Isolation and genomic characterization of the first Norway rat (Rattus norvegicus) papillomavirus and its phylogenetic position within Pipapillomavirus, primarily infecting rodents

Eric Schulz,1 Marc Gottschling,2 Gudrun Wibbelt,3 Eggert Stockfleth1 and Ingo Nindl1,4

Correspondence
Marc Gottschling
gottschling@bio.lmu.de

1Clinic for Dermatology, Venereology and Allergology, Skin Cancer Center Charité (HTCC), Charité – Universitätsmedizin Berlin, Berlin, Germany
2Department of Biology, Systematic Botany, Ludwig-Maximilians-Universität München, Munich, Germany
3Institut für Zoo- und Wildtierforschung, Berlin, Germany
4DKFZ – Charité, Viral Transformation Mechanisms, German Cancer Research Center, Heidelberg, Germany

A series of papillomavirus (PV) types have been isolated from different rodent species, and most of them belong to the genus Pipapillomavirus. We isolated and sequenced the complete genome of a novel PV type (designated RnPV) from the oral cavity of the Norway rat (Rattus norvegicus), as well as an L1 gene fragment from hair-follicle cells of the European beaver (Castor fiber). As inferred from amino acid sequence data, RnPV clustered within the \( \beta + \gamma + \pi + \zeta \)-PV supertaxon as a member of the genus Pipapillomavirus. The closest relatives of RnPV were McPV-2 and MmPV, and time estimates indicated that the genus Pipapillomavirus originated in the late Cenozoic era. The close relationship of RnPV to other murid PV types supports the hypothesis of co-divergence between members of the genus Pipapillomavirus and their hosts. However, the derived Neogene origin of the genus Pipapillomavirus is much younger than has been considered for the Rodentia as the primary hosts, indicating that alternative interpretations of the phylogenetic trees should be conceived.

Papillomaviruses (PVs) are a diverse group of small, non-enveloped, double-stranded DNA viruses. Their structurally conserved, circular genome comprises approximately 8000 bp and is organized in up to eight open reading frames (ORFs), including the four major genes E1 and E2 (i.e. early genes) and L2 and L1 (late genes). PVs infect epithelial tissues of humans and different animal species, causing cutaneotropic and mucosotropic proliferative lesions. The presence of genital PV types is the major risk factor of cervical cancer in humans (zur Hausen, 2002), and various types of \( \beta \)-PV (genus Betapapillomavirus) seem to be involved as a cofactor in the development of cutaneous squamous cell carcinoma (Nindl et al., 2007). Despite the growing knowledge of virus–host interactions, the molecular mechanisms are only partially known. Thus, further studies including animal models are warranted to examine PV-induced skin carcinogenesis.

Taking advantage of the rolling-circle amplification method (Rector et al., 2004), the number of isolated PV types has increased over the past few years. Approximately 150 PV genomes, isolated from humans and different animal species, have been cloned and completely sequenced so far (Van Ranst et al., 1992; Terai et al., 2002; Forslund et al., 2003; Rector et al., 2007). However, the number of known non-human host species is still low (approx. 40 of >4000 mammalian species), and the true diversity of animal PVs is largely unknown at this moment in time.

PV types from only four rodent species, including the hamster (HaOPV; Iwasaki et al., 1997), the multimammate rat (MnP-1, McPV-2; Müller & Gissmann, 1978; Tan et al., 1994; Nafz et al., 2008), the European harvest mouse (MmPV; Van Doorslaer et al., 2007) and the North American porcupine (EdPV; Rector et al., 2005), are known. The majority of these PV types constitute the genus Pipapillomavirus (\( \pi \)-PV); only EdPV and MnPV-1 show a distant relationship to other rodent PVs. Concomitantly, rodents are not the only hosts of \( \pi \)-PVs; a sequence of an

The GenBank/EMBL/DDBJ accession number for the complete RnPV genome sequence is GQ180114.

Supplementary material is available with the online version of this paper.
L1 fragment isolated from a gorilla is most similar to the corresponding gene region of McPV-2 (Antonsson & Hansson, 2002; Gottschling et al., 2008).

