

# Meta-Mesh Documents\_V26122012

---

## **1 OVERALL INTRODUCTION TO META-MESH**

Meta-Mesh is a well-organized metagenomic database system which was designed to automatically collect, store, integrate metagenomic samples and information, with web services that offer functions of sample comparison, searching and mining metagenomic information for microbial community researches.

All resources in Meta-Mesh can be divided into 4 layers—just as the Mesh-style-shortcuts in the home page (Figure 1). The first layer is the Data Source layer, which integrates a range of annotated and selected data from other resources with associated information into the database. The database is updated and distributed regularly.

In the second layer, a work center with analysis tools including sample structure analysis, comparison, search and other functions is offered for large data utilization and other microbial community researches. In addition Meta-Mesh has complete user management system that allows users to upload and analysis their own data safely and securely, as well as to share data with others.

All datasets and information can be downloaded from the third layer—Facilities. Furthermore, web-based APIs (Application Programming Interface) are supplied for plug-in and extension development. In the last layer-Documents, as its name, all related documents are supplied.

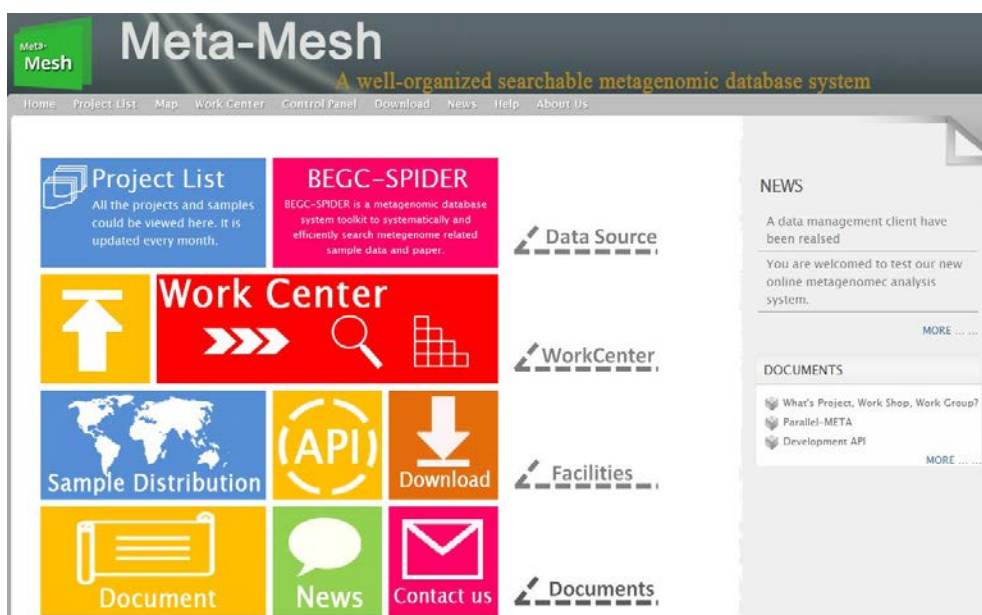


Figure 1. The home website of Meta-Mesh with mesh-style-shortcuts.

## 2 DATABASE COLLECTION AND INDEXING

### 2.1 Database collection and selection

Meta-Mesh database collected sequences and data from a range of resources by automatic information collection tool named "BEGC-SPIDER" (<http://autoupdate.metasee.org/>) to access publically available metagenome related projects and papers, and deposit them into the raw data buffer of Meta-Mesh database for annotation with their original information and meta-data that indicates the detailed basic information of metagenomic sample including project name, sample collection location, date, source tag, and sequencing methods. Annotations are added in a highly structured and unified format using controlled vocabularies for database consistency and access simplification.

As datasets and annotations in the raw data buffer are collected by unsupervised operation, all data need manual pre-selection and pre-filtering for high quality annotation and to reduce the redundancy. After the pre-selection and pre-filter, metagenomic data need to be analyzed as regard to taxonomy and phylogeny by Parallel-META for quality control of the sequences. For the aim of well-annotation, only sequence data with more than 300 rRNA sequences and 75% mapping rate could be kept and added in the database.

## **2.2 Database Indexing**

All metagenomic sample data in Meta-Mesh database are organized into projects with all annotations based on their source, and maintained by Meta-Storms which enables the efficient sample indexing for fast look up. Meta-Mesh constructs a hierarchical structure index based on the taxonomical structure of all metagenomic samples, and can also divide them into two groups of “environmental microbial communities” and “human microbial communities” for high speed search and easy browsing.

