Detection and genome characterization of two novel papillomaviruses and a novel polyomavirus in tree shrew (Tupaia belangeri chinensis) in China

Ping Liu 1†, Ye Qiu 1†, Cheng Xing 1, Ji-Hua Zhou 2, Wei-Hong Yang 2, Qiong Wang 1, Jin-Yan Li 1, Xi Han 2, Yun-Zhi Zhang 3* and Xing-Yi Ge 1*

Abstract

Background: Papillomaviruses (PVs) and polyomaviruses (PyVs) infect diverse vertebrates including human and cause a broad spectrum of outcomes from asymptomatic infection to severe disease. There has been no PV and only one PyV detected in tree shrews, though the genomic properties of tree shrews are highly similar to those of the primates.

Methods: Swab and organ samples of tree shrews collected in the Yunnan Province of China, were tested by viral metagenomic analysis and random PCR to detect the presence of PVs and PyVs. By PCR amplification using specific primers, cloning, sequencing and assembling, genomes of two PVs and one PyV were identified in the samples.

Results: Two novel PVs and a novel PyV, named tree shrew papillomavirus 1 and 2 (TbelPV1 and TbelPV2) and polyomavirus 1 (TbelPyV1) were characterized in the Chinese tree shrew (Tupaia belangeri chinensis). The genomes of TbelPV1, TbelPV2, and TbelPyV1 are 7410 bp, 7526 bp, and 4982 bp in size, respectively. The TbelPV1 genome contains 7 putative open-reading frames (ORFs) coding for viral proteins E1, E2, E4, E6, E7, L1, and L2; the TbelPV2 genome contains 6 ORFs coding for viral proteins E1, E2, E6, E7, L1, and L2; and the TbelPyV1 genome codes for the typical small and large T antigens of PyV, as well as the VP1, VP2, and VP3 capsid proteins. Genomic comparison and phylogenetic analysis indicated that TbelPV1 and TbelPV2 represented 2 novel PV genera of Papillomaviridae, and TbelPyV1 represented a new species of genus Alphapolyomavirus. Our epidemiologic study indicated that TbelPV1 and TbelPV2 were both detected in oral swabs, while TbelPyV1 was detected in oral swabs and spleens.

Conclusion: Two novel PVs (TbelPV1 and TbelPV2) and a novel PyV (TbelPyV) were discovered in tree shrews and their genomes were characterized. TbelPV1, TbelPV2, and TbelPyV1 have the highest similarity to Human papillomavirus type 63, Ursus maritimus papillomavirus 1, and Human polyomavirus 9, respectively. TbelPV1 and TbelPV2 only showed oral tropism, while TbelPyV1 showed oral and spleen tropism.

Keywords: Tree shrew, Tupaia belangeri, Papillomavirus, Polyomavirus, Tumor virus
Background

Papillomaviruses (PVs) belong to the *Papillomaviridae* family, which are non-enveloped small viruses with about 60 nm in diameter. PVs have a single molecule of circular genomic dsDNA of 5.8 to 8.6 kilobases (kb) in size. PVs infect epithelium of skin and mucosa in a variety of vertebrates, including mammals, birds, and reptiles. Recently, the *Firstpapillomavirinae* and *Secondpapillomavirinae* subfamilies in the *Papillomaviridae* family have been established. PVs in the *Firstpapillomavirinae* are classified into 53 genera, from *Alpha*- to *Zetapapillomavirus* [1–3]. In history, PV is the first tumor virus reported and was described in 1933 as an agent causing cutaneous horn-like warts in eastern cottontail rabbits [4]. PV infection leads to different outcomes, from lack of symptoms, warts, skin lesions, to carcinomas [5, 6]. Recently, 226 types of human papillomavirus (HPV) have been recorded (http://www.nordic-health.se/hpvcenter, accessed on 2019-01-10). HPV types 16 and 18 are the main cause of cervical cancer and are considered highly carcinogenic [7].

Polyomaviruses (PyVs) belong to the *Polyomaviridae* family, which are non-enveloped small viruses with about 50 nm in diameter and have a single circular genomic dsDNA of about 5 kb in size. More than 80 species of PyVs have been nominated by the International Committee on Taxonomy of Viruses (ICTV). PyVs have been classified into 4 genera including *Alpha*-, *Beta*-, *Gamma*-, and *Deltapolyomavirus*, among which, the alpha-, beta-, and deltapolyomaviruses infect birds [8]. The first rodent PyV was discovered in 1953 in the house mouse and could induce epithelial tumors at multiple sites [9]. Recently, PyV infection in fish has been reported [10]. Most PyVs rarely cause significant clinical diseases in their host. However, some PyVs can cause serious diseases particularly in immunocompromised individuals, such as Merkel cell PyV (MCPyV, *Human polyomavirus 5*) causing Merkel cell cancer, BK PyV (BKPyV, *Human polyomavirus 1*) causing nephropathy and haemorrhagic cystitis, and JC PyV (JCPyV, *Human polyomavirus 2*) causing progressive multifocal leukoencephalopathy [11–16].

