

# Extraction of Modules from the Metabolic Pathways with Phylogenetic Profile

**Takuji Yamada**

takuji@kuicr.kyoto-u.ac.jp

**Yoshihiro Yamanishi**

yoshi@kuicr.kyoto-u.ac.jp

**Susumu Goto**

goto@kuicr.kyoto-u.ac.jp

**Minoru Kanehisa**

kahehisa@kuicr.kyoto-u.ac.jp

Bioinformatics Center, Institute for Chemical Research, Kyoto University, Uji, Kyoto,  
Japan

**Keywords:** modularity, metabolic pathways, phylogenetic profile, orthology

## 1 Introduction

The concept of modularity assumes that functionality can be partitioned into a collection of modules. Each module is a discrete entity of elementary components and performs an identifiable task, separable from the functions of other modules [3]. As the modules in the protein 3D structure can be related to molecular functions, the modules in the protein network may be related to cellular functions. In fact, a recent study suggests that the modularity is also seen in the metabolic network. Here, we attempt to extract modules from the metabolic network, using the similarity of the phylogenetic profile of 77 distinct organisms.

## 2 Method and Results

### 2.1 Phylogenetic Profile

The phylogenetic profile of a protein is a string of ones and zeros that encodes the presence or absence, respectively, of the protein in fully sequenced genomes [2]. Proteins that participate in a common structural complex or metabolic pathway are likely to evolve in a correlated fashion. Hence, such proteins have similar phylogenetic profiles. In this research, we made 1100 phylogenetic profiles based on the orthologs defined in KEGG [1].

### 2.2 Cluster Analysis

We defined similarities between profiles by the Euclidean distance (ED), selected components obtained by the independent component analysis (ICA) and the tree kernel function (KNL) [5], and then performed the complete linkage analysis using each similarity measure. The thresholds of each similarity are given by the statistically significant point (top 5% or 10% of total distribution). Table 1 shows the result. Next, we extracted subclusters which also form a cluster on the pathway map in KEGG. We refer to such a cluster as a module. Table 2 shows the number of extracted modules with each similarity measure.