## **2.3 Database update**

Currently Meta-Mesh has collected more than 12,000 raw data of microbial community samples, and accumulated more than 7000 well-annotated high quality samples from 209 different projects in the database. The database is not a fixed one, but has update strategies. As the “BEGC-SPIDER” is updated weekly, candidate data in the raw data buffer are renewed regularly.

# **3 PROJECTS**

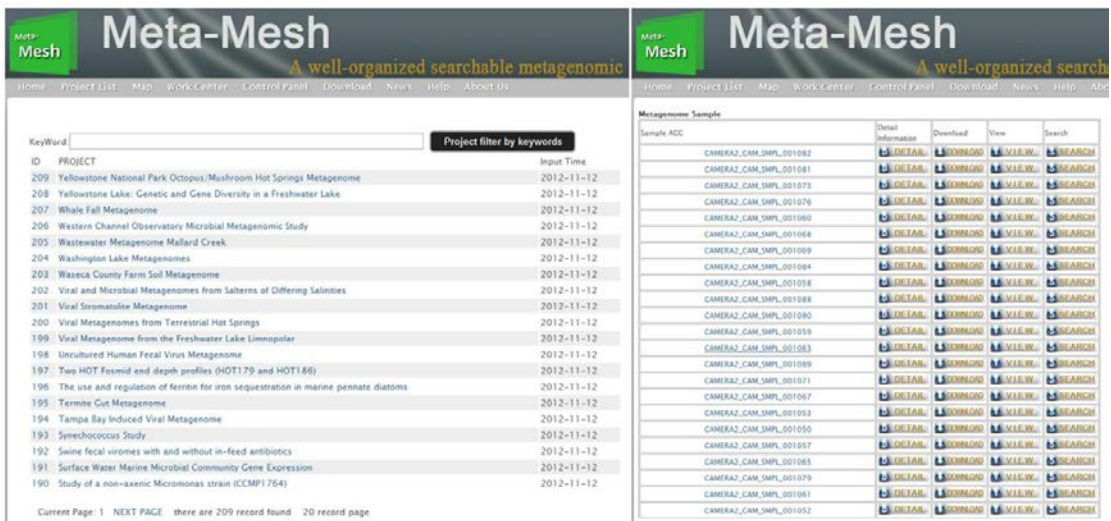
## **3.1 Project list and key-word filtering**

All open projects and samples can be freely accessed and browsed from the Meta-Mesh database website by clicking the “Project List” in the navigation bar. Listed project can be filtered by inputted key words (Figure A). In each project, annotations and meta-data, structure analysis results of samples and all related information in database are open to everyone from the metagenomic sample list (Figure B).

## **3.2 Data search**

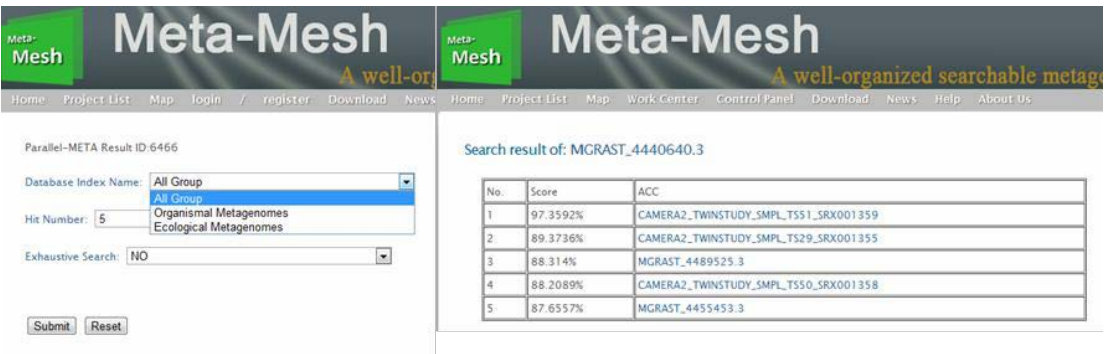
Any open sample is also searchable to find its similar ones in the database with efficient index strategy and similarity evaluation formula of Meta-Storms (Figure C). This searching can either be performed in specified group such as “environmental microbial communities” of Meta-Mesh for high speed, or among all opened samples

of Meta-Mesh for accuracy matching. Links to the target samples with high similarity to queried are provided in the results (Figure D).