Tree shrews are small mammals belonging to the genus *Tupaia*, family Tupaiidae, order Scandentia. Recent studies elucidated that the genome and some basic biological properties of the tree shrew are much closer to those of primates than those of rodents, which makes the tree shrew a valuable model animal in biomedical research [17–19]. Tree shrew has been successfully used to create animal models for the studies of hepatitis B virus, hepatitis C virus, herpes simplex virus type 1 infection, and myopia, breast cancer, depression, and others [20–22]. The Chinese tree shrew (*Tupaia belangeri chinensis*) is widely distributed in Southeast Asia, South and Southwest China. However, viruses naturally carried by *Tupaia belangeri chinensis* have been barely described. In the present study, we performed a viral metagenomic and degenerated primer-based PCR study of *Tupaia belangeri chinensis*. The epidemics of PV and PyV in *Tupaia belangeri chinensis* were revealed, and the genomes of 2 novel PVs and a novel PyV were characterized and analyzed. This is the first report of detection of PVs in tree shrew at the complete-genome level, and our result indicates that tree shrews may carry diverse PVs and PyVs. The novel viruses described here are a first step for studying the origin and evolution of PVs and PyVs in tree shrews.

Methods

Sample collection

From 2016 to 2017, 71 tree shrews, including 56 live ones and 15 dead ones, were captured in bushes and grass near the cropland ridge of Jianchuan County and Lufeng County of the Yunnan Province of China, during routine surveillance of pathogen control. Oral and rectal swabs were sampled from live tree shrews, and then the animals were released to nature, while dead individuals were dissected for tissue collection. Swabs were stored in cryopreservation tubes containing 1 mL of virus transport medium (VTM), and organs were kept in the same tubes, but without medium. VTM was composed of Hank’s balanced salt solution, pH 7.4, containing BSA (1%), amphotericin (15 μg mL⁻¹), penicillin (100 U ml⁻¹) and streptomycin (50 μg ml⁻¹) and filtered with the 0.22 μm filter. VTM was stored as 1 mL aliquots at −80 °C. All samples were immediately put in liquid nitrogen for short storage, then transported to the laboratory on dry ice and stored at −80 °C. Animal species was confirmed by sequencing the mitochondrial cytochrome b (*CytB*) gene as described previously [23].

Viral metagenomic sequencing

To perform the viral metagenomic analysis, two libraries were constructed using pooled swabs and tissues, respectively. For swabs, totally 112 swabs from 56 living tree shrews, including 56 oral swabs and 56 rectal swabs, were pooled. For tissues, 75 tissue samples from 15 dead tree shrews were pooled. To prepare swab samples, each sample was thawed on ice, homogenized by vortexing for 30 s, and centrifuged at 13,000×g for 15 min at 4 °C. 200 μl of each supernatant were pooled together, and then filtered through a 0.45 μm filter. To prepare tissue samples, 100 mg tissue of each sample were homogenized using 1 ml phosphate buffer saline, centrifuged, pooled, and filtered. Viral particles in filtrate were enriched by ultracentrifugation, and treated with DNase and RNase to digest unprotected nucleic acid. Then viral
DNA/RNA was extracted using QIAamp Viral RNA Mini Kit (Qiagen) and subjected to random PCR (rPCR) as previously described [24]. The purified rPCR products were used for library construction, and then were sequenced using the HiSeq-PE150 instrument (Illumina platform). The generated paired-end reads were debar-coded, trimmed, and de novo assembled using Geneious software package (Version10.2) [25]. The assembled contigs were aligned to viral protein database using BLASTx with an E-value less than 10−5.

Viral detection and genomic sequencing

The contigs related to PV and PyV with significant E-values were used to design specific primers for screening, amplifying, and sequencing the complete genome of novel viruses in the samples. Primer sets of PV1F/R, PV2F/R, and PyVF/R were used to detect three different viruses in 112 swabs and 75 tissues, respectively. For detection PV and PyV in the single sample, the PCR was conducted in a 25 μl reaction mix containing 2.5 μl PCR reaction buffer, 5 pmol of each primer, 50 mM MgCl2, 0.5 mM dNTP, 0.1 μl Platinum Taq Enzyme (Invitrogen) and 2 μl DNA. The amplification was performed as follows: 94 °C for 4 min followed by 40 cycles consisting of 94 °C for 30 s, 50 °C for 30 s, 72 °C for 30 s, and a final extension of 72 °C for 10 min. After PCR screening and sequencing, samples showing singly positive for TbelPV1, or TbelPV2, or TbelPyV1 were chosen to amplify the complete genomes. Primer sets of PV1F1/R1, PV1F2/R2, and PV1F3/R3 were used for genomic characterization of TbelPV1; PV2F1/R1, PV2F2/R2, PV2F3/R3 for TbelPV2; and PyVF1/R1, PyVF2/R2 for TbelPyV1 (Additional file 1: Table S1). The PCR replicons were cloned into T-Vector, and at least 3 positive clones of each fragment were sequenced by Sanger method. The obtained sequences were used to assemble the three full-length viral genomes.

Sequence analysis

The ORF Finder (National Center for Biotechnology Information) was used to predict the putative open reading frames (ORFs) and their deduced amino acid (aa) sequences. Gene sequences and encoded protein aa sequences were compared to those of known viruses with complete genomes available in GenBank. ClustalW and Geneious were used to generate and edit sequence alignment [25, 26]. The alignment data sets were used to generate the phylogenetic trees by MEGA6 under the maximum likelihood (ML) method with bootstrap values of 1000 replicates [27].

