

# ortholog —— 概念、生物信息预测方法和数据库

陈作舟<sup>1,2</sup>、朱晟<sup>2</sup>、薛成海<sup>3</sup>、陈良标<sup>\*2</sup>

1 浙江大学生命科学院

2 中国科学院遗传与发育生物学研究所

3 中国科学院自动化研究所复杂系统与智能科学国家重点实验室

orthologs指起源于不同物种的最近共同祖先的一些基因。orthologous的基因，具有相近甚至相同的功能，由相似的途径调控，在不同的物种中扮演相似甚至相同的角色，因此在基因组序列的注释中，是最可靠的选择。orthologs的生物信息预测方法主要有两类：系统发生方法和序列比对方法。这两类方法都是基于序列的相似性，但又各有特点。系统发生方法通过重建系统发生树来预测orthologs，因此在概念上比较精确，但难于自动化，运算量也很大。序列比对方法在概念上比较粗糙，但简单实用，运算量相对较小，因此得到了较广泛的应用。

## ORTHOLOG —— CONCEPT, BIOINFORMATIC INFERENCE AND DATABASES

Orthologs are genes in different species that originate from a single gene in the last common ancestor of these species. Orthologous genes are suggested to share similar functions, be regulated by similar biochemical pathways and play similar roles in different species. Thus, it is the best choice to use orthologous genes when annotating newly discovered genes. There are mainly two categories of algorithms for predictions of orthologs: phylogenetic algorithms and sequence comparison algorithms. Both of them are based on sequence similarities, whereas they have their own characteristics. Phylogenetic ways predict orthologs by reconstructing phylogenetic trees. As a result, they are conceptually accurate, but hard to automate, and demanding huge amount of computational resources. In contrast, the latter methods are conceptually less accurate but not as complex and require less computational resources, therefore, widely used.

关键词