# How to use MUSI

Introduction

MUSI is a tool for uncovering multiple peptide binding specificities from sequence data (Gfeller, D., Butty, F., Wierzbicka, M., Verschueren, E., Vanhee, P., Huang, H., Ernst, A., et al. (2011). The multiple-specificity landscape of modular peptide recognition domains. Molecular systems biology). The easiest way to use MUSI is to follow the numbered steps in the main (and only) interface window. This brief guide explains each step in a little more detail. The following diagram illustrates in flow chart form what the program does:
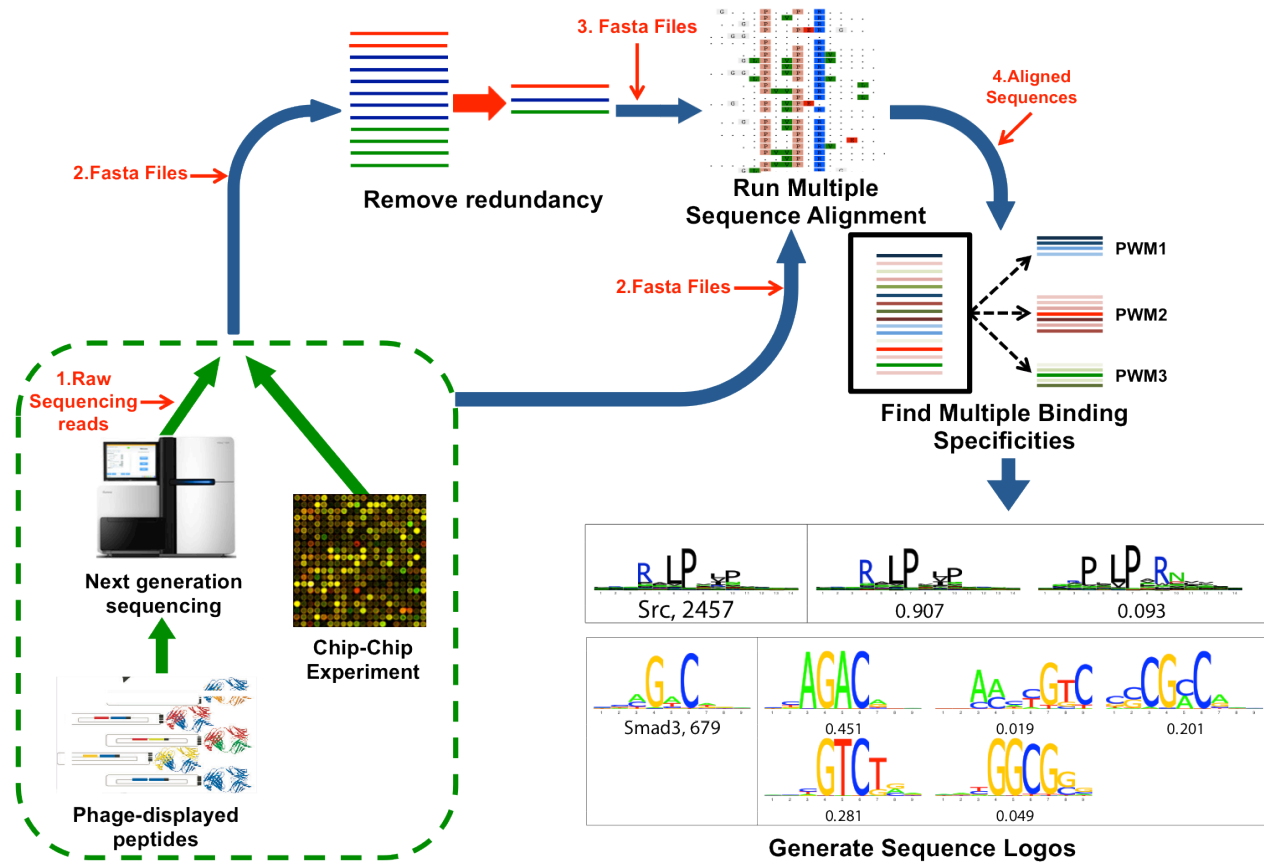
Requirements

This Mac release of MUSI requires Mac OS X 10.5 or later and Java Runtime Environment 1.5 or later. LaTeX is required if you would like to generate a PDF document containing tables of all the output sequence logos.