Results
PV and PyV in Chinese tree shrew

Two libraries, swab library and tissue library were sequenced by Illumina platform. By BLASTx searching, 15 contigs encoding proteins which are significant related to homologous proteins of PVs and PyVs were found. In detail, 10 contigs related to PV were only detected in swab library. By alignment of these contig sequences, 2 contigs with different nucleotide (nt) sequences both encode the E1 protein, indicating that 2 different PV strains were in the swab samples. These 2 PV strains were named TbelPV1 and TbelPV2.

Two long contigs almost covering the whole genome of PyV were detected in the swab library, and 3 contigs related to PyV were detected in the tissue library. These sequences of PyV contigs in the swab library and the tissue library were identical which indicates that they originate from the same PyV strain (named as TbelPyV1).

To confirm the prevalence of PV and PyV in tree shrew, the specific primer sets (Additional file 1: Table S1) designed according to the detected contig sequences of TbelPV1, TbelPV2, and TbelPyV1 were used to screen all the samples. The result showed that TbelPV1, TbelPV2, and TbelPyV1 were detected in oral swabs with 37.5, 25, and 30.4% positive rates, respectively (Table 1). Oral swabs co-infected PV/PyV negative. In tissue samples, only TbelPyV1 was detected (in 7 of 15 spleens; 1

| Samples     | PyV1a | PV1b | PV2c | PyV1/PV1 | PyV1/PV2 | PyV1/PV2 | PyV1/PV1/PV2 |
|-------------|-------|------|------|----------|----------|----------|--------------|
| Oral swab   | 21/56 (37.5%) | 14/56 (25%) | 17/56 (30.4%) | 7/56 (12.5%) | 9/56 (16.1%) | 4/56 (7.1%) | 2/56 (3.6%) |
| Rectal swab | 0/56  | 0/56 | 0/56 | –        | –        | –        | –            |
| Heart       | 0/15  | 0/15 | 0/15 | –        | –        | –        | –            |
| Liver       | 0/15  | 0/15 | 0/15 | –        | –        | –        | –            |
| Spleen      | 7/15 (46.7%) | 0/15 | 0/15 | –        | –        | –        | –            |
| Lung        | 0/15  | 0/15 | 0/15 | –        | –        | –        | –            |
| kidney      | 0/15  | 0/15 | 0/15 | –        | –        | –        | –            |

TbelPyV1; TbelPV1; TbelPV2
| Papillomavirus | GenBank number | Genome size(nt) | G+C content (%) | pairwise identities of nucleotide and amino acid sequence(nt/aa%) |
|---------------|---------------|----------------|----------------|---------------------------------------------------------------|
|               |               |                |                | Tupaia belangeri chinensis papillomavirus 1 | Tupaia belangeri chinensis papillomavirus 2 |
|               |               |                |                | L1  L2  E1  E2  E4  E6  E7 | L1  L2  E1  E2  E4  E6  E7 |
| TbelPV1       | MK443496      | 7410           | 48.3           | /    /    /    /    /    /    / | /    /    /    /    /    /    / |
| TbelPV2       | MK443497      | 7568           | 54.3           | 58/  49/  66/  45/  /    /    /    / | 54/  35/  45/  31/  27/  36/ |

### Alphapapillomavirus

| Papillomavirus | GenBank number | Genome size(nt) | G+C content (%) | pairwise identities of nucleotide and amino acid sequence(nt/aa%) |
|---------------|---------------|----------------|----------------|---------------------------------------------------------------|
| Human papillomavirus 2 | NC_001352      | 7860           | 48.4           | 55/  32/  46/  39/  32/  37/  39/  35/  31/  47/  35/  35/  | 39/    |
| Human papillomavirus 18 | NC_001357      | 7857           | 40.4           | 52/  35/  46/  41/  31/  36/  40/  35/  37/  42/  | 32/  29/  40/  |
| Rhesus monkey papillomavirus | NC_001678     | 8028           | 48.0           | 54/  33/  49/  40/  /    29/  37/  54/  34/  49/  | 33/  29/  40/  |

### Betapapillomavirus

| Papillomavirus | GenBank number | Genome size(nt) | G+C content (%) | pairwise identities of nucleotide and amino acid sequence(nt/aa%) |
|---------------|---------------|----------------|----------------|---------------------------------------------------------------|
| Human papillomavirus type 5 | NC_001531     | 7746           | 42.4           | 53/  32/  43/  38/  19/  34/  38/  35/  35/  44/  33/  26/  38/  36/  |
| Canine papillomavirus 3 | NC_008297     | 7801           | 51.5           | 55/  35/  51/  38/  /    35/  40/  57/  36/  52/  38/  37/  42/  |
| Canine papillomavirus 4 | NC_010226     | 7742           | 53.3           | 54/  35/  50/  38/  /    33/  41/  56/  35/  50/  37/  39/  43/  |