(A) Project list

(B) Sample list



(C) Sample search configuration

(D) Sample search result

Figure 3. Project

## 4 USER MANAGEMENT

The registration process is opened to everyone via the Meta-Mesh website portal. Registered users' data and information are kept in the database and automatically attributed as private data which cannot be accessed by others but only the owners.

### 4.1 Register and Login

As user management is designed for private safety and security, user should register to use the work center for data analysis (Figure 4.1).

The image shows two side-by-side web forms for the Meta-Mesh application. The top navigation bar includes the Meta-Mesh logo and the tagline 'A well-organize', along with links for Home, Project List, Map, login, register, Download, News, and Help. The left form, labeled (A) Register, contains input fields for Username, Password, Confirm your password, Email, and CheckCode, followed by a 'Register' button and a 'Reset' button. The right form, labeled (B) Login, contains input fields for Username, Password, and CheckCode, followed by a 'Log in' button and a 'Reset' button. Both forms include a 'refresh check code' link next to the CheckCode field.

(A) Register

(B) Login

Figure 4.1 Register and Login

## 4.2 Work Group

Registered users can create work group(s) and invite other registered users to join as the group owner for data sharing. For registered users' private datasets, their attribute can be modified from private to be shared to any group(s) which the data owner has joined, so the shared datasets can be accessed by all users in the appointed group(s).

## 5 WORK CENTER

Work center is an integrated toolkit for metagenomic analysis, which could process the users' uploaded data by microbial community structure analysis, sample organization and management, samples comparison and searching to improve the metagenomic research experience.

### 5.1 Work Shop

Work shop is the entry of work center for users (Figure 5.1). Registered users can create work shop(s) for integrated metagenomic analysis. Work shop can be shared to the appointed Work group in which the work shop owner has joined. Then data within the shared work shop can be accessed by all users in the work group.

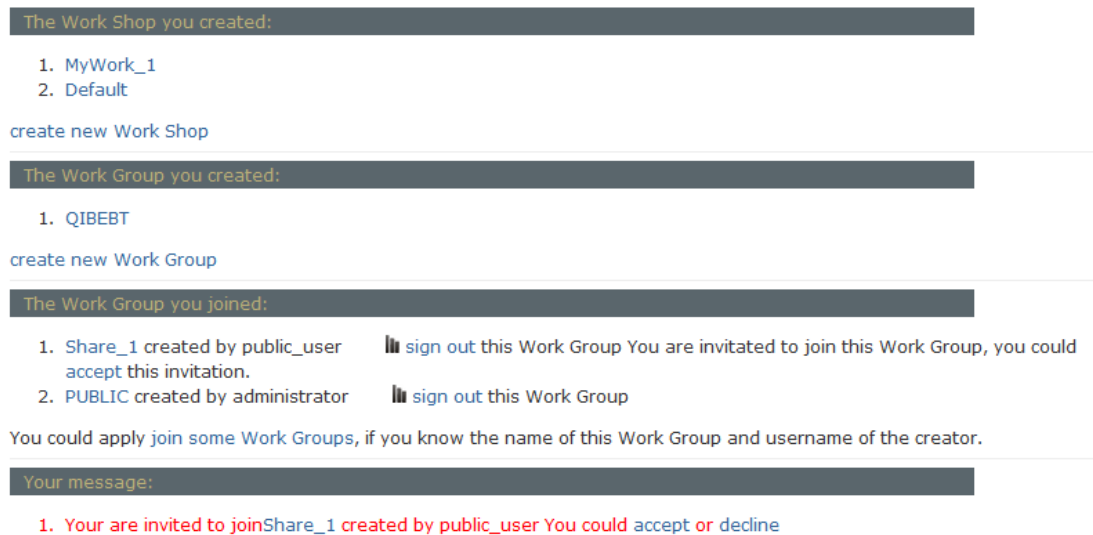


Figure 5.1. Work center with 2 workshops and 1 work group.

## 5.2 Project

In each work shop, data could be stored together (all data in a same project), or be divided into different projects (Figure 5.2) for easy management. Projects are named automatically by the system, and appointed by users at the data upload step.

