# Patterns and effects of GC3 heterogeneity and parsimony informative sites on the phylogenetic tree of genes

Shuai Ma, Qi Wu, Yibo Hu, Fuwen Wei

PII: S0378-1119(18)30173-2
DOI: doi:10.1016/j.gene.2018.02.037
Reference: GENE 42589

To appear in: *Gene*

Received date: 2 November 2017
Revised date: 27 January 2018
Accepted date: 12 February 2018

Please cite this article as: Shuai Ma, Qi Wu, Yibo Hu, Fuwen Wei , Patterns and effects of GC3 heterogeneity and parsimony informative sites on the phylogenetic tree of genes. The address for the corresponding author was captured as affiliation for all authors. Please check if appropriate. Gene(2017), doi:10.1016/j.gene.2018.02.037

# Patterns and effects of GC3 heterogeneity and parsimony informative sites on the phylogenetic tree of genes

Shuai Ma[a,b], Qi Wu[a], Yibo Hu[a,*], Fuwen Wei[a,b]

[a] Key Laboratory of Animal Ecology and Conservation Biology, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China.

[b] University of Chinese Academy of Sciences, Beijing 100049, China.

[*] Corresponding author: ybhu@ioz.ac.cn

## Abstract

The explosive growth in genomic data has provided novel insights into the conflicting signals hidden in phylogenetic trees. Although some studies have explored the effects of the GC content and parsimony informative sites (PIS) on the phylogenetic tree, the effect of the heterogeneity of the GC content at the first/second/third codon position on parsimony informative sites ($GC1/2/3_{PIS}$) among different species and the effect of PIS on phylogenetic tree construction remain largely unexplored. Here, we used two different mammal genomic datasets to explore the patterns of $GC1/2/3_{PIS}$ heterogeneity and the effect of PIS on the phylogenetic tree of genes: (i) all $GC1/2/3_{PIS}$ have obvious heterogeneity between different mammals, and the levels of heterogeneity are $GC3_{PIS} > GC2_{PIS} > GC1_{PIS}$; (ii) the number of PIS is positively correlated with the metrics of 'good' gene tree topologies, and excluding the third codon position (C3) decreases the quality of gene trees by removing too many PIS. These results provide novel insights into the heterogeneity pattern of $GC1/2/3_{PIS}$ in mammals and the relationship between GC3/PIS and gene trees.