**FASTA Manipulation Tool**

**URL:** <http://bfx3.aap.jhu.edu/avasant1/finalproj/fprocess.html>

Project files are available at: /var/www/html/avasant1/finalproj

Tar file location: **/var/www/html/avasant1/finalproj.tar**

FASTA directory: /var/www/html/avasant1/finalproj/FASTA

MySQL database: avasant1

Table: transcripts

The FASTA Manipulation Tool is an application that lets the user investigate FASTA files. The user can select a FASTA file that is saved in the FASTA repository. The program parses the FASTA file and retrieves the header information comprising of the accession number, description, and sequence length. For demo purposes, a set of ten FASTA files are uploaded into the FASTA directory. These files were downloaded from NCBI [1]. The tabular display of the FASTA header resulting from the search of the FASTA file from the dropdown, has the accession number displayed with a hyperlink, which when clicked would take the user to the NCBI page for the corresponding record in a new browser tab.

Users can also search the MySQL database (DB) providing a search string. The program searches for the search string against the description column of the transcripts table. The records are fetched and displayed in the user interface in a tabular format. This search would allow the user to see the FASTA files already saved in the database. An auto-scrollbar feature is added to the section that displays the table. This would allow the table to handle the display more efficiently in case large number of records are retrieved from the database.

The ‘Free text search’ option is provided as a second level search and is enabled only after the user performs the search with the dropdown or searches the database. This search allows users to search a text within the FASTA files using a search string. The search is not case-sensitive. It retrieves the accession, description, sequence length, and sequence information from all the files that match the search criteria and displays it in tabular format. The auto- scrollbars are provided to display the long sequence information. Each row of the output table has a “Save to DB” button, on the click of which the record will be saved in the ‘transcripts’ table of the database ‘avasant1’. User-friendly messages are provided when data is saved successfully and when the record already exists in the DB, the message tells the user to pick another record to save.

The functionality of saving the record to the database was implemented using a button ‘Save to DB’ instead of a checkbox (as mentioned in the project proposal document). This was necessary to keep the interface more user-friendly. The users now would only need to click a button to save the record, instead of selecting a checkbox and then clicking on a button to save the record (in case of checkbox implementation).

To make sure that the user interface does not become crowded with too many HTML tables, the tables generated from the dropdown list search and the DB search are mutually exclusive, in the sense that when one of them is displayed, the other table is hidden. Similarly, when the ‘Free text search’ table is displayed, the other tables on the page are hidden. But, if the user wishes to search from DB or select a file from the dropdown list with the Free text search table existing on the HTML page, the user may do so. To clear all the tables from the page and start afresh, the user must reload the page.

To ensure that the program does not terminate for lack of valid user input, appropriate error-handling has been implemented. A click on ‘submit’ button before selecting a file from the dropdown, or without entering a search term in the search boxes will result in the program displaying an error message in red asking the user to make the selection or enter the search term before clicking submit button. The usage of scrollbars to the table sections avoids the user from having to scroll the page up and down, as all the contents would be able to fit within a page.

This application may be used by anyone doing genomic studies as it provides them an easier way to save FASTA information (Accession, description, sequence length, and sequence) to a table in the DB. The information in the DB can be passed on to other downstream processing tools. The activity of downloading the necessary FASTA files from NCBI will be the only overhead in terms of effort. Once the FASTA files are in the FASTA directory, the user may run this application, and save the needed files into the DB. This eliminates the need to perform multiple insert operations on the MySQL DB table.

