



# 首届致远学术节 学生科研成果展示

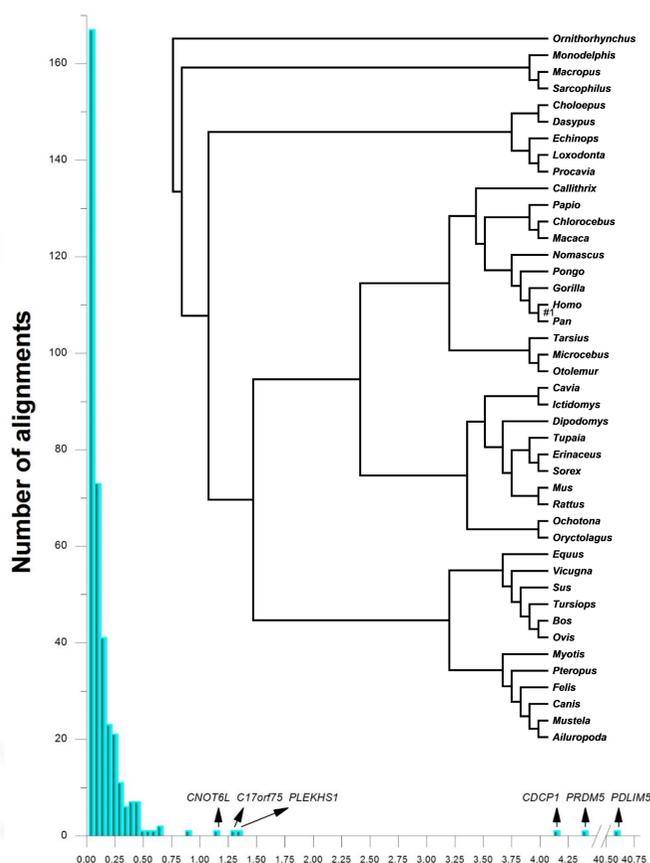
## Detecting Positively Selected Genes among Mammalian Species Using Phylogenetic Analysis of Maximum Likelihood

By Haikuo Li

Research Field: Molecular evolution, Bioinformatics

Directed by Manyuan Long, Edna K. Papazian Distinguished Service Professor, the University of Chicago

**Introduction:** The birth of new novel genes that originated from various molecular mechanisms plays an essential role in understanding and elaborating the evolution of genetic systems. Mutation and selection have different effects on synonymous silent (Ks) and nonsynonymous amino acid (Ka) substitution rates. The  $\omega$  value ( $\omega = Ka/Ks$ ) significantly greater than one is considered to be an important indicator for positive selection at the molecular level.



**Methods:** Phylogenetic Analysis of Maximum Likelihood, phylogenetic analyses, and likelihood ratio tests

**Results:** 1. Almost all of the orthologous genes are under negative purifying selection in the basic model.

- 362 of 368 genes have the  $\omega$  ratios lower than one, which means negative purifying selection, and there exist six genes with  $\omega > 1$  including one gene with  $\omega > 40$  suggesting that the gene might have undergone a positive selection in nearly all of the mammalian species.

2. Four PSGs are detected linking human species in the branch model.

- Only four genes are detected in the branch model to be significantly under positive selection and significantly variable in  $\omega$  for the branch linking with human species. The method in our research is able to lend supports to predicting well-functioning genes in species and foreseeing that it might play an important phylogenetic role in the evolution process in certain species.

**Figure 1.** An example of maximum likelihood tree of the CDS of one gene among 43 mammalian species, and the histogram of  $\omega$  values calculated from 368 genes in 43 mammalian species in the  $1\omega$  model

**Table 1.** Likelihood Ratio Statistics for the  $2\omega$  model compared with the  $1\omega$  model and the hypothesis test

Gene marker identifier	$\omega_1$	$\omega_2$	$\ln L_2$	$\ln L_1$	$2\Delta l_1$	$P_1$	$\omega_0$	$\ln L_0$	$2\Delta l_0$	$P_0$
ENSG00000172578 KLHL6	0.0021	2.3437	-37.4489	-34.6524	5.593	0.0180*	0.0021	-34.6524	5.593	0.0180*
ENSG00000120708 TGFBI	0.0049	2.3423	-193.9492	-191.2627	5.373	0.0204*	0.0050	-191.2627	5.373	0.0204*
ENSG00000127463 EMC1	0.0011	2.3426	-357.7519	-364.5068	13.510	0.0002**	0.0033	-364.5069	13.510	0.0002**
ENSG00000115839 RAB3GAP1	0.0855	2.3383	-1216.2192	-1213.5500	5.338	0.0209*	0.0863	-1213.4833	5.472	0.0193*

个人信息：2015级，生命科学  
邮箱：lihaikuo@sjtu.edu.cn