To apply for project registration, you should fill in basic information about it. Please read descriptions below carefully and provide complete and comprehensive data. All fields below are mandatory.

1. Project full name

You should start with specifying the name of your project. The "Full Name" is descriptive, and has no arbitrary restrictions (except a 40 character limit).

A Mechanistic Model for Protein Evolution

Full Name:

2. Project Purpose And Summarization

Please provide detailed, accurate description of your project and what R-Forge resources and in which way you plan to use. This description will be the basis for the approval or rejection of your project's hosting on R-Forge, and later, to ensure that you are using the services in the intended way. This description will not be used as a public description of your project. It must be written in English.

Our project’s goal is to construct a model for protein evolution that incorporates biological mechanism, population genetics and phylogenetic information. In this model, we assume that at each site of the protein there is an optimal amino acid. The functionality of a protein depends on the physiochemical distance (Grantham 1983) between its amino acids and the optimal ones as well as the selection strengths exerted on different amino acid positions. And the fitness of a protein is a function of functionality and cost of manufacturing the protein. Based on the assumption that the fixation of mutation from one protein to another takes a relatively short time, the transition rate depends on the fixation probability. Combining the expression for protein fitness and the formula for fixation probability we can find the transition rates between any pair of proteins. We can then infer the phylogenetic tree for the species whose protein sequences are drawn from GenBank.

We also want to compare the results from this model with that from the classic model used (Goldman and Yang 1993), using both simulation and real data analysis. We expect that the results will be similar, indicating that the mathematical model is biologically valid, too. This is the first attempt to incorporate biological mechanisms and population genetics with the phylogenetic approach. It will close the gap between the biological understanding and mathematical modeling of protein evolution processes.

For the data analysis we plan to use R as the main programming tool, because of its handy packages for doing phylogenetic analysis as well as versatility in programming. Since this is a group projects between two colleagues and me, we would love to share the coding along with the project progress. R-Forge is known to do a fantastic job on this and easy to use. Hence we hope that this project can be approved and carried out with the help of R-Forge.

3. License

If you selected "other", please provide an explanation along with a description of your license. Realize that other licenses may not be approved. Also, it may take additional time to make a decision for such project, since we will need to check that license is compatible with the OpenSource definition.

4. Project Public Description

This is the description of your project which will be shown on the Project Summary page, in search results, etc. It should not be as comprehensive and formal as Project Purpose description (step 2), so feel free to use concise and catchy wording. Maximum length is 255 chars.

We are constructing a protein evolution model that incorporates the biological mechanisms, population genetic and phylogenetic information. This is the first attempt to include the mechanisms in the evolution model and we intend to see if the results from the classic phylogenetic model coincide with that from the new proposed model.

5. Project Unix Name

proteinevoutk2011

In addition to full project name, you will need to choose short,"Unix" name for your project.

The "Unix Name" has several restrictions because it is used in so many places around the site. They are:

Cannot match the unix name of any other project

Must be between 3 and 15 characters in length

Must be in lower case

Can only contain characters, numbers, and dashes

Must be a valid unix username

Cannot match one of our reserved domains

Unix name will never change for this project

Your unix name is important, however, because it will be used for many things, including:

A web site at unixname.r-forge.r-project.org

A CVS Repository root of /cvsroot/unixname at cvs.unixname.r-forge.r-project.org

Shell access to unixname.r-forge.r-project.org

Search engines throughout the site

Unix Name:

As there is only one SCM system, then this will be selected automatically. scmsvn will be selected.