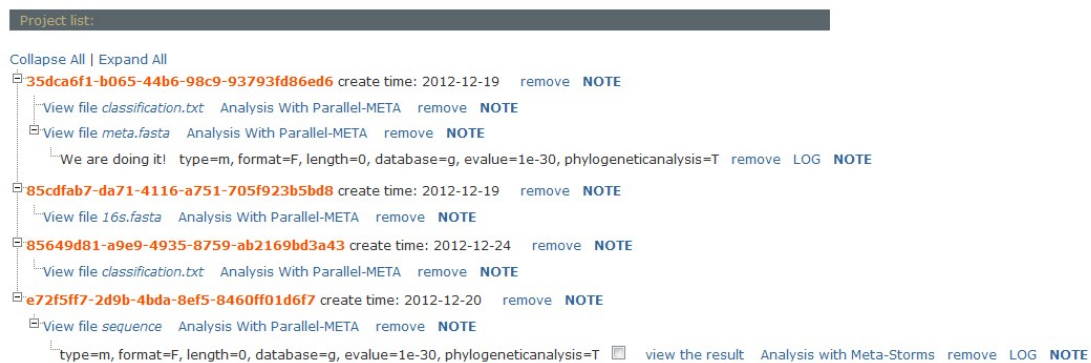


Figure 5.2 Projects. Analysis with Parallel-META means the taxonomical structure analysis, and Analysis with Meta-Storms indicates the sample search by Meta-Storms.

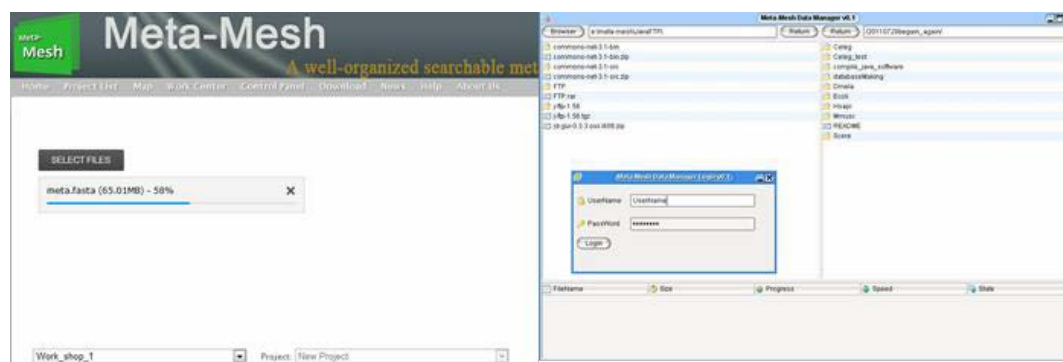
## 5.3 Data Upload

Meta-Mesh supports the data upload for registered users by direct upload from the website portal (Figure 5.3 (A))with basic operation, or by the client software

named “Meta-Mesh Data Manager”(Figure 5.3 (B))

([http://www.meta-mesh.org/news\\_content?id=2](http://www.meta-mesh.org/news_content?id=2)) that supports multi-thread high speed uploading for large size data, and data management functions designed for advanced researchers.

Before file selection, users should choose the workshop and project that the data uploaded to. Project can either be a existed one, or a new project.



(A) Upload from the website

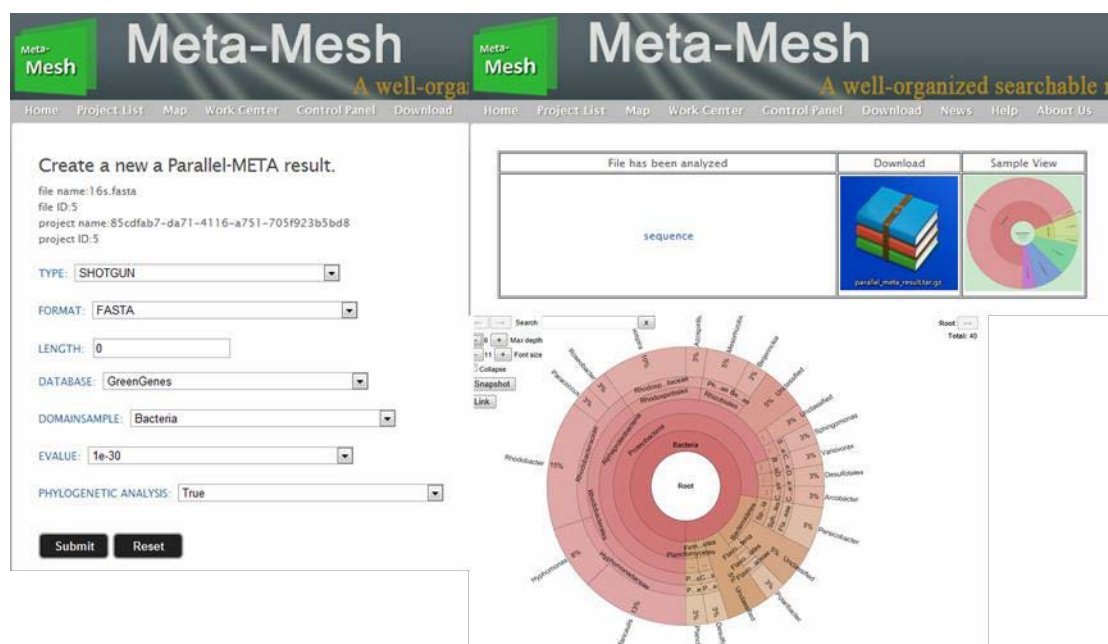
(B) Upload via the Meta-Mesh Data Manager

