

## Supplementary Data

Warden CD and Lacey SF (2012) Updated Phylogenetic Analysis of Polyomavirus-Host Co-Evolution. *J Bioinfo Res.* 1(4): 46-49

## Methods

PyV Genome Sequences: Polyomavirus genome sequences for AGMPyV (NC\_004763), BFPyV (NC\_004764), BKV (NC\_001538), BPyV (NC\_001442), CPyV (NC\_007922), CSLPyV (GQ331138), FPyV (NC\_007923), GggPyV1 (HQ385752), GHPyV (NC\_004800), HPyV6 (HM011558), HPyV7 (NC\_014407), HPyV9 (HQ696595), JCV (NC\_001699), KIV (NC\_009238), MCPyV (NC\_010277), MPyV (NC\_001515), MptV (NC\_001505), MWPyV (JX262162), MyPyV (NC\_011310), OraPyV1 (NC\_013439), OraPyV2 (FN356901), PtvPyV1 (HQ385746), PtvPyV2 (HQ385748), SquiPyV (NC\_009951), SV12 (NC\_012122), SV40 (NC\_001669), TSV (GU989205), and WUPyV (NC\_009539) were downloaded from Genbank.

PyV Phylogenetic Analysis and Tanglegram: PyV genome sequences were aligned using CLC Bio Genomics Workbench version 4.7.2 (CLC Bio, Katrinebjerg, Denmark) with a gap open cost of 10, gap extension cost of 1, end gap cost set “as any other” and the CLC Bio alignment algorithm was set on the “Very accurate” setting. Phylogenetic trees were constructed in CLC Bio Genomics Workbench using the Unweighted Pair Group Method with Arithmetic Mean (UPGMA) algorithm [1] and statistical significance was assessed via bootstrap analysis with 100 replicates.

Host-Parasite Tanglegram: The host phylogenetic tree was created using data from the phylogenetic analysis of Bininda-Emonds et al. [2] as well as the Tree of Life project [3]. The tanglegram used to visualize virus-host coevolution was created using TreeMap 3 [4, 5]. For the tanglegram comparing the UPGMA phylogenetic tree to the host tree, nodes with less than 95% bootstrap support were removed to better represent the ambiguity for certain regions of the genome alignment. The untangle function was used to improve clustering of viruses and hosts in the tanglegram.

Mauve PyV Alignment: Polyomavirus genome sequences were aligned using progressiveMauve [6] and locally collinear blocks (LCBs) from the alignment visualized in Mauve alignment viewer [7]. The phylogenetic tree representing the backbone for the Mauve alignment was created using the *newick2phylog* function in the *ade4* package [8] in R [9].

## References

1. Michener C & Sokal R (1957) A Quantitative Approach to a Problem in Classification. *Evolution* 11(2).
2. Bininda-Emonds ORP, Cardillo M, Jones KE, *et al.* (2007) The delayed rise of present-day mammals. *Nature* 446(7135):507-512.
3. Maddison DR & Schulz K-S (2007) The Tree of Life Web Project. eds Maddison DR & Schulz K-S.
4. Page RDM (1994) Parallel phylogenies: reconstructing the history of the host-parasite assemblages. *Cladistics* 10(2):155-173.
5. Page RDM & Charleston MA (1998) Trees within trees: phylogeny and historical associations. *Trends in ecology & evolution (Personal edition)* 13(9):356-359.
6. Darling AE, Mau B, & Perna NT (2010) progressiveMauve: Multiple Genome Alignment with Gene Gain, Loss and Rearrangement. *PLoS ONE* 5(6):e11147.
7. Darling ACE, Mau B, Blattner FR, *et al.* (2004) Mauve: Multiple Alignment of Conserved Genomic Sequence With Rearrangements. *Genome Research* 14(7):1394-1403.
8. Dray S & Dufour AB (2007) The ade4 package: implementing the duality diagram for ecologists. *Journal of Statistical Software* 22(4):1-20.
9. R Development Core Team (2011) R: A Language and Environment for Statistical Computing. (R Foundation for Statistical Computing, Vienna, Austria).