Novel polyomaviruses in shrews (Soricidae) with close similarity to human polyomavirus 12

Alma Gedvilaite,1 Morten Tryland,2 Rainer G. Ulrich,3 Julia Schneider,3† Vaida Kurmauskaite,1 Ugo Moens,4
Hannah Preugschas,5 Sébastien Calvignac-Spencer6 and Bernhard Ehlers7,*

Abstract

Shrews (family Soricidae) have already been reported to host microorganisms pathogenic for humans. In an effort to search for additional infectious agents with zoonotic potential, we detected polyomaviruses (PyVs) in common shrew, crowned shrew, and pygmy shrew (Sorex araneus, S. coronatus and S. minutus). From these, 11 full circular genomes were determined. Phylogenetic analysis based on large T protein sequences showed that these novel PyVs form a separate clade within the genus Alphapolyomavirus. Within this clade, the phylogenetic relationships suggest host-virus co-divergence. Surprisingly, one PyV from common shrew showed a genomic sequence nearly identical to that of the human polyomavirus 12 (HPyV12). This indicated that HPyV12 is a variant of a non-human PyV that naturally infects shrews. Whether HPyV12 is a bona fide human-tropic polyomavirus arising from a recent shrew-to-human transmission event or instead reflects a technical artefact, such as consumable contamination with shrew material, needs further investigation.

Polyomaviruses (PyVs; family Polyomaviridae) infect many vertebrates, including fish, birds and mammals [1]. It is thought that PyVs, like many other viruses with a double-stranded DNA (dsDNA) genome, have been associated to their hosts over extended timescales [2]. Host-virus co-divergence has been suggested as a major driver of PyV diversification, both at very deep and shallow evolutionary timescales [2–4]. However, host-switching events must be taken into account and a recent study suggested that the importance of this process may have been generally underestimated for dsDNA viruses, including PyVs [5]. De facto, the budgerigar fledgling disease polyomavirus (Aves polyomavirus 1) provides a striking example of a PyV infecting multiple host species, i.e. capable of host switch. It is therefore possible that PyVs infecting other vertebrates are occasionally transmitted to humans and contribute to the overall zoonotic burden [6]. Amongst the mammals that have been shown to be reservoirs for zoonotic pathogens, primates, bats and, to a lesser extent, rodents have already revealed a considerable diversity of PyVs [4, 7–17]. Yet, almost none of these viruses seem to represent a direct ancestor of any of the human polyomaviruses identified to date. Exceptions might be the two human PyVs most recently identified, New Jersey PyV [18] and Lyon IARC PyV [19]. In both cases, the authors put their human origin into question, and both PyVs have so far not been detected in follow-up studies.

Shrews (family Soricidae) host a number of zoonotic agents such as diverse hantaviruses [20–23], classical borna disease virus [24], Anaplasma phagocytophilum [25], Bartonella...
opposite strand, separated by a non-coding control region and displayed 81%–82% nucleotide sequence identity to PyV1-GER_#9778 (acc. no. MF401582), were 97.1 ScorPyV1-GER_#9777 (acc. no. MF375001), and Scor- PyV1-GER_#7615 (acc. no. MF624714), SminPyV1-GER_#7607 (acc. no. MF401583), SminPyV1-GER_#7614 (acc. no. MF624714), and SminPyV1-GER_#7613 (acc. no. MF624714), were 99–100% identical and were tentatively named SaraPyV1-GER_#4608 (Fig. S1). The start/stop codon positions and splice sites of all coding sequences are completely conserved between HPyV12 and SaraPyV1-GER_#4608. One silent nucleotide exchange is present in both the LTAg and VP1 coding sequence (CDS). More prominent differences were observed in the NCCR which displays five deletions of 2×2, 12, 13 and 33 base pairs in the HPyV12 genome (compared to Sara-pyV1-GER_#4608 genome) (Fig. S1). As these deletions might affect sites important for DNA replication and gene expression, e.g. origin of replication and binding sites for LTAg and host transcription factors [36, 37], it is of note that the AT-rich motif in the origin of replication and the LTAg binding motifs are not affected. As determined with Allgen_PROMO [38], putative binding sites for human transcription factors are not removed in HPyV12 NCCR, except for the 33 base pairs deletion. The functional implications of altered transcription factor binding sites in HPyV12 NCCR compared to the SaraPyV1-GER_#4608 NCCR remain to be investigated.