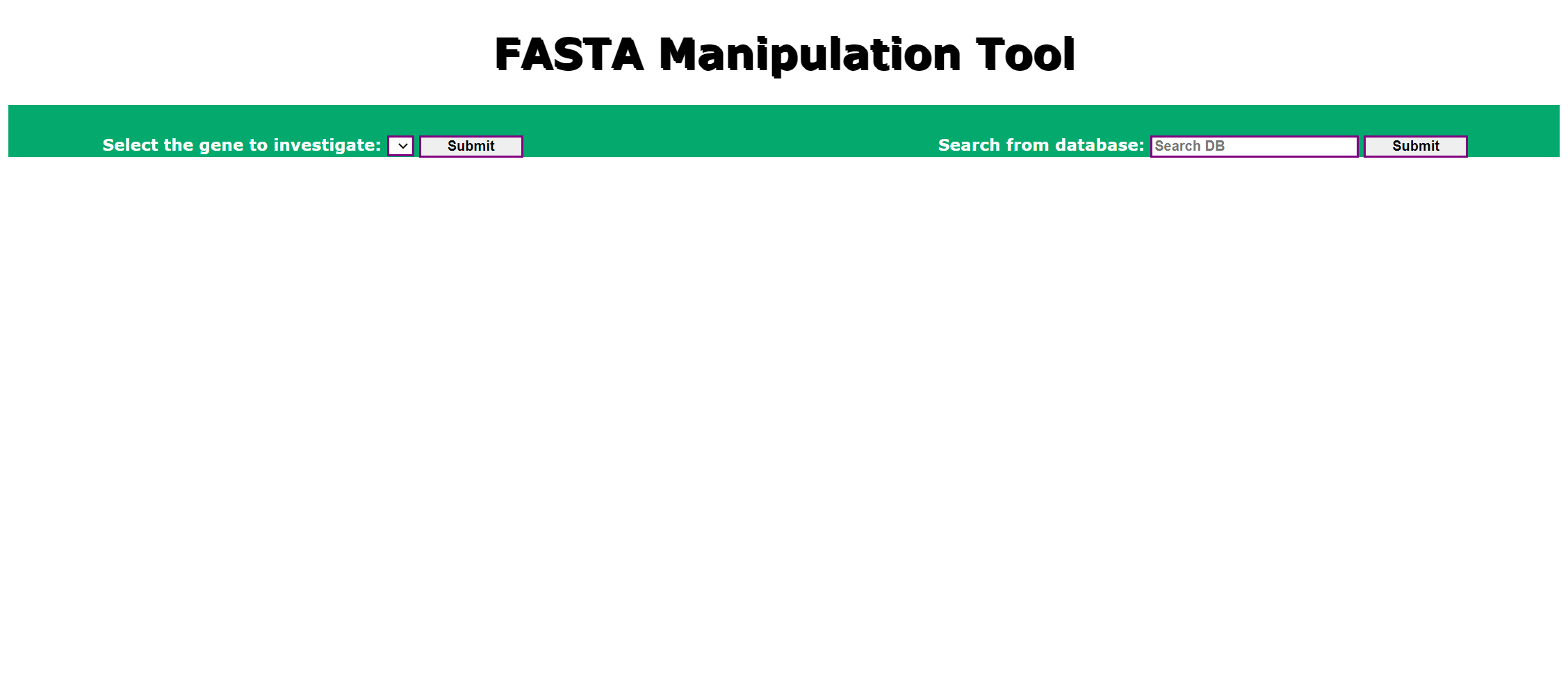
This application touches upon all the aspects covered in this course from writing CGI scripts, connecting to MySQL, AJAX calls, JSON objects, dynamically generating the output for display, and CSS. Therefore, the implementation of this application served as a skill test for me. Developing this application has given me a better grasp of programming in Python using CGI scripts. The calls to the CGI scripts from the JavaScript functions were made using AJAX calls and data was sent back from the CGI using JSON objects. One of the challenging aspects of the implementation was working with JSON and AJAX. Both were new to me and hence a little challenging at first to get things working. This project has also greatly improved my skills in CSS. I am now familiar with most of the aspects of styling basic elements of the HTML page, and dynamically changing the display features by selectively showing/hiding them.

For demo purposes, there are two records that are already saved in the database. They are:

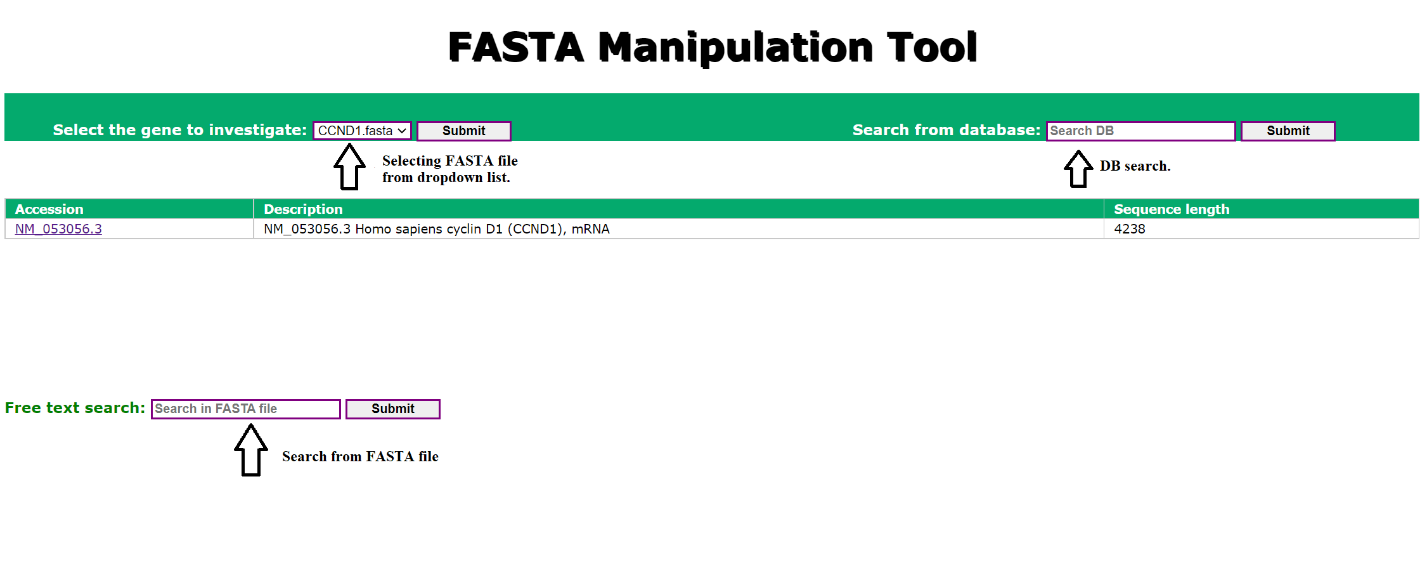
1. NM\_001382430.1- NM\_001382430.1 Homo sapiens AKT serine/threonine kinase 1 (AKT1), transcript variant 4, mRNA
2. NM\_000521.4 - NM\_000521.4 Homo sapiens hexosaminidase subunit beta (HEXB), transcript variant 1, mRNA

Below are some of the screenshots from the application:

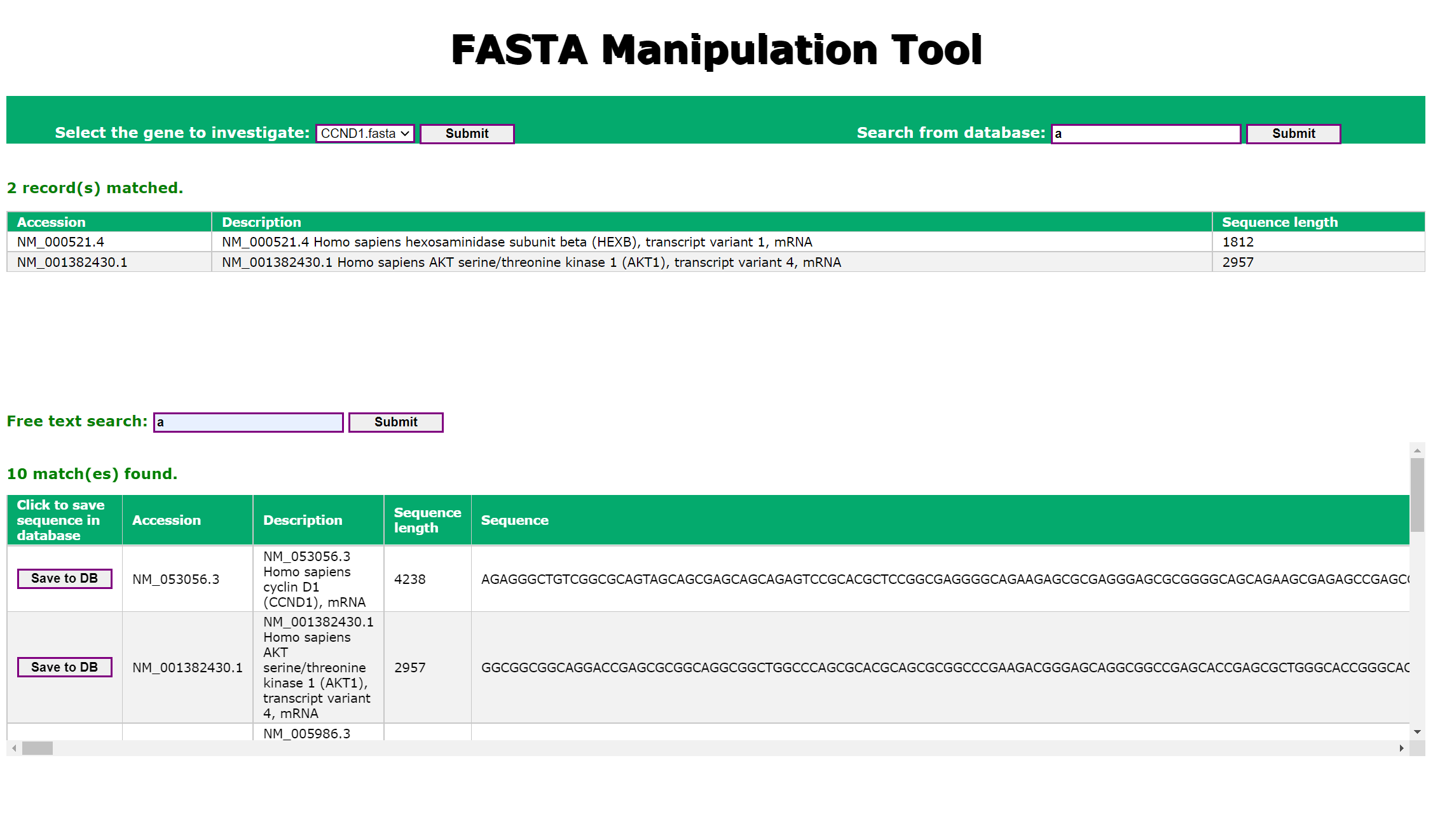
Landing page of the application:



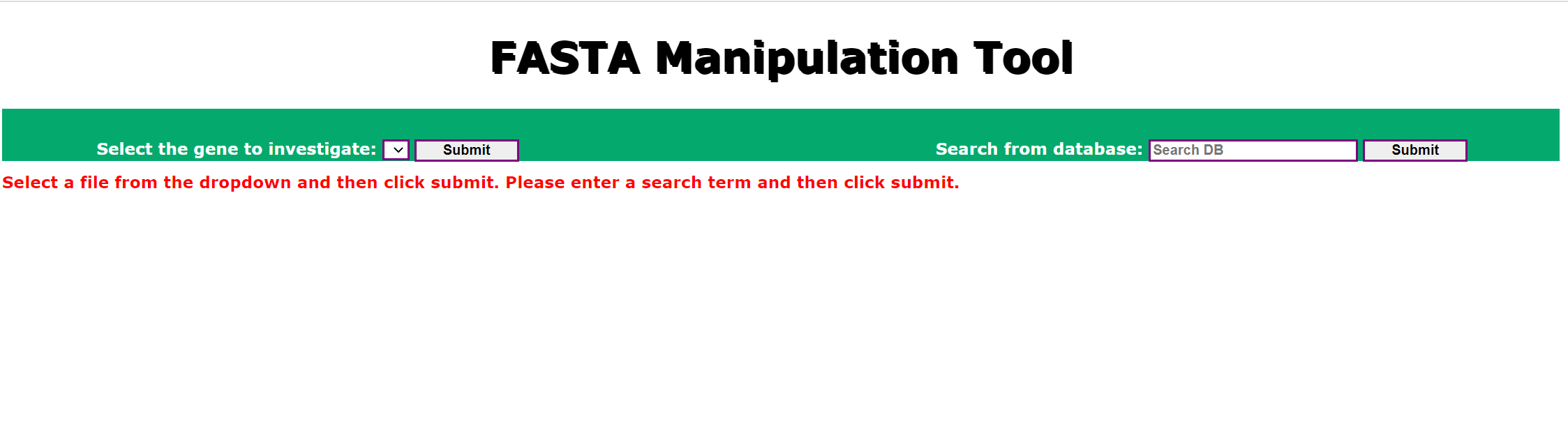
Output with FASTA file selection, Free text search is made visible.