Figure 5.3 Data Upload

#### 5.4 Taxonomical structure analysis by Parallel-META

In metagenomic structure analysis, taxonomical and phylogenetic structure analysis (Figure 5.4) is supported to uses' uploaded metagenomic samples by Parallel-META as samples in the database with configurable parameters in Meta-Mesh. Parallel-META extracts 16S rRNA or 18S rRNA fragments by HMM algorithm, and maps these fragment sequences to reference database for structure identification, taxonomical annotation and phylogenetic analysis. Before submission please select the correct type (Shotgun or targeted rRNA), format (FASTA or FASTQ), Length (rRNA extraction length filter), reference database (GreenGene, RDP, Sliva or Oral Core), dominance (Bacteria or Eukaryota), and choose the suitable E-value. More information about Parallel-META please go to <http://www.computationalbioenergy.org/parallel-meta.html>





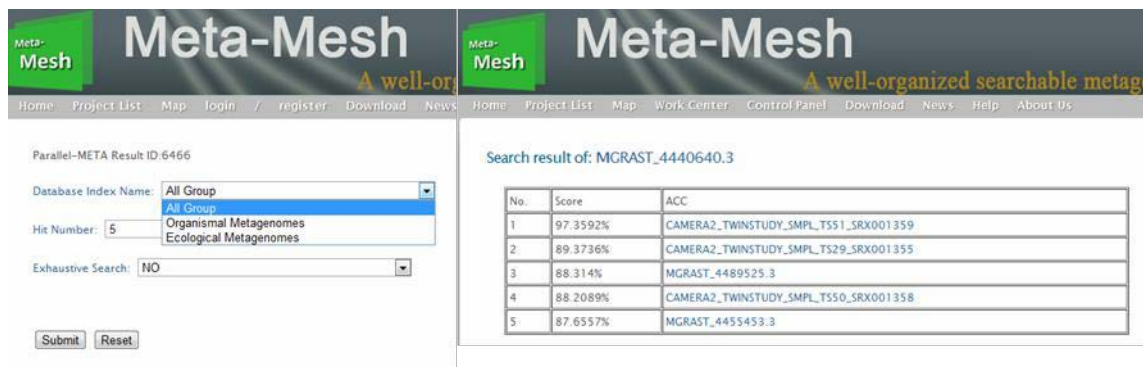
(A) Structure analysis configuration (B) Structure analysis results

Figure 5.4 Structure analysis by Parallel-META

### 5.5 Sample search by Meta-Storms

Users can also search the uploaded samples in the database of Meta-Mesh for high similarity samples (Figure 5.5). This novel function is helpful for the identification and property prediction of the users' data with the complete annotation of the samples in the search results. For sample searching, query sample pre-process by metagenomic structure analysis in work center is also required. This searching can either be performed in specified group such as "environmental microbial communities" of Meta-Mesh for high speed, or among all opened samples of Meta-Mesh for accuracy matching. Links to the matched samples with similarity values are provided in the results. More information about Meta-Storms please go to <http://www.computationalbioenergy.org/meta-storms.html>