In the past years, phylogenetic analytical methods have been applied more rigorously for the reconstruction of PV evolution (García-Vallvé et al., 2005; Narechania et al., 2005; Rector et al., 2007). Four PV supertaxa have been identified, namely $\alpha + \infty$, $\beta + \gamma + \pi + \xi$, $\delta + \varepsilon$ and $\kappa + \lambda + \mu + \upsilon + \sigma$-PVs (Gottschling et al., 2007b), and most known PV types can be classified into one of them. PVs are specific to their hosts and are generally considered to have co-diverged closely with mammals and other vertebrates (Van Ranst et al., 1995; Bernard et al., 2006). However, other evolutionary mechanisms such as recombination (Narechania et al., 2005; Varsani et al., 2006) and infection across species borders (’zoonosis’; Myers et al., 1998) may contribute to their diversification (Gottschling et al., 2007).

If congruence between phylogenies of viruses and their hosts is shown, it is possible to infer viral substitution rates from the divergence times of the hosts. DNA viruses are considered to evolve slowly, and several attempts have been made to calculate rates of evolution for PVs. A rough rate of $10^{-7}$ nucleotide substitutions per base position year$^{-1}$ (s/b/y) has been inferred from the correspondence of approximately 2% difference in the L1 gene between major variants of mucosotropic types (Ong et al., 1993; Halpern, 2000; Chen et al., 2005) and an extended reduction of the human population (’bottleneck’) that may have ended 100 000 years ago (Gibbons, 1995; Harpending et al., 1998). Most reliable evolutionary rates for PVs, ranging between $1.44 \times 10^{-8}$ and $2.69 \times 10^{-8}$ s/b/y, have been inferred from congruent phylogenies between members of the Carnivora and the genus Lambdapapillomavirus (’$\lambda$-PV; Rector et al., 2007). However, timing other viral groups may prove difficult, even if, for example, rodent phylogenetic trees (Huchon et al., 2002; Montgelard et al., 2008) and PV trees were indeed congruent: the age of the Rodentia, and therefore a possible calibration point, is controversially debated at present (Adkins et al., 2003).

In this study, we present the complete genome and its organization of a novel PV type isolated from the oral cavity of a Norway rat [Rattus norvegicus (Berkenhout, 1769); Muridae; Rodentia]. Moreover, we describe an L1 gene fragment sequence of a novel PV type isolated from a European beaver (Castor fiber Linnaeus, 1758; Castoridae; Rodentia). To some degree, the internal topology of $\pi$-PVs is congruent with the phylogeny of rodents, as inferred from maximum-likelihood (ML) analyses and Bayesian inference, and we therefore compare the derived evolutionary rate with previously published data.

Cells of a healthy, free-ranging Norway rat (R. norvegicus; female, adult, field sample, Berlin, Germany) from the oral, genital and anogenital region were collected by using sterile Q-Tips (see Supplementary Table S1, available in JGV Online). From the same animal, 10–15 hairs (eyebrow hairs and whiskers) were collected under sterile conditions. Moreover, approximately 30 hairs were removed under sterile conditions from the facial region of a European beaver (C. fiber; female, adult; see Supplementary Table S1). DNA isolation and cloning, protein prediction and phylogenetic analyses of amino acid alignments were basically performed as reported previously (Schulz et al., 2009) and are described in detail in the Supplementary Methods (available in JGV Online).

To consider the maximal known diversity of $\pi$-PVs [i.e. five rodent and one primate PVs, including the new sequences of RnPV and the FAP fragment (derived by using the FA primers; see Supplementary Methods) from the European beaver], a nucleotide alignment of the late gene FAP region was constituted for timing divergence events by using BEAST v. 1.4.8 (Drummond & Rambaut, 2007; http://evolve.zoo.ox.ac.uk/beast/; three independent chains each of $2 \times 10^7$ generations, sampled every one-thousandth iteration). In the absence of a reliable geological age of the Rodentia or a subordinate taxon (Adkins et al., 2003), we applied an evolutionary rate of $1.84 \times 10^{-8}$ nucleotide substitutions per site year$^{-1}$ (n/s/y) to the data matrix, using a log-normal relaxed clock and partitioned into codon positions. This rate was published for the L1 gene, inferred from congruent host and PV phylogenies between carnivores and $\lambda$-PVs (Rector et al., 2007).