### Deltapapillomavirus

| Papillomavirus | GenBank number | Genome size(nt) | G+C content (%) | pairwise identities of nucleotide and amino acid sequence(nt/aa%) |
|---------------|---------------|----------------|----------------|---------------------------------------------------------------|
| Bos grunniens papillomavirus type 1 | JX174437       | 7946           | 44.4           | 54/  30/  47/  39/  34/  28/  27/  52/  28/  46/  32/  | 32/  31/  31/  27/  26/  24/  19/  |
| Ursus maritimus papillomavirus 1 | NC_010739     | 7582           | 48.4           | 56/  31/  51/  39/  39/  /    59/  32/  50/  37/  39/  /    –  |
| Feline papillomavirus type 2 | EU796884      | 7899           | 53.0           | 55/  35/  51/  34/  38/  40/  57/  36/  52/  36/  38/  44/  |

### Dydeltapapillomavirus

| Papillomavirus | GenBank number | Genome size(nt) | G+C content (%) | pairwise identities of nucleotide and amino acid sequence(nt/aa%) |
|---------------|---------------|----------------|----------------|---------------------------------------------------------------|
| Morelia spilota papillomavirus 1 | HQ262535       | 7048           | 40.9           | 54/  31/  51/  44/  34/  29/  33/  53/  30/  48/  36/  | 32/  34/  |
| Equine papillomavirus 2 | NC_012123     | 7802           | 56.1           | 56/  32/  39/  38/  /    25/  33/  54/  31/  40/  36/  26/  35/  |

### Gammapapillomavirus

| Papillomavirus | GenBank number | Genome size(nt) | G+C content (%) | pairwise identities of nucleotide and amino acid sequence(nt/aa%) |
|---------------|---------------|----------------|----------------|---------------------------------------------------------------|
| Human papillomavirus type 4 | NC_001457     | 7353           | 38.5           | 53/  33/  51/  43/  41/  37/  37/  52/  34/  | 48/  36/  36/  36/  |
| Mastomys natalensis papillomavirus | NC_001605     | 7687           | 50.2           | 55/  36/  50/  32/  /    33/  37/  58/  33/  50/  34/  39/  33/  |
| Cottontail rabbit papillomavirus | NC_001541     | 7868           | 46.4           | 58/  32/  49/  44/  44/  23/  39/  54/  32/  | 49/  35/  21/  17/  |
| Rabbit oral papillomavirus | NC_002232     | 7565           | 44.4           | 59/  34/  49/  43/  38/  22/  43/  54/  36/  48/  | 34/  24/  42/  42/  |

### Iotapapillomavirus

| Papillomavirus | GenBank number | Genome size(nt) | G+C content (%) | pairwise identities of nucleotide and amino acid sequence(nt/aa%) |
|---------------|---------------|----------------|----------------|---------------------------------------------------------------|
| Crocuta crocutapapillomavirus 1 | NC_018575     | 8344           | 38.4           | 55/  36/  52/  45/  34/  29/  40/  52/  32/  48/  | 35/  34/  38/  |
| Procyon lotor papillomavirus 1 | NC_007150     | 8170           | 45.9           | 59/  38/  52/  45/  39/  32/  38/  55/  36/  52/  36/  33/  38/  28/  25/  38/  |
Table 1). These data suggest that TbelPV1 and TbelPV2 exhibit oral tropism, and TbelPyV1 has oral and spleen tropism.

Characterization of TbelPV1, TbelPV2, and TbelPyV1 genomes

To characterize the viral genomes, 3 samples singly positive for TbelPV1, or TbelPV2, or TbelPyV1 were chosen to amplify the genomic fragments (Additional file 1: Table S1). The replicons were cloned, sequenced, and assembled to obtain the complete genomes. Genomes of TbelPV1, TbelPV2, and TbelPyV1 were deposited in GenBank under the accession numbers: MK443496, MK443497, and MK443498.

Analysis of genomes of TbelPV1 and TbelPV2

The circular genomic DNA of TbelPV1 and TbelPV2 display a length of 7410 bp and 7526 bp, with the GC content of 48.3 and 54.3%, respectively (Table 2). By BLASTx and BLASTp search, homologues of E1, E2, E4, E6, E7, L1, and L2 were found in TbelPV1 genome, and E1, E2, E6, E7, L1, and L2 were found in TbelPV2 genome. As shown in Fig. 1, the coding sequences of TbelPV1 and TbelPV2 are clockwise arranged on the same strand of the circular double stranded DNA, which is similar to other PVs. The accurate locations and lengths of the predicted ORFs, as well as the molecular weights of the corresponding proteins, are listed in Table 3.

The E1 is the most conserved protein among PVs. The E1 protein, an ATP-dependent helicase, binds specifically to the origin of replication and is required for both the initiation and elongation of viral DNA synthesis [28]. The conserved ATP-binding sites of the ATP-dependent helicase were detected at the carboxyl terminus of E1 proteins in TbelPV1 and TbelPV2 with GNSNTGKS and GQPNTGKS sequences, respectively. E2, an auxiliary factor, is not essential for origin-dependent DNA replication, but E2 greatly enhances the ability of E1 to initiate DNA replication [28]. The leucine zipper domain in the E2 protein is involved in DNA-binding and dimerization of the E2 protein. The consensus sequence is L-X6-L-X6-L-X6-L in most E2 proteins of PVs. In the E2 proteins of TbelPV1 and TbelPV2, the variant sequences L-X7-L-X8-L-X6-L and L-X8-L-X6-L-X7-L were found, respectively. An E4 ORF harbored in the E2 ORF was detected in the genome of TbelPV1. The TbelPV1 E4 has a high percentage of cytosine di-, tri- and
tetranucleotides, resulting in a high-proline region with content of 50%.

