

# Getting Started with SATCHMO

## Quick Start Guide

SATCHMO is a novel bioinformatics method that simultaneously constructs a phylogenetic tree and multiple sequence alignments (MSA) [1]. The input is a set of protein sequences in FASTA format. SATCHMO-JS (Jump Start SATCHMO) is a new variant of SATCHMO (see here <http://makana.berkeley.edu/satchmo/help#faq>) for an explanation of the differences between SATCHMO and SATCHMO-JS). This Guide will help you get started using the SATCHMO server (<http://phylogenomics.berkeley.edu/satchmo-js>), which is one of many tools from Berkeley Phylogenomics Group (<http://phylofacts.berkeley.edu/front/>). In this tutorial, we use the “globins” link from the “Sample Datasets”.

### Table of Contents

- [Input](#)
- [How to submit data](#)
- [How to know that the job has submitted successfully](#)
- [When the job is completed](#)
- [Output](#)
  - How to use the 4 options to view the output
  - 1. [Phyloscope Viewer](#)
  - 2. [Jalview Viewer](#)
  - 3. [Downloads](#)
  - 4. [Satchmo Viewer](#)



This symbol denotes a warning to users for potential pitfalls

### Input

The SATCHMO server accepts the following as input:

- a. Unaligned protein sequences in [FASTA format](#)
- b. SATCHMO **does not** accept DNA sequences
- c. The minimum number of sequences is 4
- d. The maximum number of sequences is 300
- e. The maximum size of a sequence is 2000 residues

### How to submit data



**DO NOT** cut and paste your sequences from a Microsoft Word document as this could cause problems due to hidden formatting. Use a text editing program like **Notepad** (Windows) or **TextEdit** (MacOS).

**SATCHMO**

**SATCHMO: Simultaneous Alignment and Tree Construction using Hidden Markov Models**

SATCHMO simultaneously constructs a tree and a set of multiple sequence alignments, one for each internal node of the tree. ([more](#))

[View tutorial](#)

Paste multiple sequence protein FASTA data (up to 300 sequences).

Example Input and Output

Paste FASTA:

```
>g11707910|sp|P51535.1|GLB2_NIPBR RecName: Full=Myoglobin; AltName:
  Full=glabin, body wall isoform
  RASTAATVSSMGPEQYKNSLAALSVAGLTTPKRVQKDFPKYLFTRPQVCKYFKGAESFTADDVQK
  SDKFAVQGNALLSYRLADTYDTEMTFAPVQIMNRKRGGLGQKKEFWITFKFESRRPETA
  >g11707910|sp|P51535.1|GLB2_NIPBR RecName: Full=Myoglobin; AltName:
  Full=glabin, body wall isoform
```

reCAPTCHA:

Type the two words:

Bank Worthy

Email to:

Email subject: SATCHMO server results

Go

1. Paste your sequence in [FASTA format](#) in the text field.
2. Type information in the reCAPTCHA box. This verifies that you are a person and not an automatic script. In the example, the random words are "Bank" and "worthy". These random words change for every new submission.
3. Providing an email address is optional. If provided you will get an email with a link to your final results. Or you can bookmark the submitted page and return to view the results later.
4. Click the "Go" button.

## How to know that the job has submitted successfully

**SATCHMO**

**SATCHMO v2.0 (Jumpstart Satchmo) is executing; submitted 2010-01-30 22:56:58.**

45%

This page will refresh in 11 seconds

- 1. Setup
- 2. MAFFT alignment
- 3. QuickTree
- 4. Making KERF cuts
- 5. Converting KERF output to SATCHMO input
- 6. Executing SATCHMO
- 7. Convert SATCHMO
- 8. Run RAXML
- 9. Incorporate branch lengths
- 10. Clean up
- 11. Parse SATCHMO Viewer File