The RnPV genome (GenBank accession no. GQ180114) consisted of 7378 bp. As inferred from protein-prediction software programs and sequence evaluation in a comparative alignment, seven ORFs were encoded on the same strand and in the same orientation, namely the genes E6, E7, E1, E2, E4, L2 and L1. The potential E4 ORF was nested within the E2 ORF and was characterized by proline-rich stretches. All gene descriptions and protein primary sequence analyses are summarized in Table 1. The non-coding region (NCR) was located upstream of the E6 gene and comprised 556 nt (7.5% of the complete genome). A compilation of the predicted transcription factor-binding sites and other non-coding elements, as well as of structural elements of the RnPV proteins, is provided in Supplementary Tables S2 and S3 (available in JGV Online). RnPV (L1 and/or E1) was only detected in the oral cavity and was absent in hairs and cells of the (ano-)genital areas examined.

The L1 sequence of RnPV showed highest similarity to that of MmpPV (84.7%), as inferred from a sequence-similarity matrix (available at http://htcc.pt-dlr.de/dateien/SchulzRnPv.xls). The corresponding region in the complete genome was identical to an L1 fragment sequence initially generated by using FA primers (see Supplementary Methods). The alignment for the multigene phylogenetic analysis was 2352 aa in length (Table 1); 1618 of these sites (69%) were parsimony-informative (21.9 per terminal taxon). Fig. 1 shows the best-scoring ML tree ($-\ln=165305.02$) with the statistical-support values for the two phylogenetic approaches used.

E. Schulz and others
Mammalian PVs were monophyletic, irrespective of whether the data were analysed under the likelihood criterion (bootstrap-support value from ML analysis: 100 LBS) or the Bayesian approach (Bayesian posterior probability: 1.00 BPP). They basically segregated into the four monophyletic supertaxa $\alpha + \omega$-PV (89 LBS, 1.00 BPP), $\beta + \gamma + \pi + \xi$-PV (69 LBS, 1.00 BPP), $\delta + \psi$-PV (100 LBS, 1.00 BPP) and $\kappa + \lambda + \mu + \nu + \sigma$-PV (97 LBS, not supported by the Bayesian analysis). Rodent PVs were found in three distantly related lineages in the phylogenetic tree: (i) EdPV (Iotapapillomavirus $\sigma$-PV) within the supertaxon $\kappa + \lambda + \mu + \nu + \sigma$-PV as closest relative of human papillomavirus 41 (100 LBS, 1.00 BPP), (ii) MnPV-1 (Iotapapillomavirus $\pi$-PV) with a basal and uncertain phylogenetic position, and (iii) $\pi$-PV (100 LBS, 1.00 BPP) within the supertaxon $\beta + \gamma + \pi + \xi$-PV. RnPV belonged to the $\pi$-PVs and constituted a monophyletic group with other PV types isolated from Muridae, namely MmPV [from Micromys minutus (Pallas, 1771)] and McPV-2 [from Mastomys coucha (Smith, 1834)] (100 LBS, 1.00 BPP).

Using an evolutionary rate of $1.84 \times 10^{-8}$ n/s/y, the origin of monophyletic $\pi$-PVs (1.00 BPP) was dated to the Oligocene–Miocene boundary about 23 million years (Ma) ago (Fig. 2), but 95% confidence intervals were wide. An L1 fragment isolated and sequenced from a European beaver (GenBank accession no. GQ352456) constituted the sister group of the remaining $\pi$-PVs (0.97 BPP). The divergence between RnPV and the most similar relative, MmPV (0.96 BPP), took place about 12 Ma ago. The majority of $\pi$-PVs were found on rodents, but an L1 fragment isolated and sequenced from a gorilla was the closest relative of McPV-2 (1.00 BPP). The divergence between them was dated to approximately 12 Ma.