The E6 and E7 genes are involved in inducing transformation of infected cells. This feature permitted the mapping of the E6 and E7 as oncogenes for the high-risk PV types [7]. The zinc binding domains with conserved 36 aa sequence of C-X-X-C-X29-C-X-X-C was characterized as the zinc binding domain and the E6 and E7 proteins bind zinc through these cysteine residues [29]. For both TbelPV1 and TbelPV2, two of these domains were found in E6 protein and one in E7. The E7 proteins of both TbelPV1 and TbelPV2 contain a zinc binding domain containing the L-X-C-X-E motif which is sufficient for binding of pocket protein like the retinoblastoma tumor suppressor protein [29].

A long control region (LCR) is located between the L1 ORF and the E6 ORF. The length of TbelPV1 and TbelPV2 LCRs are 545 bp and 463 bp, respectively. In the LCR of TbelPV1, 4 E2 binding sites (E2BS) were found with the consensus sequence AAC-N6-GTT. As for the LCR of TbelPV2, 4 E2BS were detected with the sequence ACC-N6-GGT.

**Gene similarities and phylogenetics of TbelPV1 and TbelPV2**

The sequence similarities between TbelPV1, TbelPV2, and known PVs are shown in Table 2. The L1 is the most conserved gene among all the PVs. The identity of nt sequence in the L1 ORF between TbelPV1 and TbelPV2 was 58%. For the L1 ORF, TbelPV1 showed the highest similarity to *Human papillomavirus type 63* (HPV-63) with a maximum of 60% nt sequence identity, followed by *Rabbit oral papillomavirus* (RoPV) and *Procyon lotor papillomavirus 1* (PIPv1) both with 59% nt identities. For the L1 ORF, TbelPV2 showed the highest similarity to *Ursus maritimus papillomavirus 1* (UmPV1) with a maximum of 60% nt sequence identity, followed by *Mastomys natalensis papillomavirus* (MnPV) with 58% nt identity.

A phylogenetic tree was constructed based on the L1 sequences of TbelPV1, TbelPV2, together with other representative animal and human PVs in the established genera. In the tree, TbelPV1 was most closely related to members of the genera *Kappapapillomavirus* and *Mupapillomavirus*, but did nest within *Kappapapillomavirus* or *Mupapillomavirus*, suggesting an early evolution. Likewise, TbelPV2 formed a separate evolutionary branch and showed a long evolutionary distance to the members of genus *lotapapillomavirus* (Fig. 2).

**Analysis of TbelPyV1 genome**

The circular genome of the TbelPyV1 was 4982 bp with overall GC content of 42.6% (Table 4). By genomic comparing and ORF searching, 5 genes and 2 noncoding regions homologous to earlier described PyVs were detected in the TbelPyV1 genome (Fig. 1 and Table 5). Two earlier genes, the small T antigen (ST) and the large T antigen (LT) were arranged on one strand, while 3 later genes including the capsid VP1, VP2, and VP3 were arranged on the opposite strand. An upstream regulatory region (URR) of 463 bp was located between the beginning of the early genes and the beginning of the late genes. A short noncoding region of 66 bp with an AT-rich region was located between LT and VP1 (Fig. 1 and Table 5). More details about the sizes and positions of predicted genes, as well as the deduced protein sequences, were listed in Table 5.

The URR contains specific DNA sequences meditating the origin of PyV replication [29]. In the URR of TbelPyV1 genome, 2 copies of the perfectly consensus pentanucleotide LT binding site (GAGGC) and their reverse complement (GCCTC) were present. Between the pentanucleotide and reverse complement, a central perfect octanucleotide palindrome (TCCCTTCT and AGAAGGGA) was found. The possible sequence of core origin region of replication is GCCCTCGAGCTCCTCC TTCCTTTAGTCAGAGGAGGAGCGCAGAGG.

Previously studies demonstrated that some typical elements in T antigens are necessary to fulfill infectious cycle of PyVs [30]. The predicted ST protein of TbelPyV1 is 189 aa in length and contains two CXCX2C consensus sequences (CFCITC, CFCYSC) for protein
phosphatase 2A binding. The ST and LT sharing an N-terminus region of approximately 80 residues is a common feature for all PyVs. In the N-terminus of ST and LT of TbelPyV1, the highly canonical residues HPDK was found, which binds and activates the ATPase activity of host cell HSC70 [31]. In the LT of TbelPyV1, the conserved residues LXCXE (LLCSE) was detected that binds direct to the retinoblastoma family of tumor suppressor and are crucial for DNA replication [30, 32].