Given these observations, we then aimed at characterising further the evolutionary history of the shrew PyVs and understand how closely they were associated to their hosts. For this, we first performed maximum likelihood (ML) phylogenetic analyses of an amino acid sequence alignment of the LTAg including representatives of all PyV species currently recognised by the International Committee on Taxonomy of Viruses (https://talk.ictvonline.org/ictv-reports/ictv_9th_report/dsdna-viruses-2011/w/dsdna_viruses/129/ polyomaviiridae). We performed the model selection with Prottest v3 [39] and the phylogenetic analysis with PhyML v3 [40]. This revealed that (i) the novel shrew PyVs formed a single, deep-branching clade within the clade of the alphapolyomaviruses (genus Alphapolyomavirus) and (ii) as expected, HPyV12 nested within the shrew PyV clade (Figs 1 and S2). Neither additional LTAg-based (Fig. S2) nor VP1-based (not shown) phylogenetic analyses including more sequences revealed any other closely related PyV.

To determine whether the diversification process of the hosts had influenced that of their PyVs, we performed ML phylogenetic analyses on (i) a nucleotide sequence alignment of the LTAg coding region including all unique shrew PyV sequences as well as the HPyV12 sequence and (ii) a nucleotide sequence alignment of 19 host mitochondrial cytochrome b sequences (complete or nearly complete)
which were selected with T-Coffee [41] from a total of 546 publicly available sequences so as to maximize the representation of host genetic diversity (the partial cytochrome \( b \) sequences that we determined from the PyV-positive shrews nested all in the respective clade; data not shown). Here we performed model selection using jmodeltest v2 [42] and
also built ML phylogenetic trees using PhyML v3. The PyV phylogeny recapitulated the host phylogeny, i.e. (S. minutus, S. araneus, S. coronatus) (Fig. 2a, b; [43, 44]).

Since this pattern was suggestive of host-PyV co-divergence, we also investigated the relative divergence dates of the PyV infecting common shrews and crowned shrews and of their hosts themselves. To do this, we performed uncalibrated molecular clock analyses of the shrew PyV LTAg and host mitochondrial cytochrome b sequences and examined the ratios of the time to the most recent common ancestor (tMRCA) of the PyV in common shrews and crowned shrews or the tMRCA of common shrews and crowned shrews themselves to the age of the root of the tree (respectively represented by nodes 2 and 1 in the ML phylogenetic trees; Fig. 2). These analyses were performed with BEAST v1.8.2 [45]. Under a constant population size coalescent model, this ratio was larger in the PyV LTAg [median 0.63, 95% highest posterior density: (0.47–0.88)] than in the cytochrome b analyses [median 0.24, 95% HPD (0.12–0.41)]. The same was observed using a birth-death speciation model (data not shown). While these results do not pinpoint a synchronicity of the divergence of PyV and their hosts, the divergence of the PyV lineages prior to the divergence of their hosts is still compatible with the notion of co-divergence. Indeed, the time to coalescence of co-diverging viral lineages can only exceed the time of divergence of their respective hosts [3]. All in all, the shrew PyV clade uncovered in our study is likely to represent a lineage of PyV which has been associated with shrews over evolutionary timescales. Accordingly, HPyV12 is, in fact, a shrew-derived virus, and originates in the species S. araneus. We note here that, by calibrating our shrew PyV LTAg analysis using the divergence time of their hosts (ca. 6.7 My ago for the tMRCA of the three species; [43]), we obtained nucleotide substitution rates [median 1.48E-8 sub.s-1.y-1, 95% HPD (1.03E-8–2.02E-8) under a constant population size model; similar values under a birth-death model] very close to published estimates for PyV long-term substitution rates [2, 46].

The data presented here indicate that with HPyV12 a shrew polyomavirus was detected in humans, either, because HPyV12 was recently transmitted to humans and is now circulating in the human population, or tends to spill over to humans. However, artefactual detection due to a technical failure has to be taken into account. We initially detected HPyV12 nucleic acid sequences in 16 samples of human gastro-intestinal tract (liver, colon, rectum) and 1 stool sample [47]. Since some shrew samples were collected, stored, and extracted in our laboratory before we processed the HPyV12-positive human samples, the identification of HPyV12 in human samples may represent an artefact, i.e. DNA contamination or sample mixing. We could exclude a mix-up of human and shrew samples by amplifying with cytochrome b-generic PCR and sequencing of human cytochrome b sequence from the DNA of the human sample that gave rise to the published complete HPyV12 genome [47] and two other DNA extracts of the above mentioned HPyV12-positive set that were still available (data not shown). Furthermore, DNA contaminations occur most frequently from PCR products. Such product contamination can also be excluded here, as the HPyV12 genome was amplified years before any of the shrew PyV genomes. The third possibility for an artefactual origin of HPyV12 was contamination of the 16 human DNA extracts with DNA from common shrews. DNA extraction columns and other reagents used in PCR have been reported to be contaminated with eukaryotic, viral, and bacterial DNA [48–51]. We therefore performed specific PCR targeting common shrew (but not human) cytochrome b sequences. This assay was positive for common shrew DNA extracts (as confirmed by sequencing), but negative for the 3 HPyV12-positive human DNA extracts and blank DNA extracts. Specific PCR, targeting human (but not shrew) cytochrome b sequences gave the opposite result, i.e. was positive for the human but not the shrew DNA extracts (data not shown).