Despite the efforts that have been made to isolate and characterize novel PV types from different host species during the past years, knowledge about PV diversity generally is still sparse. PV types from classical laboratory animals such as Mus musculus Linnaeus, 1758 and R. norvegicus have not been identified to date. In this study, we present the complete genome and its organization of a novel PV type isolated from the oral cavity of a Norway rat. It can be assumed that RnPV shares similar properties to McPV-2, which induces anogenital tumours in Mastomys coucha, because these PV types are closely related. So far, Mastomys coucha and the cottontail rabbit are the only animal models that have been used to examine PV-induced carcinogenesis (Amtmann et al., 1984; Wettstein et al., 1987; Naftz et al., 2007). A rat model might additionally prove useful to study the molecular mechanisms of PV-induced carcinogenesis in future.

Despite limited knowledge of PV diversity, the viruses have been generally regarded as co-diverging with their corresponding hosts (Van Ranst et al., 1995; Bernard et al., 2006). However, the overall co-phylogenetic hypothesis has been rejected for rodent PVs, because they do not constitute a monophyletic group. The polyphyletic positions of $\pi$-PVs (comprising four rodent PV types), $\omega$-PVs (including MnPV-1) and $\sigma$-PVs (including EdPV) indicate that alternative evolutionary mechanisms may have contributed to PV diversification, such as the establishment of new ecological niches by adaptive radiation (Garcia-Vallé et al., 2005; Jackson, 2005).

RnPV clearly belongs to the $\pi$-PVs and constitutes a monophyletic group with other murid PV types, although the closest relative cannot be determined unambiguously. If co-divergence drove $\pi$-PV diversification, RnPV should be allied closely to MmPV, because R. norvegicus is without doubt allied more closely to Micromys minutus than to Mastomys coucha (Rowe et al., 2008). However, this hypothesis is only supported by separate analyses of the genes E6, E7 and L2, as well as the Bayesian L1 fragment analysis. It remains to be determined whether the possible close relationship between RnPV and McPV-2, as inferred from the multigene analysis, indeed reflects the phylogeny of the group or rather results from methodological artefacts. Thus, the possible co-phylogenetic structure between Rodentia and $\pi$-PVs should still be considered with some reservation.

Timing divergence events has become increasingly popular during the past years, including in PV evolutionary virology (Halpern, 2000; Chen et al., 2005). Under the most

### Table 1. Primary sequence analysis of the RnPV genes

<table>
<thead>
<tr>
<th>Gene</th>
<th>All</th>
<th>E6</th>
<th>E7</th>
<th>E1</th>
<th>E2</th>
<th>E4</th>
<th>L2</th>
<th>L1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length (bp)</td>
<td>7378</td>
<td>423</td>
<td>306</td>
<td>1815</td>
<td>1164</td>
<td>345</td>
<td>1641</td>
<td>1536</td>
</tr>
<tr>
<td>Length (aa)</td>
<td>–</td>
<td>140</td>
<td>101</td>
<td>604</td>
<td>387</td>
<td>114</td>
<td>546</td>
<td>511</td>
</tr>
<tr>
<td>Molecular mass (kDa)</td>
<td>–</td>
<td>15530.3</td>
<td>10994.6</td>
<td>68464.1</td>
<td>43854.1</td>
<td>12704.3</td>
<td>58058.7</td>
<td>57757.2</td>
</tr>
<tr>
<td>Isoelectric point</td>
<td>–</td>
<td>8.62</td>
<td>6.88</td>
<td>5.69</td>
<td>8.78</td>
<td>9.15</td>
<td>4.57</td>
<td>8.21</td>
</tr>
<tr>
<td>Alignment length (aa)</td>
<td>3640</td>
<td>304</td>
<td>255</td>
<td>854</td>
<td>325</td>
<td>368</td>
<td>920</td>
<td>613</td>
</tr>
<tr>
<td>No. parsimony-informative positions (percentage; per terminal taxon)</td>
<td>2231</td>
<td>166</td>
<td>122</td>
<td>615</td>
<td>260</td>
<td>Excluded</td>
<td>613</td>
<td>455</td>
</tr>
<tr>
<td>(61%; 30.15)</td>
<td>(55%; 2.24)</td>
<td>(48%; 1.65)</td>
<td>(72%; 8.31)</td>
<td>(80%; 3.51)</td>
<td>(67%; 8.28)</td>
<td>(74%; 6.15)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Fig. 1. RnPV belongs to the genus Pipapillomavirus, members of which primarily infect rodents. An ML tree of PVs is shown, comprising a representative set of 20 human and 54 non-human types, as inferred from a combined E6–E7–E1–E2–L1 amino acid analysis (1618 parsimony-informative amino acid positions). Generic PV clades (de Villiers et al., 2004) are indicated by Greek lettering. RnPV is highlighted by a pink arrow and the genus Pipapillomavirus by a shaded box. Abbreviations: ART, ‘Artiodactyla’; CAR, Carnivora; CET, Cetacea; CHI, Chiroptera; INS, ‘Insectivora’; LAG, Lagomorpha; PAS, Passeriformes; PER, Perissodactyla; PRI, Primates; PSI, Psittaciformes; ROD, Rodentia; SIR, Sirenia; TES, Testudines. Supertaxa in the sense of Gottschling et al. (2007b) are coloured red (α+β-PV), green (β+γ+π+ι-PV), blue (δ+ε+ζ-PV) and ochre (κ+λ+μ+ν+σ-PV), respectively. Branch lengths are drawn to scale. Bar, 0.6 amino acid substitutions per site. Numbers on branches are bootstrap-support values to clusters of the right of them (above, ML criterion; below, BPP; values <50 and <0.90, respectively, are not shown).
Radiation of the genus Pipapillomavirus took place in the late Cenozoic era. A dated phylogeny of π-PVs is shown, as inferred from an L1 fragment sequence analysis (including new PV sequences isolated from R. norvegicus (RnPV) and C. fiber (PV106)). Ultrametric maximum clade-credibility tree with node ages from Bayesian uncorrelated log-normal analysis. Bars correspond to 95% confidence intervals and colours of the branches indicate evolutionary (median) rate variation (green, slow; red, fast). Absolute ages are given in million years (Ma), and epochs (i.e. Eocene, Oligocene, Miocene, Pliocene and Pleistocene) are coloured differently.