**Protein sequence similarities and phylogenetics of TbelPyV1**

The putative LT, ST, VP1, and VP2 aa sequences of TbelPyV1 were aligned with those of related PyVs. With aa identities of 64–77.5% and 61.5–76.4%, LT, ST, VP1, and VP2 of TbelPyV1 shared the highest aa sequence identities with the corresponding proteins of 2 primate PyVs, *Human polyomavirus 9* (HPyV9, GenBank no. NC_015150) and *African green monkey polyomavirus* (AGMPyV, GenBank no. NC_004763), respectively (Table 4). To further determine the genetic relationship between TbelPyV1 and the other PyVs, phylogenetic tree was constructed based on the alignment of LT. In the tree, TbelPyV1 fell within the clade of genus *Alphapolyomavirus* and clustered with HPyV9 and AGMPyV (Fig. 3).

**Discussion**

Both PVs and PyVs have double-stranded circular DNA genomes and share many structural features. Because of that, PVs and PyVs were formerly classified into the same family of *Papovaviridae* [33]. However, the genomes of PVs and PyVs are quite different in size and organization, and contain no major sequence homology to each other. In addition, PVs and PyVs show other different biological properties, like PV transcription is unidirectional, but PyV transcription is bidirectional. Because of these differences, PVs and PyVs were classified in the year 2000 into two different families, *Papillomaviridae* and *Polyomaviridae* [8, 34].

![Fig. 2 Phylogenetic analysis of TbelPV1 and TbelPV2. The tree was constructed based on alignment of complete amino acid sequences of L1 by the maximum likelihood method with 1000 bootstrap replicates. Bootstrap values above 50% are shown. For all PVs included in the tree, GenBank accession numbers, full names, and genera of are shown. TbelPV1 and TbelPV2 detected in this study are indicated in bold with solid triangle.](image)
Recently, the Papillomaviridae family has been divided into 2 subfamilies, Firstpapillomavirinae and Secondpapillomavirinae, due to the extension of host range and phylogenetic relationships of PVs [34, 35]. The Firstpapillomavirinae subfamily consists of 53 genera including PVs infecting mammals, reptiles, and birds. The Secondpapillomavirinae subfamily has only one Alefpapillomavirus genus including PVs infecting fishes. The L1 is the most conserved gene among PVs. To distinguish genus, less than 60% sequence identity of the pairwise alignments of the L1 genes was used as the criterion. Practically, visual inspection of phylogenetic tree was also considered [36].

The L1 gene of TbelPV1 showed the highest similarity to that of viruses in genera of Mupapillomavirus, Kap-papapillomavirus and Lambdapapillomavirus. However,
TbelPV1 does not cluster with any of the above 3 genera, but branches off at the root of the common branch of the Mupapillomavirus and Kappapapillomavirus genera (Fig. 2). As for the TbelPV2, though it was clustered with genus Iotapapillomavirus according to the phylogenetic tree, it showed a long evolutionary distance with MnPV1 in Iotapapillomavirus. Due to these facts, we suggest that TbelPV1 and TbelPV2 could be considered as the representative species of two novel genera in Firstpapillomavirinae.

PVs are highly species-specific and have a specific tropism for squamous epithelial cells [37], and previous evidence showed that a large spectrum of different PVs exhibit skin tropism. The majority of PV infections are subclinical, while certain PVs are associated with an increasing number of squamous cell carcinomas at specific sites, for example, HPV types 16, 18, 31, 33, 35 [7, 38]. By phylogenetic analysis, TbelPV1 clustered with members of genera Mupapillomavirus and Kappapapillomavirus which indicates a common ancestor. 

### Table 5 Predicted open reading frames (ORF) of TbelPyV1, and the molecular masses (kiloDalton) and isoelectric points (pI) of potential encoded proteins

| ORFs      | position | length | MW   | pI   |
|-----------|----------|--------|------|------|
|           | nt       | aa     | kDa  | pI   |
| URRa      | 1-463    | –      | –    | –    |
| VP2       | 464–1438 | 975    | 324  | 35.82| 5.02 |
| VP3       | 821–1438 | 618    | 205  | 23.96| 6.71 |
| VP1       | 1395–2501| 1107   | 388  | 40.1 | 6.07 |
| NCRb      | 2502-2567| –      | –    | –    |
| LT        | 2568–4385| 2058   | 685  | 78.34| 6.67 |
| ST        | 4413–4982| 570    | 189  | 22.37| 8.31 |

*a*Upstream regulatory region

*b*Noncoding region

Fig. 3 Phylogenetic analysis of TbelPyV1. The tree was constructed based on alignment of complete amino acid sequences of LT by the maximum likelihood method with 1000 bootstrap replicates. Bootstrap values above 50% are shown. For all PyVs included in the tree, GenBank accession numbers, full names, and genera of are shown. TbelPyV1 detected in this study is indicated in bold with solid triangle. Another shrew PyV, Tupaia glis polyomavirus, is indicated in bold.
**Mupapillomavirus** genus, HPV1 is the causative agent of human deep palmo-plantar warts, and HPV63 causes punctate keratotic lesions of the foot [39, 40]. In **Kappapapillomavirus** genus, Rabbit oral papillomavirus 1 causes benign epithelial infections in the oral cavity and Cottontail rabbit papillomavirus 2 infects various sites of skin of cottontail rabbit and causes warts [4, 41, 42]. The TbelPV2 showed close genetic relationship to MnPV1 which infects different skin sites of *Mastomys natalensis* and induces tumors [43]. In this study, the TbelPV1 and TbelPV2 were only detected in oral swabs, which indicates that both viruses exhibit oral epithelium tropism. Whether these viruses could cause any lesions needs further investigation.

