

Using the Sequence Viewer

Biobike Computing

Mark Slupesky

mdslup@gmail.com

Preface

Biobike created this document to teach biologists and biology students who have little or no programming experience how to use the Sequence Viewer.

The sequence viewer is a tool for examining the nucleotide structure of DNA sequences. It allows the user to see each nucleotide and whether it is part of a gene. The user can also find sections of nucleotides using various search options.

Contents

Preface.....	2
Section 1: Accessing the Sequence Viewer.....	4
Logging into Biobike.....	4
Accessing the Sequence Viewer from the Weblistener.....	4
Accessing the Sequence Viewer from the VPL.....	4
Section 2: Viewing Sequences.....	5
By organism.....	5
By contig.....	5
Section 3: Navigating within a Sequence.....	7
Jumping to a coordinate.....	7
Jumping to a coordinate and beyond.....	7
Jumping to a range of coordinates.....	7
Section 4: Using the Search Box.....	8
Finding a subsequence.....	8
Finding a gene.....	8
Section 5: Understanding the Display.....	9
Gene names.....	9
Gene annotations.....	9
Directionality.....	9
Colors.....	9

Section 1: Accessing the Sequence Viewer

You must log in to Biobike before you can access the sequence viewer. To log in:

- 1) Open a Firefox web browser.
- 2) In the address bar, type "edwards.sdsu.edu:7003/biologin" (without the double quotes) and press Enter.
- 3) Choose a user name and type it into the *LOGIN NAME* box.
- 4) To receive email updates about Biobike, type your email address into the *EMAIL ADDRESS* box.

If you do not want to receive these email updates, leave the box blank.

- 5) Click one of the buttons to choose your login method.
 - Lisp: Access the weblistener in Lisp mode. You cannot use Biobike's specially defined functions
 - BBL: Access the weblistener in BBL mode. You can use Biobike's specially defined functions.
 - VPL: Access the Visual Programming Language.
- 6) To connect to a previous session, click the *Previous Session* button. To start a new session, click the *New Login* button.

After you log in, you can access the sequence viewer:

- From the Weblistener
- From the VPL

To access the sequence viewer from the Weblistener:

- 1) Type "(sequence-viewer)" (without the double quotes) in the text box and press Enter.
- 2) Click the Sequence Viewer Link that appears.

To access the sequence viewer from the VPL:

- 1) From the *Strings-Sequences* palette menu, choose *Sequence-of*. A Sequence-of box appears in the palette.
- 2) In the *entity* box, type 't' (without the single quotes) and hit Enter.
- 3) From the Sequence-of box's menu, click *Execute*.
The sequence viewer window appears.

Section 2: Viewing Sequences

With the Sequence Viewer, you can view the nucleotide sequence of any organism in the Biobike system.

The Sequence Viewer has three ways to choose an organism:

- Type the organism's name or nickname.
- Choose an organism from the pulldown menu of most popular organisms.
- Choose an organism from the master list.

To choose an organism by typing its name:

- 1) Type the organism name or one of its nicknames into the *Organism* box.
- 2) Click the *Go* button.

To choose an organism from the list of most popular organisms:

- 1) Move your mouse over the word *Organism*.
A list of organism names appears.
- 2) Click the organism whose sequence you want to view.
- 3) Click the *Go* button.

To choose an organism from the master list:

- 1) Click the *Select org/contig* link below the *Organism* box.
The master list of all organisms appears.
- 2) Click the organism whose sequence you want to view.
The Sequence Viewer Window reappears, showing the sequence.

You can also view a particular contiguous sequence (contig) of an organism. The Sequence Viewer has three ways to choose a contig:

- Type the contig's name.
- Choose the contig from the menu of most popular organisms.
- Choose a contig from the master list.

To choose a contig by typing its name:

- 1) Type the contig's name into the *Contig* box next to the *Organism* box.
- 2) Click the *Go* button.

To choose a contig from the list of most popular organisms:

- 1) Move your mouse over the word *Organism*.
A list of organism names appears.
- 2) Move your mouse over the organism that contains that contig you want to view.
A list of the organism's contigs appears.
- 3) Click the contig whose sequence you want to view.
- 4) Click the *Go* button.