Among myriads of L1 fragments that have been isolated (Adkins et al., 2003) and sequenced primarily from humans and other primates (Antonsson & Hansson, 2002; Adkins et al., 2003). Two interpretations of this conflict in our time estimates are conceivable: either the apparent congruence between PV and host phylogenies does not correspond to the true evolution of the group (due to the very limited knowledge of rodent PV diversity) or evolutionary rates are significantly faster in π- than in λ-PVs. The relatively balanced branch lengths of PV phylogenetic trees given by Gottschling et al. (2007b) make the latter assumption rather implausible.

As reliable evolutionary rate for PV available at present (Rector et al., 2007), the origin of π-PVs is much younger than has been considered for their primary hosts, namely the Rodentia (Huchon et al., 2002; Adkins et al., 2003). The close relationship of this conflict in our time estimates are conceivable: either the apparent congruence between PV and host phylogenies does not correspond to the true evolution of the group (due to the very limited knowledge of rodent PV diversity) or evolutionary rates are significantly faster in π- than in λ-PVs. The relatively balanced branch lengths of PV phylogenetic trees given by Gottschling et al. (2007b) make the latter assumption rather implausible.

Fig. 2. Radiation of the genus Pipapillomavirus took place in the late Cenozoic era. A dated phylogeny of π-PVs is shown, as inferred from an L1 fragment sequence analysis (including new PV sequences isolated from R. norvegicus (RnPV) and C. fiber (PV106)). Ultrametric maximum clade-credibility tree with node ages from Bayesian uncorrelated log-normal analysis. Bars correspond to 95% confidence intervals and colours of the branches indicate evolutionary (median) rate variation (green, slow; red, fast). Absolute ages are given in million years (Ma), and epochs (i.e. Eocene, Oligocene, Miocene, Pliocene and Pleistocene) are coloured differently.

In conclusion, RnPV belongs to the π-PVs, which primarily infect rodents. The close relationship of RnPV to other PV types isolated from different murid species suggests that evolutionary co-divergence with hosts is important for PV diversification. Time estimates indicate that π-PVs originated in the late Cenozoic era. RnPV is the first completely sequenced PV type isolated from a classical laboratory animal, providing a promising basis for future studies on the complex molecular mechanisms of virus–host interactions in the natural host.

References