The **Polyomaviridae** family contains 4 genera *Alpha*-, *Beta*-, *Delta*-, and *Gammapolyomavirus*. The novel species of PyV infecting fish have not yet been assigned to a genus [8]. Polyomavirus-like sequences that did not exhibit the typical genome organization of PyVs, have also been found in arthropods, which indicates that PyVs infect or previously infected invertebrates, probably indicating an ancient evolutionary history [44]. In order to define PyV species, the ICTV established 5 criteria [8]. The TbelPV1 characterized in this study showed 60.8% nt identity to the most closely related species HPyV9 in the LT coding sequence, which is much less than the criterion (<85%). The TbelPV1 genome displays an organization typical for PyVs, and the natural host has been determined. In the phylogenetic tree, TbelPV1 clustered with other PyVs in the genus *Alphapolyomavirus*, which suggests that TbelPV1 is a novel species in this genus.

For most mammalian PyVs, the target cells for initial entry, and the exact routes of infection and transmission, are unclear. Previous studies indicate MCPyV, HPyV6, and HPyV7 may initially infect skin and lead to skin-to-skin contact transmission [45]. The MWPyV, HPyV10, STLPyV, HPyV11, and HPYV12 have been detected in the gastrointestinal tract which suggests a fecal-oral transmission route [46]. The JCPyV, BKPyV, KIPyV, and WUPyV are frequently detected in tonsillar tissue and respiratory aspirates, suggesting a route of respiratory tract transmission. In addition, the JCPyV has been found in lymphoid tissues including bone marrow and spleen, BKPyV in lung and salivary glands, WUPyV and KIPyV in plasma and urine [47, 48]. These findings suggest that PyVs can infect various tissues and use diverse transmission routes. TbelPV1 was most closely related to the HPyV9 and a PyV derived from African green monkey (*Chlorocebus aethiops*) (AGMPyV) (Fig. 3). AGMPyV is lymphotropic and HPyV9 was found in the serum of a kidney transplant patient under immunosuppressive treatment [49]. Here, TbelPV1 was detected in oral and spleen samples with high rates, which indicates that oral tissues like tonsil and spleen are the major sites of infection for TbelPV1. However, we did not perform pathological examination due to limited tissue samples. Therefore, the pathogenicity of TbelPV1 remained unclear.

PVs and PyVs evolve remarkably slowly (1.95 × 10^{-8} for PVs and 8 × 10^{-9} for PyVs, measured in nucleotide substitutions per site per year) and are assumed to be ancient viruses that have co-evolved with their host species at least over half a billion years [44, 50]. These characteristics make PVs and PyVs fascinating candidates for development of viral evolutionary and virus-host co-evolutionary models. In addition, phylogenetic analysis based on genomic level showed that Scandentia are most closely related to primates and may be the evolutionary transition from insectivores to primates [17, 51]. Our first description of 2 novel PVs and a novel PyV in tree shrews indicates that tree shrews may carry diverse PVs and PyVs. The novel genomes of PVs and PyV characterized in this study, together with the additional PVs and PyVs discovered in tree shrews in the future, would promote the study of viral evolution and virus-host co-evolution, and might shed light on the origin of PVs and PyVs of primates.

**Conclusion**

Two novel papillomaviruses (TbelPV1 and TbelPV2), and one novel polyomavirus 1 (TbelPyV1) were detected in Chinese tree shrews, and their genomes were characterized. Phylogenetic analysis based on the LT indicated that TbelPV1 and TbelPV2 represent members of two novel PV genera. Phylogenetic analysis based on the LT indicated that the TbelPV1 represents a new species within genus *Alphapolyomavirus*. The epidemiological study showed that TbelPV1 and TbelPV2 were prevalent in the oral cavity, while TbelPV1 was prevalent in the oral cavity and spleen. Whether these viruses cause disease in tree shrews needs further investigation.

**Additional file**

**Additional file 1: Table S1.** Primers used for detection and sequencing of TbelPV1, TbelPV2, and TbelPV1 in tree shrew. (DOCX 17 kb)

**Abbreviations**

BLAST: Basic Local Alignment Search Tool; HPB: Human papillomavirus HPV; ICTV: The International Committee on Taxonomy of Viruses; LCR: Long control region; LT: Large T antigen; ORF: Open reading frame; PCR: Polymerase chain reaction; PV: Papillomavirus; PyV: Polyomavirus; ST: Small T antigen; TbelPV1: Tupaia belangeri chinensis papillomavirus 1; TbelPV2: Tupaia belangeri chinensis papillomavirus 2; TbelPyV1: Tupaia belangeri chinensis polyomavirus 1; URR: Upstream regulatory region

**Acknowledgements**

Not applicable.

**Funding**

This work was jointly funded by the National Key Research and Development Program of China (2017YFD0500104), the National Natural...
Science Foundation of China (31470260, 81401672, and 81660558), and the Hu-Xiang Youth Talents Scholar Program of Hunan Province (2017RS3017).

