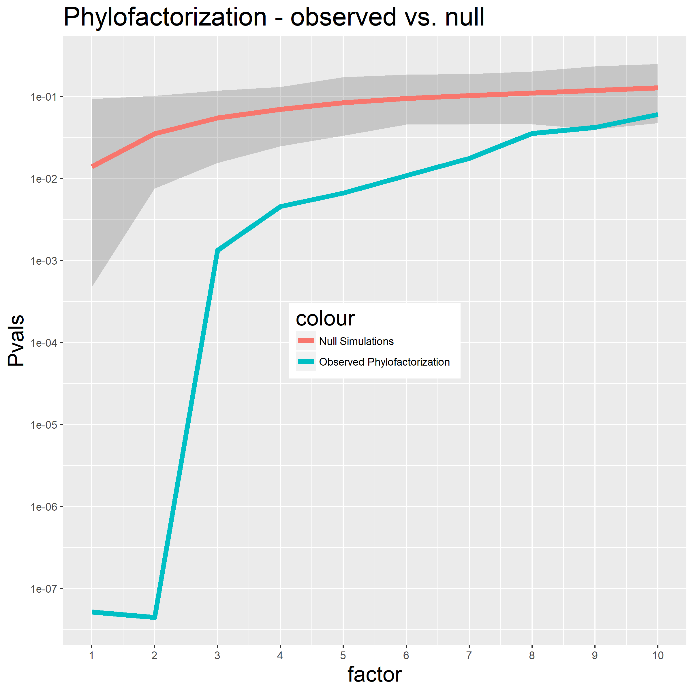
Phylogenetic factors of zoonosis in mammalian viruses

Alex Washburne, Daniel Becker, Daniel Crowley & Raina Plowright

Many emergent infectious diseases are zoonoses, pathogens capable of infecting more than one host, and one major goal in monitoring and managing emergent infectious diseases is to identify potential zoonotic viruses and survey their abundances in wildlife populations. Olival et al. (2017) curated a dataset of all known mammalian viruses and classified zoonotic viruses using viral traits, such as host phylogeny, enveloped/non-enveloped, nucleic acid composition and others.

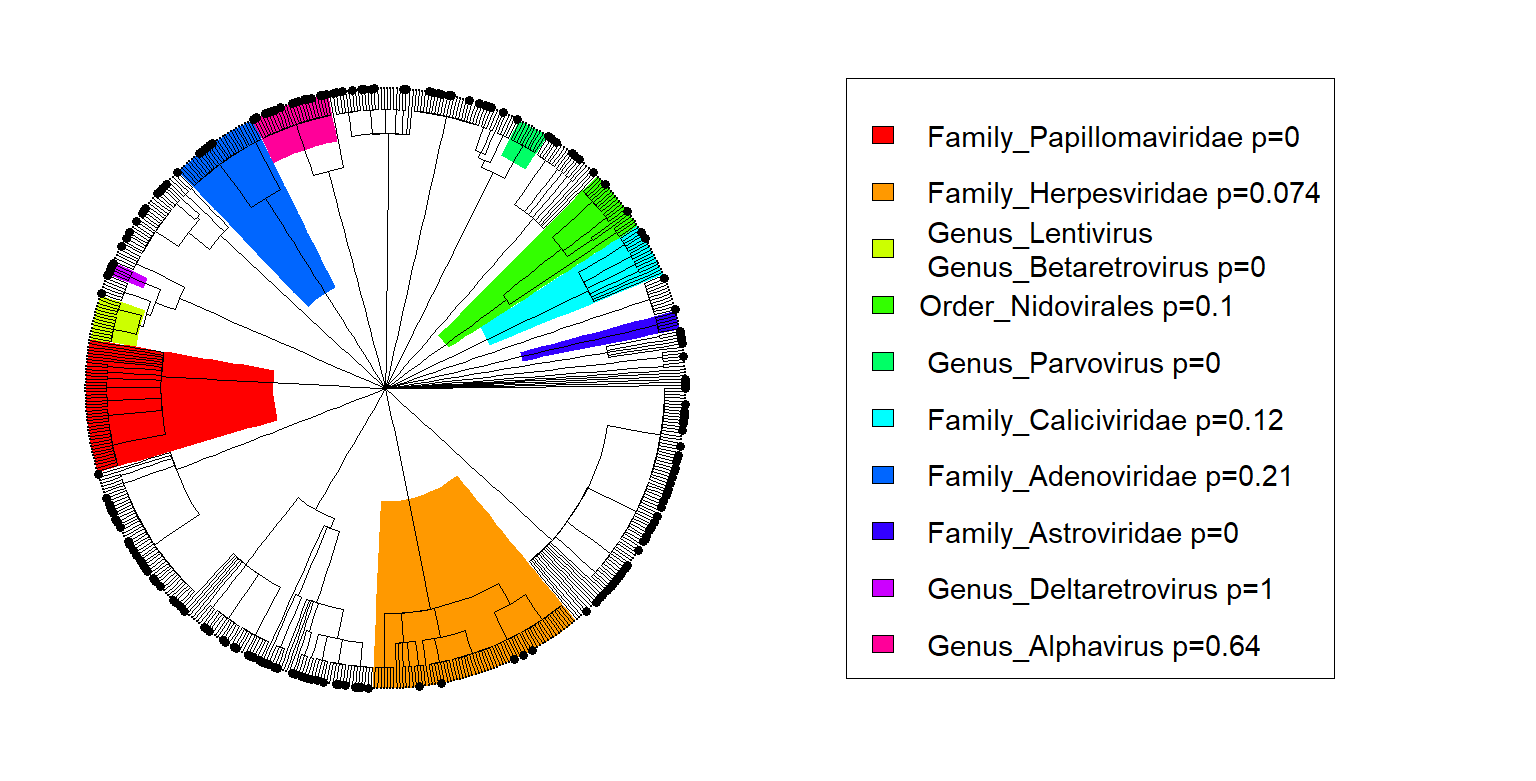
However, classification of zoonotic viruses using trait datasets is sensitive to precisely which traits are included. If trying to understand which traits classify “flight” in vertebrates, we might identify feathers, hollow bones and wings. Do feathers make birds fly, or hollow bones, or wings? A more concise & accurate analysis is to say that birds fly, and many traits together assist their flight.

