FastGroup II: A web-based bioinformatics platform for analyses of large 16S rDNA libraries

Yanan Yu[1] , Pat McNairnie[1] and Forest Rohwer[1,2]

Key words: prokaryote, bacteria, 16S rDNA, biodiversity, richness, rarefaction

1 Introduction

Prokaryotes are the most abundant and diverse components of the biosphere. Most of these microbes cannot be cultured in the lab and are therefore studied by sequencing their ribosomal RNA genes (16S rDNA) [1]. We have been using high-throughput 16S rDNA sequencing to study the diversity of *Bacteria* and *Archaea* associated with reef-building corals. This has created a data glut because there are few bioinformatic tools for the analyses of large numbers of 16S rDNA sequences. To address this problem, FastGroup II has been developed.

2 Method

FastGroup II is a web-based software tool to de-replicate large 16S rDNA libraries. After cloning and sequencing, the 16S rDNA sequences are imported into FastGroup II as individual text files or as one FASTA file. The sequences are then trimmed to commonly used conserved sites [2], or to sites specified by the user. Poor-quality sequence, as determined by user-defined criteria is also eliminated. Then, using one of several different algorithms, the 16S rDNA sequences are grouped together. This grouping de-replicates the library, displays one representative sequence for all sequences that are >97% identical, and produces a table noting the number of sequences that fell into each group. The user has the ability to change both the algorithm and the grouping criteria used.

When the grouping is completed, FastGroup II automatically calculates standard diversity and richness indices, including the Shannon index, Chao1, and rarefaction of each 16S rDNA library. The result is also visualized in the rank-abundance curve using web-based graphical views. The data from each library is then stored in a MySQL database for later retrieval and comparison against other 16S rDNA libraries.

3 Results

FastGroup II is user-friendly and is currently being used in our laboratory for several studies of coral-related *Bacteria* and *Archaea*. This software package is built using Objected-Oriented Perl and new methods and functions are continuously being added. This client-server program will soon be available for public use.

---

[1]San Diego State University, Department of Biology, LS316, 5500 Campanile Drive San Diego, California 92182-4614 USA
[2]Center for Microbial Sciences, San Diego State University, 5500 Campanile Drive San Diego, California 92818-4614 USA
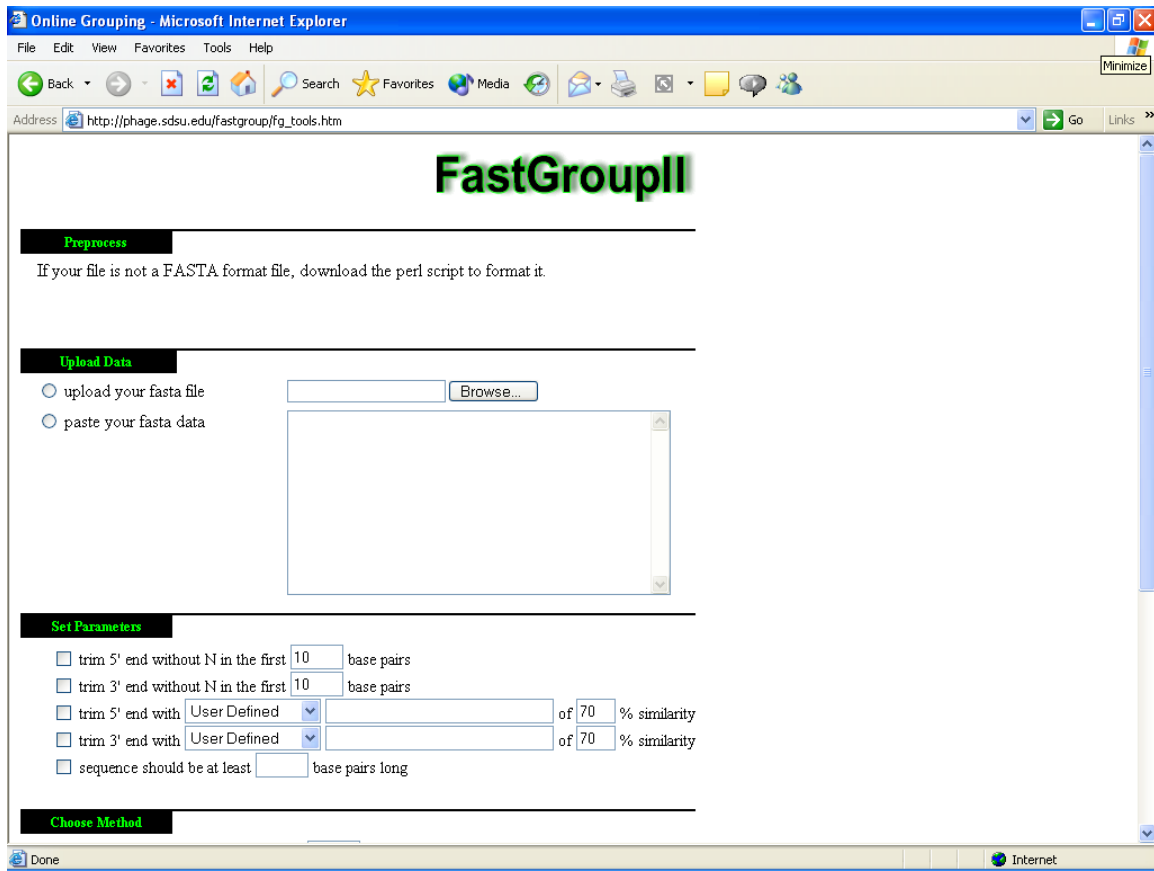
Figure 1. Part of FastGroupII user interface.

4 References
[1] Hugenholtz P, Goebel BM, Pace NR: Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *Journal of Bacteriology 1998, 180: 4765-4774*
[2] Seguritan, V, Rohwer, F: FastGroup: A program to dereplicate libraries of 16S rDNA. *BMC Bioinformatics 2001, 2:9*