Availability of data and materials
The genomic sequences of TbelPV1 and TbelPV2, and the TbelPyV1 were deposited in GenBank under the following accession numbers: MK443496, MK443497, and MK443498.

Authors’ contributions
YZ-Z, XY-G, JH-Z, WH-Y, X-H performed the experiments. XY-G, Y-Q, and YZ-Z analyzed the data and prepared the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate
The manuscript. All authors read and approved the final manuscript.

Consent for publication
Not applicable.

Published online: 18 March 2019

References
1. Truchado DA, Williams RA, Benitez L. Natural history of avian papillomaviruses. Virus Res. 2018;252:58–67.
2. Eleni C, Corteggio A, Altamura G, Meoli R, Cocumelli C, Rossi G, Friedrich KG, Di Cerbo P, Borzacchiello G. Detection of papillomavirus DNA in cutaneous squamous cell carcinoma and multiple Papillomas in captive reptiles. J Comp Pathol. 2017;157(1):23–6.
3. Muhr LSA, Bzhalava D, Lagheden C, Eklund C, Johansson H, Forslund O, Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol. 2013;30(12):2725–9.
4. Abbate EA, Berger JM, Botchan MR. The X-ray structure of the SV40 T-antigen origin binding domain in complex with DNA. PLoS Biol. 2007;5(2):e23.
5. DeCaprio JA. Garcea RL. A cornucopia of human polyomaviruses. Nat Rev Microbiol. 2013;11(4):264–76.
6. Gardner SD, Field AM, Coleman B. New human papovavirus (Bk) isolated from urine after renal transplantation. Lancet. 1971;1(7172):1253–7.
7. van der Meijden E, Janssens RW, Lauber C, Bouwes Bavinck JN, Gorbalenya AE, Feltkamp MC. Discovery of a new human papovavirus associated with trichodysplasia spinulosa in an immunocompromised patient. PLoS Pathog. 2010;6(7):e1000124.
37. Biryukov J, Meyers C. Papillomavirus infectious pathways: a comparison of systems. Viruses. 2015;7(8):4303–25.
38. Cubie HA. Diseases associated with human papillomavirus infection. Virolology. 2013;445(1–2):21–34.
39. Danos O, Katinka M, Yaniv M. Human papillomavirus 1a complete DNA sequence: a novel type of genome organization among papovaviridae. EMBO J. 1982;1(2):231–6.
40. Egawa K, Delius H, Matsukura T, Kawashima M, de Villiers EM. Two novel types of human papillomavirus, HPV 63 and HPV 65: comparisons of their clinical and histological features and DNA sequences to other HPV types. Virolology. 1993;194(2):789–99.
41. Gill I, Danos O, Yaniv M. Genomic structure of the cottontail rabbit (Shope) papillomavirus. Proc Natl Acad Sci U S A. 1985;82(6):1580–4.
42. Christensen ND, Cladel NM, Reed CA, Han R. Rabbit oral papillomavirus complete genome sequence and immunity following genital infection. Virolology. 2000;269(2):451–61.
43. Muller H, Grissmann L. Mastomys natalensis papilloma virus (MnPV), the causative agent of epithelial proliferations: characterization of the virus particle. The Journal of general virology. 1978;41(2):2315–23.
44. Buck CB, Van Doorslaer K, Peretti A, Geoghegan EM, Tisza MJ, An P, Katz JP, Pipas JM, McBride AA, Camus AC, et al. The ancient evolutionary history of polyomaviruses. PLoS Pathog. 2016;12(4):e1005574.
45. Liu W, Yang R, Payne AS, Schowalter RM, Spurgeon ME, Xu X, Buck CB, You J. Identifying the target cells and mechanisms of Merkel cell polyomavirus infection. Cell Host Microbe. 2016;19(6):775–87.
46. Schowalter RM, Pastrana DV, Pumphrey KA, Moyer AL, Buck CB. Merkel cell polyomavirus and two previously unknown polyomaviruses are chronically shed from human skin. Cell Host Microbe. 2010;7(6):509–15.
47. Kean JM, Rao S, Wang M, Garcea RL. Seroepidemiology of human polyomaviruses. PLoS Pathog. 2009;5(3):e1000363.
48. Gossai A, Waterboer T, Nelson HH, Michel A, Willhauck-Fleckenstein M, Farzan SF, Hoen AG, Christensen BC, Kelsey KT, Marst CJ, et al. Seroepidemiology of human polyomaviruses in a US population. Ann J Epidemiol. 2016;183(1):61–9.
49. Scuda N, Hofmann J, Calvignac-Spencer S, Ruprecht K, Liman P, Kuhn J, Hengel H, Ehlers B. A novel human polyomavirus closely related to the african green monkey-derived lymphotropic polyomavirus. J Virol. 2011;85(9):4586–90.
50. Van Doorslaer K. Evolution of the Papillomaviridae. Virolology. 2013;445(1–2):11–20.
51. Lindblad-Toh K, Garber M, Zuk O, Lin MF, Parker BJ, Washietl S, Kheradpour P, Ernst J, Jordan G, Maucler E, et al. A high-resolution map of human evolutionary constraint using 29 mammals. Nature. 2011;478(7370):476–82.