(A) Sample search configuration (B) Sample search result

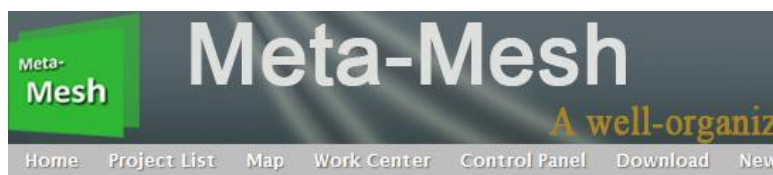
Figure 5.5 Sample search by Meta-Storms

## 5.6 Sample comparison by Meta-Storms

Two or more samples can be compared in Meta-Mesh by Meta-Storms. Result of similarity value (for single sample pair comparison) or similarity value matrix (for multi sample comparison) is supplied after the computation. This comparison requires that all related samples should be pre-process by metagenomic structure analysis in work center. More information about Meta-Storms please go to <http://www.computationalbioenergy.org/meta-storms.html>

## 6 DOWNLOAD

All data and information in the Meta-Mesh database are downloadable by clicking the “Download” in the navigation bar via the ftp (Figure 6).



## Download

1. The raw sequence data and phylogenetic analysis result
  1. CAMERA2
  2. MGRAST
2. The compiled database for similar sample search
 

Download
3. Parallel-Meta
4. META-Storms
5. MetaSee

Figure 6. Download

## 7 API

To make Meta-Mesh database could be widely used for further development, we provide several web-based APIs, which is in XML or (and) JSON format, to use the database or tools of Meta-Mesh. In addition, programmatic access to data and search results is provided via simple HTTP requests. Moreover, we have developed a web service for metagenome global distribution which shows the locations of samples in the database as a demo (“Map” in the navigation bar, Figure 7) by combined Meta-Mesh APIs and Google Map API. Such extended web service enables the research on relationship between the sample similarity and specific annotation information (location, date, temperature, ocean current, etc.) by data mining with the Meta-Mesh tools provided in work center.

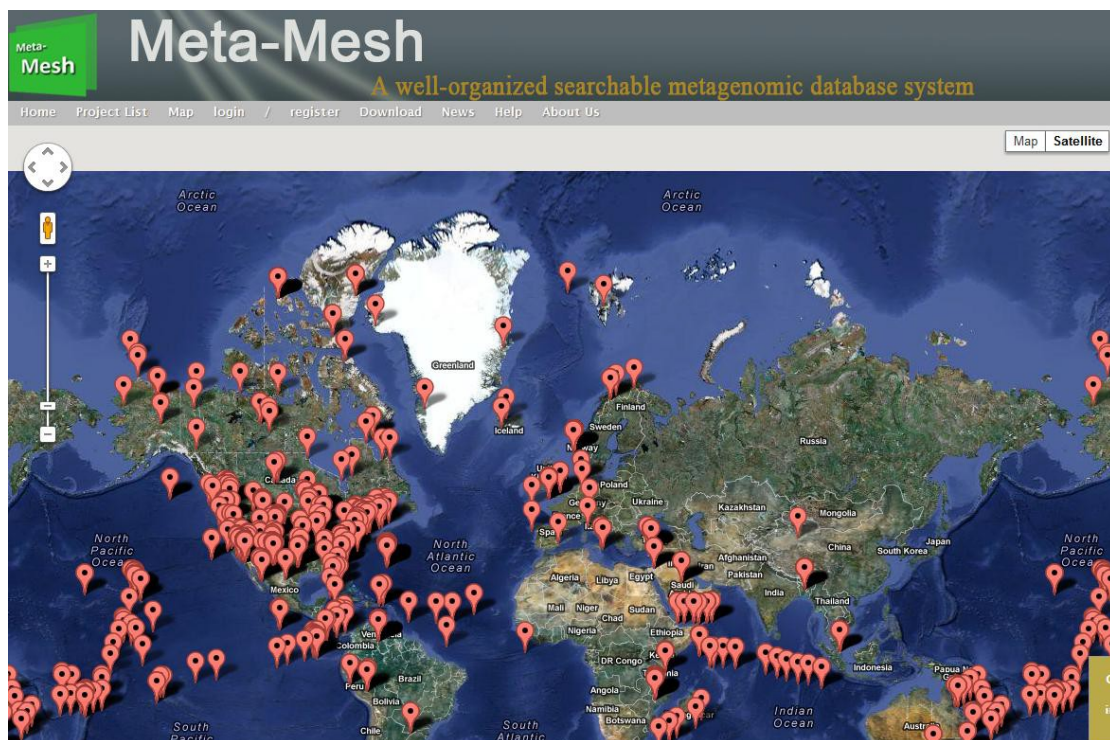


Figure 7 . Web service of metagenomic sample global distribution based on  
Meta-Mesh APIs.