

PROJECT INTRODUCTION

Objectives

1. Provide a convenient resource for annotation extraction and sequence analysis
2. Capitalize on the availability of cluster and Grid Computing to speed up the process.

Project Investigator / Manager

Adrian Danker
Temasek Polytechnic
dankeran@TP.EDU.SG

Lim Yun Ping
Bioinformatics Institute
get@bii.a-star.edu.sg

Period of Project

01 Apr 2005 – 19 Aug 2005

Website

<http://get.bii.a-star.edu.sg>

Abstract

Grid Enabled Tools suite is developed for biologists to access computing resources via a user friendly web interface for high throughput bioinformatics analysis.

PROJECT DETAILS

Description

As bioinformatics analysis becomes more complex, we have developed a suite of grid enabled tools (GET) which integrate several commonly used and freely available applications such as BLAST, EMBOSS and ClustalW to run in combination in a workflow. This aims to (1) provide a convenient resource for annotation extraction and sequence analysis (2) capitalize on the availability of cluster and Grid Computing to speed up the process.

The GET suite consists of GetAnno for annotation extraction, GETSeq for sequence extraction, GetEMBOSS for sequence analysis and GetMSA for multiple sequence comparison using ClustalW. The results from these analyses are consolidated for easy retrieval. It is targeted for biologists who would like to access computing resources via a user friendly web interface for high throughput bioinformatics analysis.

GetAnno allows users extract functional annotation based on sequence similarity search to databases such as Pfam (protein family), Refseq (Reference Sequence), Gene Ontology and SwissProt. GetAnno may be used to obtain functional information for microarray and proteomics data. GetEmboss comprises of sequence analysis programs such as getorf, eprimer3, garnier, iep, pepcoil, pepstats, restrict, tmap and transeq. GetMSA consists of CLUSTALW for multiple sequence alignment and HMMBuild to build a Hidden Markov Model (HMM) profile based on the alignment generated. GetSeq allows users to extract sequences submitted for further analysis.

GetAnno is developed using PERL while GetEmboss and GetMSA, developed using phpMyAdmin. The system is implemented on a LSF (Load Sharing Facility) Multicluster set up.