

---

# **Tripal Sequence Similarity Search Documentation**

***Release 7.x-1.0***

**Sean Buehler**

**Jun 03, 2019**



---

## Contents:

---

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Administration</b>	<b>3</b>
2.1	Prerequisites . . . . .	3
2.2	Installation . . . . .	3
2.3	Configuration . . . . .	3



# CHAPTER 1

---

## Introduction

---

This module supports sequence similarity search on a Tripal website through a new dual application option. The Tripal module provides access to the speed increase available through Diamond for BLASTP/BLASTX style searches as well as traditional NCBI BLAST for BLASTN. Both applications are integrated into a single interface that provides file upload or copy/paste sequence support for the query and access to formatted databases for NCBI BLAST or Diamond. The target databases can be customized for the categories of whole genome, gene, protein, and transcriptome/unigene. The administration interface allows the admin user to set what pre-indexed databases are available (which show up in a dropdown menu). The module supports execution of the searches on a remote machine so that the search is not running directly on the limited resources typically associated with web servers.



### 2.1 Prerequisites

- **Installed on the server:**
  - Diamond, version 0.9.24
  - NCBI's BLAST+, latest
- **Drupal Modules:**
  - Tripal Remote Job

### 2.2 Installation

1. Install all prerequisites. Diamond and BLAST executables must be on the remote server where jobs will be run. 2. Download/clone the github repository to somewhere within your site's `sites/all/modules` folder 3. If you have Drush installed, run

```
drush pm-enable tripal_seq
```

If drush is not installed, you can enable the module by navigating to your Modules page and enabling it there

4. You may want to clear cache:

```
drush cc all
```

to be safe.

### 2.3 Configuration

All configuration is done through the tabs at `/admin/tripal/extension/tseq/config`.

**Available databases** Tripal Sequence Similarity Search has the ability to provide searches against databases that are provided by the website. The site administrator can make these available by adding them to the list on this page. Click “Add an existing database” to offer these. The following pieces of information are needed:

1. Name (a unique name to identify this to the user in a list)
2. Type (Protein, Genome, or Gene)
3. Category (see below)
4. Version (some version info that makes sense for the user)
5. **The full path on disk to the file. It must be available** on the remote server (See Tripal Remote Job documentation).
6. **Web location (optional) If the file is available for download** on the website or on an FTP server, provide that URL here.
7. Genes, Proteins, or Scaffolds count (optional)

From the list view, individual Database file entries can be edited or deleting by clicking the appropriate button.

### Categories

The Categories tab allows the admin to define categories that databases can be listed under. Upon installation, the “Standard” category is generated. When a user goes to submit a search query, each category will have a drop-down menu. Using categories is useful if the admin wants to delineate between various sources or methods used to generate the sequence. Whole categories can be enabled/disabled.

### Job Settings

The Job Settings tab lists a number of options for the admin to set based on the server settings and capacities.

1. **Threads - the most important setting. More threads means faster searches.** Make sure your server can handle this many threads. (Default: 1)
2. **File Expiration - how long should search results be saved on the server.** Currently not implemented, defaults to 30 days
3. **Preferred Remote Resource. If you have multiple Remote Servers defined** in Tripal Remote Job, choose which one you want to use for Diamond/BLAST jobs.
4. **BLAST Executable location. Optional - will always look in the Remote User’s \$PATH first.** You can use this setting to use a specific version.
5. **Diamond Executable location. Optional - will always look in the Remote User’s \$PATH first.** You can use this setting to use a specific version.
6. **Diamond Executable version - This setting is required to meet certain** versions so that additional functionality is provided

### Submission Defaults

The Submission Defaults tab allows the admin to set the default “Advanced options” for when a user tries to submit a search. Some reasonable defaults are provided when the module installs, but they may be tweaked.