Although these findings demonstrated the absence of Sorex DNA from the human samples they do not strictly exclude contamination with Sorex PyV DNA. In addition, the presence of shrew PyV DNA in those extraction columns or PCR reagents that had been used for extraction of the HPyV12-positive human samples could not be investigated, as these or materials from the same batches were not available anymore.

To determine whether a shrew PyV with a strict genome identity to HPyV12 is present in our shrew samples, we used specific primers to amplify and sequence the NCCR, where most differences between HPyV12 and its closest shrew-infecting relative had been observed. None of 16 NCCR of SaraPyV1 was identical to the HPyV12 NCCR, as HPyV12 presents three unique deletions of >10 nucleotides (Fig. 3). However, this does not firmly prove the absence of a PyV genome with the exact HPyV12 sequence in our shrew samples, and it does not exclude contamination of the human samples with HPyV12-identical PyV DNA from other sources.

The hypothesis of a shrew-to-human transmission is supported by serological studies that reported a HPyV12 seroprevalence in humans of 23% (VP1 pentamer-based ELISA of 373 human sera [47]) or 19% [yeast-derived VP1 virus-like particle based ELISA of 627 sera (Alma Gedvilaite, unpublished data)], and 50% (LTAg-based ELISA on 186 sera [52]). These lines of evidence however do not end up with unambiguously identifying HPyV12 as a virus circulating in humans, as cross-reactivity was frequently observed in serological assays targeting PyV (e.g. [53]).

In summary, we showed here that shrews of the family Soricidae naturally harbor closely related PyVs, with which they likely co-diverged. We also showed that HPyV12 originates in shrews. While this may be a first hint for animal-to-human PyV transmission, it is not possible at this stage to exclude an environmental contamination of the human samples with shrew material. An important step forward...
Fig. 2. Maximum likelihood trees of LTAg-encoding nucleotide sequences from shrew (Sorex spp.) polyomaviruses (a) and cytochrome b nucleotide sequences from Sorex spp. (b). Note that the sequence disrupting the monophyly of S. araneus in B (DQ630406) was obtained from a shrew collected in China and may therefore represent another species. Branch robustness was assessed with Shimodaira-Hasegawa-like approximate likelihood ratio tests (SH-like aLRT); branches with SH-like aLRT values <0.95 are shown in grey. Root positions were determined from the family-level analysis (a) and from the literature [58]; the same root positions were confirmed by a search for the best-fitting root with Tempest v1.5 [59]. The scales are in nucleotide substitutions per site.
would be detection of HPyV12 in human samples by a laboratory that does not work with shrew material. Of note, HPyV12 was recently detected with a commercial kit in 1/20 squamous cell carcinoma [54], at a rate of 5,734 viral copies/10,000 cells. Although interesting it is a singular finding that was not verified by sequencing and not confirmed in another laboratory. In other recent studies on PyVs in kidney transplant patients [55] and in non-malignant tonsillar tissue and tonsillar carcinoma [56], HPyV12 was not detected. Future studies are thus needed to investigate if HPyV12 is a shrew virus that was artefactually detected in human samples, or is a bona fide human virus that arose from a transmission event.

Funding information
The authors received no specific grant from any funding agency.

Acknowledgements
The excellent technical assistance of Sonja Liebmann, Cornelia Walter, Eva M. Breines, Dörte Kaufmann, Anke Mandelkow and Ellinor Hareide is gratefully acknowledged. Thanks to Dr Tore Sandvik and Dr Ronny Ehlers for contributing to the sampling of shrews in Norway and Germany, respectively. The sample characterization and transfer was supported by the German Federal Ministry of Education and Research (BMBF) through the German Research Platform for Zoonoses (Network Rodent-borne pathogens’, FKZ 01KI1018 and 01KI1303, granted to RGU).

Conflicts of interest
The authors declare that there are no conflicts of interest.

Ethical statement
This article does not describe experimental work with humans.

References


---

**Five reasons to publish your next article with a Microbiology Society journal**

1. The Microbiology Society is a not-for-profit organization.
2. We offer fast and rigorous peer review – average time to first decision is 4–6 weeks.
3. Our journals have a global readership with subscriptions held in research institutions around the world.
4. 80% of our authors rate our submission process as ‘excellent’ or ‘very good’.
5. Your article will be published on an interactive journal platform with advanced metrics.

*Find out more and submit your article at microbiologyresearch.org.*