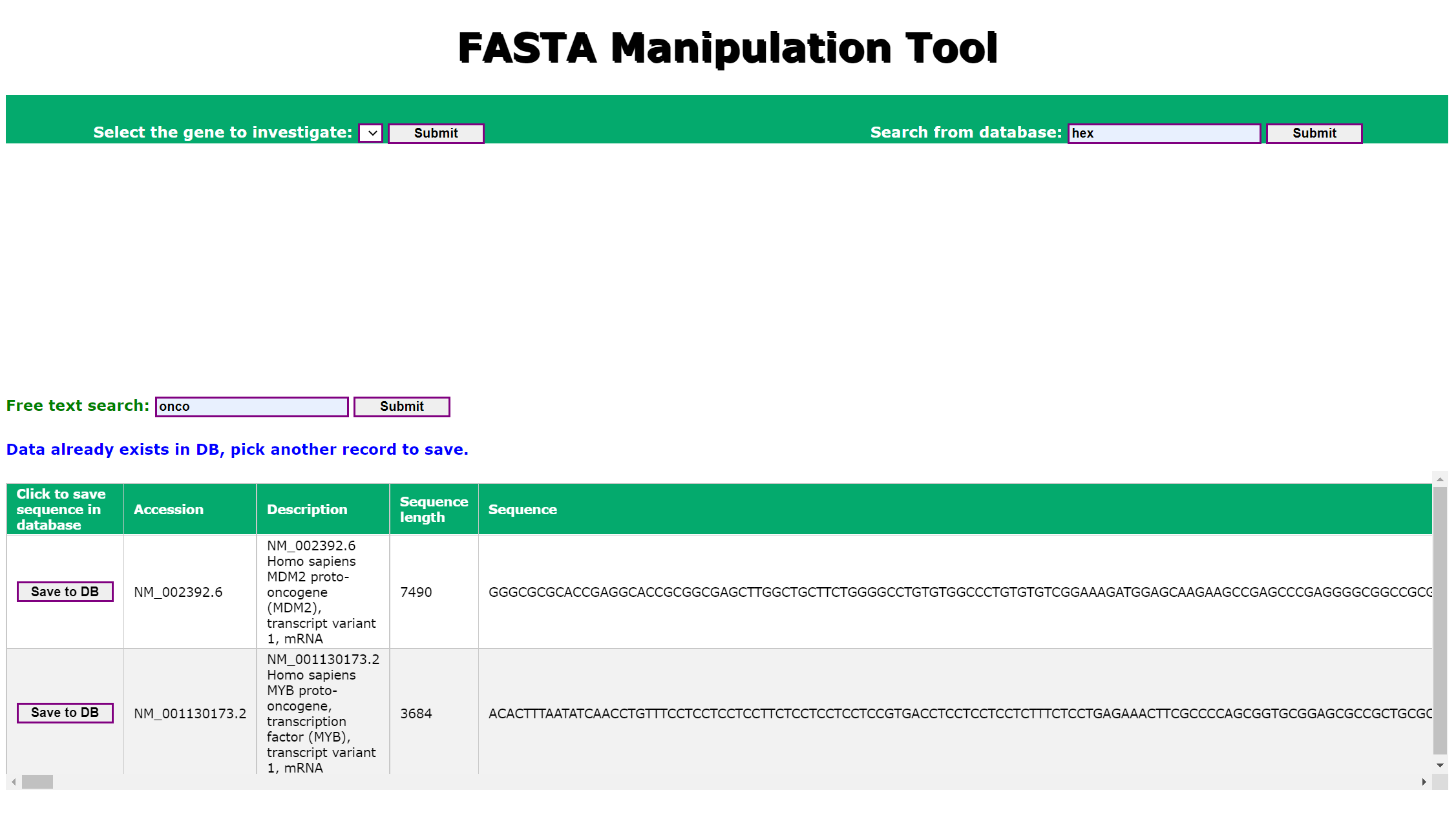
Output with DB search results as well as Free search results:



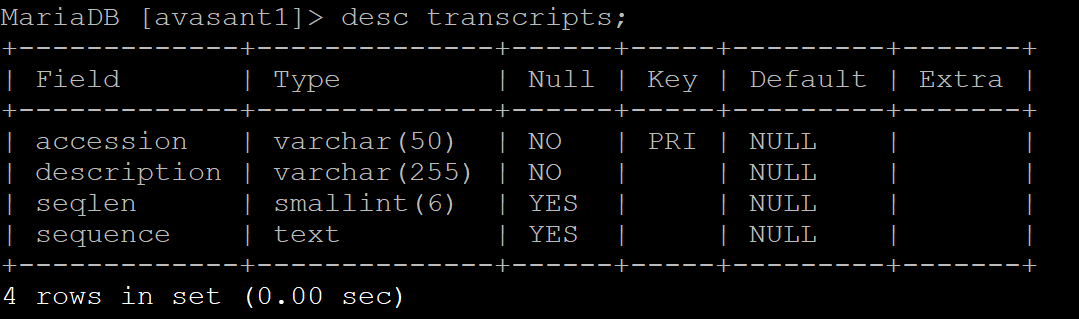
Display of error messages:



Output with message when user clicks “Save to DB” button:



Below is the description of the columns in the table transcripts:



References:

[1]. https://www.ncbi.nlm.nih.gov/search/