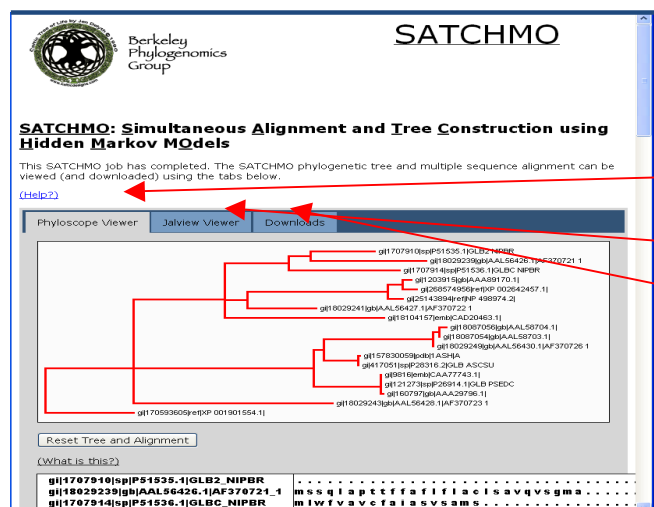
Time	Message
2010-01-30 23:28:42	Started

[Show Detailed Step Output](#)

When the job is submitted successfully you will see this screen that displays the SATCHMO server pipeline. You will be able to track:

1. The date and time of submission
2. The percentage of work completed.
3. How soon before the page is refreshed.
4. Which stage of the pipeline the job is in.

## When the job is completed



When the job is completed, you will see this page with the results of your submission and 3 ways to view them.

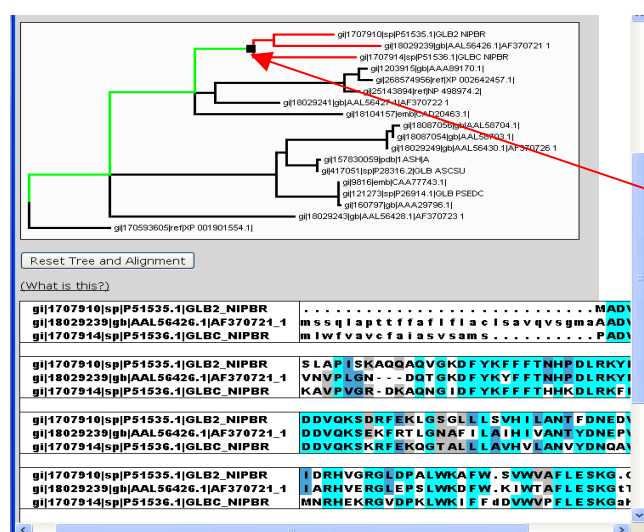
1. Phyloscope Viewer- this is the default view that you see displaying the tree and MSA
2. Jalview Viewer- an interactive Java Application for viewing the tree and MSA
3. Downloads - downloadable alignment and tree files

### Output

There are 4 different ways to view the results - depending upon your needs and personal preferences:

1. The Phyloscope Viewer where you can view the generated tree and MSAs already shown on the page.
2. The Jalview Viewer can be used if you have Java on your computer.
3. Downloads allows you to download the alignment and tree files for use in other bioinformatics programs.
4. You can download the SATCHMO combined tree and alignment file for concurrent interactive examination of the MSA and tree. This requires the SATCHMO Viewer (Windows only). The SATCHMO viewer software is available for download (The author of the free SATCHMO Viewer software is Robert Edgar. We do not maintain this software).

## 1. Phyloscope Viewer

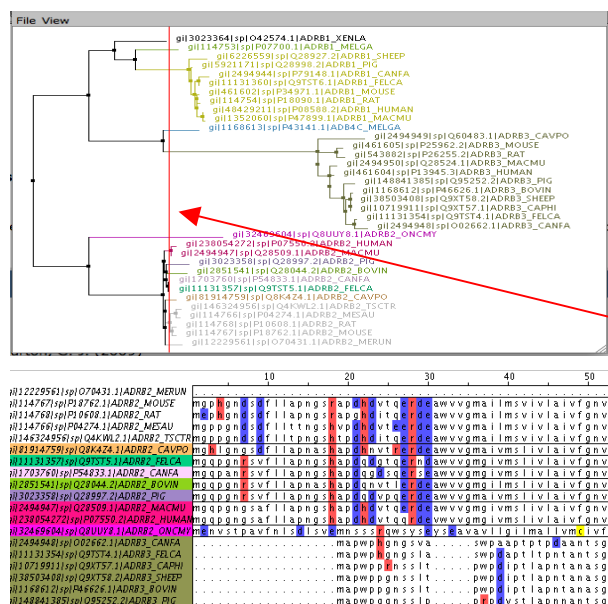


The Phyloscope Viewer is the default viewer from the Berkeley Phylogenomics Group Phyloscope software. It gives you a quick preview of the results. Some of the features are:

1. You can mouse over the nodes of the tree to highlight subtrees (the red subtree is highlighted in the example).
2. SATCHMO estimates a different MSA for each node of the tree. As you select portions of the tree, the MSA below updates to reflect your selection.

(HELPFUL HINT: For very large inputs, the Phyloscope program may take some time to process and may give an error.)

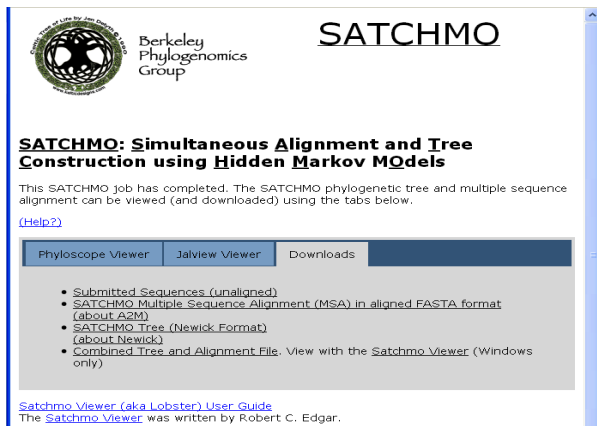
## 2. Jalview Viewer



The Jalview viewer ([www.jalview.org](http://www.jalview.org)) requires Java. Jalview will display both the MSA and the SATCHMO tree but you cannot examine the changes in alignments at internal nodes of the tree. When you click on the "View Multiple Sequence Alignment and Tree" button 2 windows with the tree and MSA are displayed. Some of the features are:

1. The red vertical bar can be moved to cut the tree in different places. The alignment viewer reflects these cuts by highlighting the sequences.

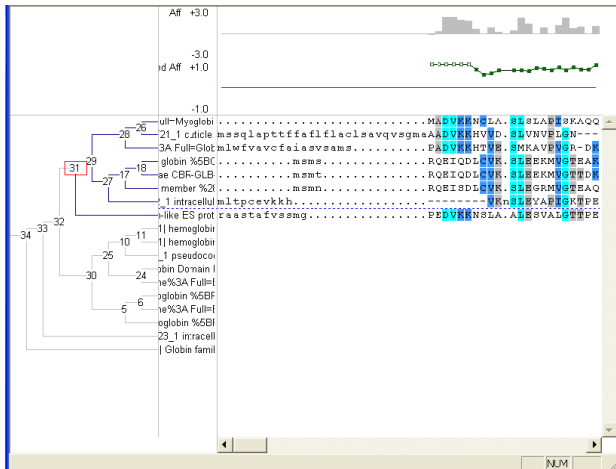
### 3. Downloads



Downloads provides the different SATCHMO outputs that you can download:

- Submitted Sequences (unaligned) - the original set of sequences you submitted
- SATCHMO Multiple Sequence Alignment (MSA) in aligned FASTA format ( [aligned FASTA format](#))
- SATCHMO Tree ([Newick Format](#)) - Newick is the standard format for phylogenetic trees. It can be used by most phylogenetic tree viewers.
- Combined tree and Alignment File - This file format is viewable using [SATCHMO viewer](#).

### 4. SATCHMO Viewer



The [SATCHMO Viewer](#) allows the user to simultaneously view a phylogenetic tree and its alignment. The user can select a node on the tree and instantly view the corresponding sub-alignment. Unfortunately, the SATCHMO Viewer is currently only available for Windows.

### Reference

1. Edgar, R., and Sjölander, K., "SATCHMO: Sequence Alignment and Tree Construction using Hidden Markov models," Bioinformatics. 2003 Jul 22; 19(11):1404-11. [Oxford University Press access](#).