Steps

Load Data

You can select a single FASTA or FASTQ file containing peptide sequences, a directory containing multiple such files, or you may enter (or copy) sequence data manually into a textbox (click Load File, Load Folder, or Manual Input Window, respectively). When loading a directory, be sure that the selected directory only contains files with valid sequence data.
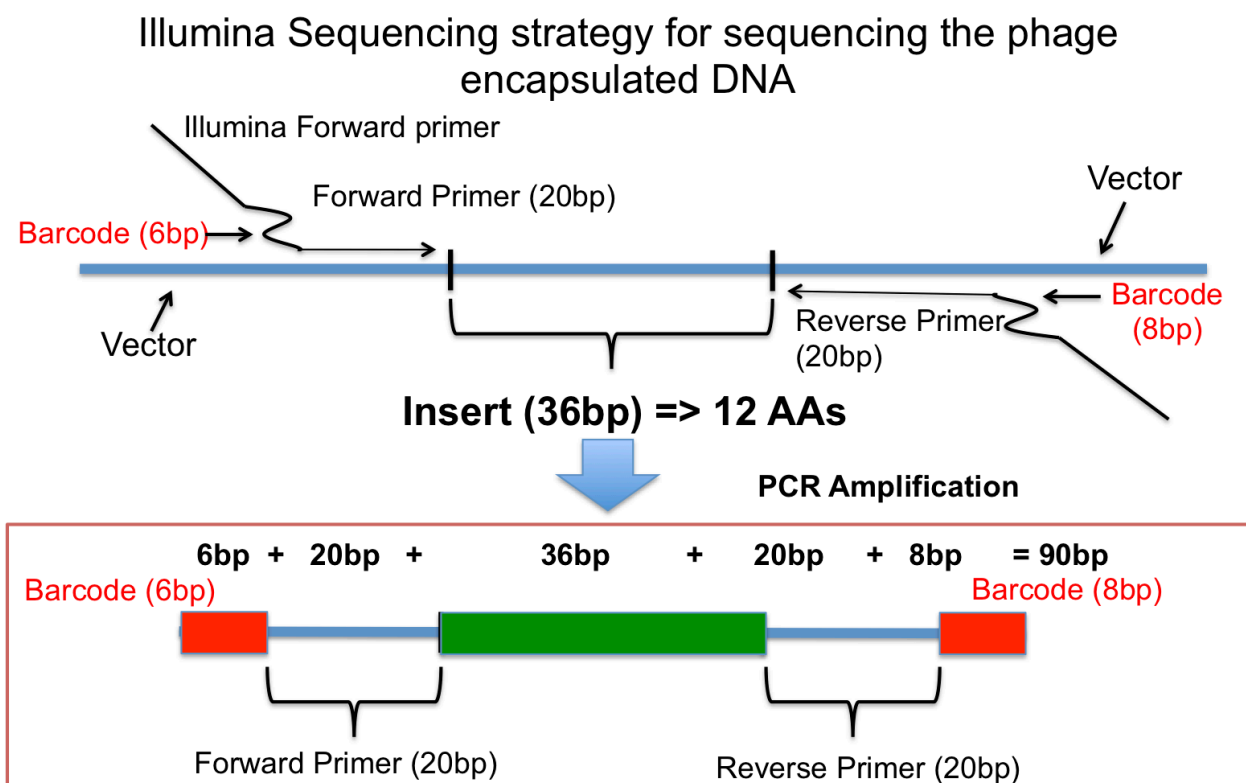
Select input format

Specify whether the data format is FASTQ or FASTA. FASTQ data will be filtered for read quality and duplicates. The FASTA option will also filter for duplicate sequences. The following option, FASTA (keep duplicates), obviously will not. All three of those options will run the data through a multiple sequence alignment program (MAFFT). The last option is for prealigned sequence data.

Set FASTQ options

If the input data is in FASTA format, skip this section (or fill it out for fun, but any values entered will be ignored).

For FASTQ files, click the FASTQ Settings button to open a window with more settings. These fields tell MUSI how to filter the data. This diagram should give you an idea of what the fields for lengths and positions refer to:

Illumina Sequencing strategy for sequencing the phage encapsulated DNA

Illumina Forward primer

Forward Primer (20bp)

Vector

Barcode (6bp)→

Vector

Reverse Primer (20bp)

Barcode (8bp)

Insert (36bp) => 12 AAs

PCR Amplification

6bp + 20bp + 36bp + 20bp + 8bp = 90bp

Barcode (6bp)

Barcode (8bp)

Forward Primer (20bp)

Reverse Primer (20bp)

The quality fields allow you to specify the minimum required read quality (as a Phred score). All sequences with reads below the threshold will be removed at the filtration stage.

Use the two buttons at the bottom of the options window to import the permissible barcodes. Any sequences containing barcodes not found in these lists are assumed to be bad data and will be removed at the filtration stage.

Specify sequence type

This is straightforward. Select DNA if the data contain DNA sequences, select RNA if they contain RNA sequences, and select Protein if they contain amino acid sequences.

Enter advanced options (optional)

Everything in this section can be safely ignored. The maximum number of PWMs field limits the number of logos the algorithm will split each group of sequences into. The strict number of PWMs field forces the algorithm to always split each group of sequences into a specific number of logos. When both of these fields are left blank, the algorithm will decide how many PWMs to generate.

Generating a PDF requires LaTeX. MUSI generates HTML tables by default and will do so even if this option is selected.

Choose output location and name

Choose a location for MUSI's output. MUSI creates a folder inside the directory you specify for its output, so there's no need to create a new subfolder. The name field is just the title of the final HTML document (the default is "Tables").

Run MUSI

We highly recommend that you click this button.

Feedback

Email musi@kimlab.org for feedback and bug reports. If reporting a problem, please include the "pipeline.log" file. If the problem is specific to certain data, it might be helpful to include at least a small portion of that dataset.