To sequentially infer & control for phylogenetic patterns of zoonosis, we use a novel method called “phylofactorization” (Washburne et al. 2017). Phylofactorization is a graph-partitioning algorithm, also viewed as a graph-constrained clustering algorithm, which partitions the phylogeny into clades with high within-clade similarity and maximum between-clade differences. Phylofactorization ultimately makes inferences on edges in the phylogeny along which hypothesized trait differences arise.

For example, if classifying vertebrates based on land/water habitat association, a few edges in the phylogeny suffice to capture most of the variation in the dataset: tetrapods, cetaceans, pinnipeds, penguins & amphibians. Phylofactorization can identify 5 edges in the phylogeny which capture most of the variation in vertebrates’ land/water habitat association, edges which correspond to the evolution of major traits.

We performed phylofactorization to identify phylogenetic bins of zoonotic viruses in the Olival et al. dataset. We used the viral taxonomy tree as a coarse phylogeny. The highest taxonomic level of each lineage was connected to the root to allow contrasts across lineages in the graph-partitioning scheme. The taxonomic graph was partitioned by iteratively identifying the edges which maximized Fisher’s exact test statistic contrasting the fraction of zoonotic viruses on opposing sides of the edge. We compared the resultant sequence of P-values from Fisher’s test with those obtained from 1,000 phylofactorizations of null datasets constructed by random assignment of zoonosis to viruses independent of taxonomy.

Phylofactorization yielded 10 significant edges partitioning the tree into 11 different bins with a range of zoonotic potential. Several clades at a range of taxonomic scales – the families Papillomaviridae and Astroviridae, the order Nidovirales and the genus betaretroviruses, for example – have few to no zoonotic representatives. Other clades, such as the genera deltaretroviruses and alphaviruses, have an unusually high fraction of zoonotic representatives. The alphaviruses, interestingly, specialize in arthropod intermediates – mosquitoes & ticks – and thus, like birds & flight, may have many traits which assist their ability to generalize between hosts. Their ability to generalize between hosts may ultimately have evolved from selection by their historical exposure to the sera from multiple animals.



By identifying significant taxonomic bins of low or high spillover risk, we aim to propose surveillance strategies by identifying primer sets that target high-priority clades, i.e. clades with a high propensity for spillover. We will repeat the trait-based analyses by phylogenetic bin to obtain key predictors of spillover within clades having a common zoonotic pattern and understand how inclusion of Papillomaviridae, for example (a large clade with not one zoonotic representative) confounds the analysis of traits predicting zoonotic risk.