To choose a contig from the master list:

- 1) Click the *Select org/contig* link below the *Organism* box.
The master list of all organisms appears.
- 2) Find the organism that contains the contig you want to view.
- 3) Click the contig's name.
The Sequence Viewer Window reappears, showing the sequence.

Section 3: Navigating within a Sequence

After you choose a contig or organism, the Sequence Viewer defaults to showing the first 6,000 nucleotides of that sequence. Most sequences are longer than 6,000 nucleotides, so the Sequence Viewer provides a way to navigate huge sequences:

- To move to the next 6,000 nucleotides, click the *Next* button.
- To move to the previous 6,000 nucleotides, click the *Prev* button.
- To move to the start of the sequence, click the *Start* button.
- To move to the end of the sequence, click the *End* button.

The Sequence Viewer has three options for displaying a certain range of nucleotides:

- Start at a certain coordinate.
- Start at a certain coordinate, and view a certain number of nucleotides after that coordinate.
- Select a range of coordinates.

To start viewing the sequence at a certain coordinate:

- 1) Type the coordinate into the *Go to* box.
- 2) Click the *Go* button.

To view the sequence starting at a coordinate and a certain number of nucleotides after that coordinate:

- 1) Type the beginning coordinate into the *Go to* box.
- 2) Type a plus sign (+) into the *Go to* box.
- 3) Type the number of nucleotides after the start coordinate into the *Go to* box.
- 4) Click the *Go* button.

For example, to view the 10,000 nucleotide and the 2,000 nucleotide after it, type "10000+2000" into the *Go to* box.

To view a range of coordinates:

- 1) Type the beginning coordinate into the *Go to* box.
- 2) Type a dash (-) into the *Go to* box.
- 3) Type the ending coordinate into the *Go to* box.
- 4) Click the *Go* button.

For example, to view nucleotides 10,000 through 16,000, type "10000-16000" into the *Go to* box.

Section 4: Using the Search Box

Biobike has several search options to help you find a particular subsequence within an organism's or contig's sequence. When a sequence appears onscreen, you can search for:

- A particular sequence of nucleotides. For example: "AGCTGGAA".
- A gene name. For example: "NPF0001".
- A gene annotation. For example: "Chromosome replication".

To search for a particular subsequence of nucleotides:

- 1) Display a sequence, as described in Section 2.
- 2) Make sure the *Go to* box is empty.
- 3) Type the subsequence into the *Search* box.
- 4) To search only within the sequence currently displayed on screen, click the *Display only* box.

To search over the entire sequence, leave this box unchecked.

- 5) Click the *Go* button.
The first matching subsequence is highlighted.
- 6) To go to the next match, click the *Next match* button.

To go to the previous match, click the *Prev match* button.

To search for a particular gene:

- 1) Display an organism sequence, as described in Section 2.
- 2) Make sure the *Go to* box is empty.
- 3) To search for a gene by its name, type all or part of the gene's name into the *Search* box.

To search for a gene by its annotation, type all or part of that gene's annotation into the *Search* box.

- 4) Click the *Go* button.
The Sequence Viewer displays the sequence of the gene.

Section 5: Understanding the Displaying

Sequences are very long, and certain subsequences encode for particular genes.

Sequences are displayed in rows of 60 nucleotides. If the nucleotide that begins to encode a gene appears in a line, the name of the gene being encoded appears to the right of the sequence itself.

The name of the gene is a link to the gene's frame. Click this link to learn more about the gene:

- Its type.
- A description of its function.
- A history of its annotation.

The starting and ending coordinates of a gene appear next to its name.

- A forward arrow (→) indicates that the gene is encoded in the forward direction.
- A backward arrow (←) indicates that the gene is encoded in the backward, complemented direction.

The gene's annotation appears underneath the gene's name and coordinates. Annotations do not have to contain any specific information, but might describe the gene's function, type, or discoverer.

Nucleotides are colored:

- Black - This nucleotide is part of the sequence, but does not encode a gene.
- Red or purple - This nucleotide encodes for a gene in the forward direction.
- Blue or green - This nucleotide encodes for a gene in the reversed, complemented direction
- Pink - This nucleotide encodes for two genes.