGerald PC, Bliskovsky V, Pastrana DV, Buck CB. Genome Sequence of a Fish-Associated Polyomavirus, Black Sea Bass (*Centropristis striata*) Polyomavirus. Genome Announc. 2015; 3 (1):1–2. pii: e01476-14. doi: 10.1128/genomeA.01476-14

78. Gaynor AM, Nissen M D, Whiley DM, Mackay IM, Lambert SB, Wu G, et al. Identification of a novel polyomavirus from patients with acute respiratory tract infections. PLoS pathog. 2007; 3(5), e64. PMID: 17480120

79. An P, Brodsky JL, Pipas JM. The conserved core enzymatic activities and the distinct dynamics of polyomavirus large T antigens. Arch Biochem Biophys. 2015; 573:23–31. doi: 10.1016/j.abb.2015.02.033 PMID: 25752954

80. Huchon D, Madsen O, Sibbald MJ, Ament K, Stanhope MJ, Catzeflis F, et al. Rodent phylogeny and a timescale for the evolution of Glires: evidence from an extensive taxon sampling using three nuclear genes. Mol Biol Evol. 2002; 19:1053–1065. PMID: 12082125

81. Meerburg BG, Singleton GR, Kijlstra A. Rodent-borne diseases and their risks for public health. Crit Rev Microbiol. 2009; 35: 221–270. doi: 10.1080/10408410902989837 PMID: 19548807