**Test data and analysis ReadMe**

**Overview**

For each analysis referred to in the main text there is a corresponding folder containing the files used to generate the results. Data files and trees from the references cited in the main text were obtained from Gregg Thomas with the number of taxa reduced to facilitate reasonable computation time. The commands issued for each analysis are given below with additional details as appropriate. Each analysis was run multiple times to ensure convergence of all parameter estimates and -ln likehood values.

**Primate Analysis**

Analysis of 12 primate taxa under increasing numbers of discrete gamma rate categories was performed as below (with multiple runs to ensure convergence).

cafexp -t 12Primates.tree -i 12Primate\_genefamilies.tab -k 1 -o 1RATE\_Primates\_RealDat

…

…

cafexp -t 12Primates.tree -i 12Primate\_genefamilies.tab -k 10 -o 10RATE\_Primates\_RealDat

The count file with the value of *K* for which the highest likelihood was estimated (*K*=4) was used to generate an empirical root distribution for use in simulation.

grep Fam Gamma\_count.tab|wc -w

24 (last column contains the root values, so we can cut, count, and sort).

grep -v -P "\<" Gamma\_count.tab|cut -f24|sort|uniq -c|sort -r|awk '{print $2"\t"$1}'|sort -n >Root\_dist.txt

Command used to simulate data ( “simulation\_k4\_L.00453\_A.62.txt”)

cafexp -f Root\_dist.txt -t 12Primates.tree -k 4 -s -l 0.00453 -a 0.62 -o 12Primates\_Sim

Simulated data was analyzed as above.

cafexp -t 12Primates.tree -i simulation\_k4\_L.00453\_A.62.txt -k 1 -o 1RATE\_Primates\_SimDat

…

…

cafexp -t 12Primates.tree -i simulation\_k4\_L.00453\_A.62.txt -k 7 -o 7RATE\_Primates\_SimDat

**Hymenoptera Analysis**

Analysis of 10 Hymenoptera taxa under increasing numbers of discrete gamma rate categories was performed as below (with multiple runs to ensure convergence).

cafexp -t 10Hymenoptera.tree -i 10Hymenoptera\_genefamilies.tab -k 1 -o 1RATE\_Hymen\_RealDat

…

…

cafexp -t 10Hymenoptera.tree -i 10Hymenoptera\_genefamilies.tab -k 10 -o 10RATE\_Hymen\_RealDat

The count file with the value of *K* for which the highest likelihood was estimated (*K*=6) was used to generate an empirical root distribution for use in simulation.

grep Fam Gamma\_count.tab|wc -w

20

grep -v -P "\<" Gamma\_count.tab|cut -f20|sort|uniq -c|sort -r|awk '{print $2"\t"$1}'|sort -n >10Hymen\_rootdist.txt

cafexp -f 10Hymen\_rootdist.txt -t 10Hymenoptera.tree -k 6 -s -l 0.00375 -a 0.3725 -o 10Hym\_Sim

Simulated data was analyzed using the following commands with multiple runs to ensure convergence.

cafexp -t 10Hymenoptera.tree -i simulation\_k6\_L.00375\_A.37.txt -k 1 -o 1RATE\_Hym\_SimDat

…

…

cafexp -t 10Hymenoptera.tree -i simulation\_k6\_L.00375\_A.37.txt -k 7 -o 7RATE\_Hym\_SimDat

**Birds of Paradise Analysis**

cafexp -t 8BoP.tree -i 8BoP\_genefamilies.tab -k 1 -o 1RATE\_BoP\_RealDat

…

…

cafexp -t 8BoP.tree -i 8BoP\_genefamilies.tab -k 10 -o 10RATE\_BoP\_RealDat

The count file with the value of *K*=7 was used to generate an empirical root distribution for use in simulation.

grep Fam Gamma\_count.tab|wc -w

16

grep -v -P "\<" Gamma\_count.tab|cut -f16|sort|uniq -c|sort -r|awk '{print $2"\t"$1}'|sort -n >8BoP\_rootdist.txt

Data was simulated using 7 rate categories instead of the maximum likelihood value of *K*=10 as the lambda value for this dataset was nearly identical for *K*=2-10 and the difference in likelihood was minimal. Command used to simulate data

(“simulation\_K7\_L.00226\_A.98.txt”)

cafexp -f BoP\_rootdist.txt -t 8BoP.tree -k 7 -s -l 0.00226 -a 0.98 -o 8BoP\_Sim

Simulated data was analyzed using the following commands

cafexp -t 8BoP.tree -i simulation\_K7\_L.00226\_A.98.txt -k 1 -o 1RATE\_BoP\_SimDat

…

…

cafexp -t 8BoP.tree -i simulation\_K7\_L.00226\_A.98.txt -k 7 -o 7RATE\_BoP\_SimDat