

Q1. What's the meaning of Vikodak?

Ans. 'Vikodak' word is originated from 'Sanskrit' language, which means 'Decoder'.

Q2. What is the utility of Vikodak?

Ans. In simplest terms, Vikodak is a tool for obtaining functional inferences from 16S data.

Validation experiments on over 1300 metagenomic samples have confirmed the utility of Vikodak in:

- a) deciphering enzyme abundance profiles of any KEGG metabolic pathway
- b) functional resolution of distinct metagenomic environments
- c) inferring patterns of functional interaction between resident microbes
- d) automating statistical comparison of functional features of studied microbiomes.

Novel features incorporated in Vikodak also facilitate automatic removal of false positives and spurious functional predictions.

Q3. Why is Vikodak called a modular framework/package? What are the utilities of each module?

Ans. Vikodak has 3 modules, each developed with an aim to logically extend the utility of the other modules:

Global Mapper Module - Global Mapper module enables (a) an in silico estimation of the relative abundance of various metabolic pathways in an environmental sample, (b) quantifying the contribution of each of the microbes (in that sample) to the predicted functions (at all three tiers of KEGG hierarchy), and (c) identification of the core set of metabolic functions defining a particular environment.

This module thus infers the functional features of a Metagenomic environment at 3 levels of KEGG Hierarchy (Pathways, Pathway Classes and Pathway Super-classes).

Inter Sample Feature Analyzer (ISFA Module) - The ISFA module in Vikodak is a logical extension of the Global Mapper module and is designed for performing a rigorous (pair-wise) comparative statistical analysis of the (inferred/predicted) function profiles generated from two or more environments.

So, if the user wants to compare two environments in terms of their functional characteristics, then all he/she has to do is to feed the functional inferences obtained from Global Mapper module to ISFA, and a comparison will be made using a Wilcoxon Rank Sum test based bootstrapped

approach.

Local Mapper Module - The Local Mapper module further enables end-users to probe in greater detail the enzyme abundance profile(s) of individual metabolic pathway(s) identified as (a) the 'core' in one or more environments, or as (b) differentially abundant between two or more environments.

To cite an example of the utility of Local Mapper - assume that based on the functional inferences obtained from Global Mapper, ISFA differentiates two environments in terms of 'Alanine, aspartate and glutamate metabolism' pathway (amongst others). In such a case, a user can utilize Local Mapper and probe deeper into the enzymatic profiles of this Pathway.

Q4. Global Mapper provides users with different choices of Algorithms, what is the basis of each of these algorithms

Global Mapper provides to end-users the choice of two distinct algorithmic work-flows viz. 'Co-metabolism' and 'Independent Contributions'.

The principle behind these two work-flows is as follows:

The **Co-metabolism algorithm** is based on the underlying assumption that the genes/enzymes expressed by various microbes residing in an environment may pool together and contribute to the functioning of specific metabolic pathway(s). Therefore, the effective abundance of a metabolic pathway (in an environment with co-metabolising microbes) is expected to be a function of the total enzyme pool contributed by the co-metabolising microbes.

In contrast, the **Independent contributions algorithm** assumes the independent existence of microbes in the environment. Under this assumption, the effective abundance of a metabolic pathway is the sum total of the pathway abundances computed from individual microbes residing in that environment. Independent Contributions algorithm can therefore enable a user to look into the functional inferences from an entirely different perspective. He/She can find out the 'Individual Contributions' of each resident microbe towards various functions inferred for the metagenomic environment

Q5. There are three options for Independent-Contributions algorithm of Global Mapper, namely Independent Contributions (M), Independent Contributions (X) and Independent Contributions (RC). What is the meaning of these terminologies and the basis of these options?

Ans. Terms M, X and RC refer to Median, Mean and Relative Contributions respectively. In the context of Independent Contributions algorithm, each of the option signifies the underlying method

of deducing the 'Individual Contributions' of resident microbes towards inferred functions.

If option Independent Contributions (M) is chosen, then the individual contribution of a given microbe towards a given function in a metagenomic environment would be deduced by taking the median of its contribution towards that function in each individual sample of the environment. Similarly, if option RC and X are chosen, then the individual contribution of a given microbe towards a given function in a metagenomic environment would be deduced by taking the Relative Contributions and Mean (respectively) of its contribution towards that function in each individual sample of the environment.

Q6. What is PEC value?

Ans. A major limitation of the current methods for function prediction is the inability to account for the constituents of a metabolic pathway. In other words, a metabolic pathway might be effected by the joint expression of over 30 genes/enzymes, but mere expression/presence of 1-5 genes/enzymes might not effect the expression of the associated pathway. It is thus crucial to define a parameter for filtration of pathways based upon the proportion of various pathway associated genes/enzymes expressed by the microbiota. Considering the need for such a parameter, Pathway Exclusion Cut-off (PEC) value has been defined in both algorithms of Global Mapper. PEC value is defined as the minimum percentage of genes/enzymes belonging to any metabolic pathway, that must be expressed by a given microbiota for tagging that pathway as being expressed by the microbiome. For example, a PEC value of 30 would mean that a given microbe/microbiota must express atleast 30% of genes/enzymes belonging to any metabolic pathway for considering that pathway as being expressed by that microbe/microbiome. Both modules of Global Mapper have been developed in such a way that apart from providing the raw hits (where raw hits refer to the most inclusive criteria wherein even the expression of single enzyme/gene is considered for presence of a metabolic pathway) they provide the KEGG pathway (all three level) expression profiles for the microbial abundance data at various PEC values (30-90) as well.

If you have any other questions related to the functionality/ operation of Vikodak, feel free to mail at sunil.nagpal@tcs.com. The subject-line of the mail should preferably be “Vikodak | Query”.