The fall-out and fruits of big-data in primate research: disentangling evolutionary cause and effect

David Schruth, University of Washington
Marc Myers, Primate Conservation Inc.
Noel Rowe, Primate Conservation Inc.

Examination into hypothetical origins of primate morphology and behavior can be challenged by many possible confounding ecological factors. Statistical modeling with multivariate and phylogenetic control can be a fruitful approach if tools and data are readily accessible. Human readable, webpage-style data sources are useful, but may not be machine readable or amenable to usage in statistical software programs. Recent systematic efforts in tabulating data on primates exist in the form of individual silos of smaller, single-table compilations or via the All The World's Primates [AWP] larger relational SQL database. While much effort has gone into organizing and curating such datasets, the downloading of data on multiple-species can be difficult or infeasible without specific technical training. With the increasing popularity of statistical analysis environments, we can bridge this gap between these modern analysis tools and the machine scalable data archives which are often coded in non-human readable ways. We introduce here a package for use in the R statistical programing environment containing tools for accessing tabular extractions from AWP along with aggregated summary datasets and demonstrations of common analysis routines. As an example, we illustrate how the pervasive hazard of arboreal fall-risk in primates is relatively neglected in studies because the data (e.g. canopy height utilization) is often sparse or absent. We further demonstrate recoding routines to transform variables (e.g. diet) into more efficiently recoded equivalents.