

Bio-Suite-T Software Documentation

Software name: Bio-Suite-T

Version: 1.0.0

TABLE OF CONTENTS

Previous steps

Software overview

- Requirements to run the software (Recommended)
- Initial configuration
 - Local server/local computer/external server
 Home computer

Functions

- Function 1. DNA property analysis Summary
- Function 2. Analysis of Protein Properties
 Summary
- * Function 3. Transcription to mRNA
 - * Summary
- * Function 4. Reverse transcription
 - ° Summary
- * Function 5. mRNA or DNA to protein translation
 - * Summary
- * Alignment of sequences in pairs

- Summary:
- Function 7. BLAST
 - Summary
- Function 8. PDB Viewer•

Summary

• Function 9. Analysis of PDBs•

Summary

• Function 10. Phylogenetic trees•

Summary

• Function 11. MOTIFS analysis•

Summary

• Function 12. Regular Expressions•

Summary

About Regular Expressions

Usage Guide

- Step 4 Open the application
- Step 4 Use any of the modules of the application

Frequently asked questions

- Approach
- Questions, inquiries and

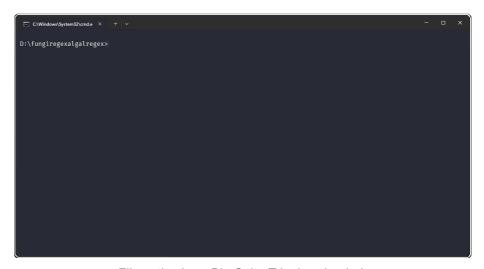
support Support

Previous steps

Before running the software, make sure you meet the prerequisites and execute the following commands:

- 1. You receive the compressed folder with the complete code, so it is necessary that you have it downloaded.
 - a. Python 3.11.
 - b. 7-Zip to decompress.
- 2. Unzip the folder.

- 3. With the console of your O.S. navigate to the path of the folder where you have downloaded, as shown in the images:
 - a. Windows.



File path where Bio-Suite-T is downloaded.

a. MacOS.

```
fungiregexalgalregex — -zsh — 80x24

Last login: Mon Nov 20 11:52:34 on ttys000
gustavodeavila@MacBook-Air-de-Gus ~ % cd downloads
gustavodeavila@MacBook-Air-de-Gus downloads % cd fungiregexalgalregex
gustavodeavila@MacBook-Air-de-Gus fungiregexalgalregex %

gustavodeavila@MacBook-Air-de-Gus fungiregexalgalregex %
```

File path where Bio-Suite-T is downloaded.

- 4. Once inside the folder type the following commands to activate the virtual environment:
 - a. Windows:

```
1  cd env
2  cd Scripts
3  activate
4  cd ...
5  cd ...
```

a. MacOS

```
source env/scripts/activate
```

5. It is required to install all the necessary dependencies for the project to work, so run the following command:

```
1 pip install -r requirements.txt
```

6. On MacOS you have experienced dificulties in the installation process associated with incompatibilities with certain packages, if this happens install each package showing the error separately using the following syntax:

```
1 pip install "package_name
```

- 7. If the problem persists try to skip the installation of the package by deleting it from the requirements.txt file. If you still have problems at this stage, you can contact me directly at the following email: \(\subseteq\text{victor.terron@cimat.mx}\)
- 8. After installing the packages you can run the program using the following command:

```
1 python .manage.py runserver
```

9. The software should show you the following output on the console:

```
November 20, 2023 - 13:39:13
Django version 4.1.13, using settings 'biosuite.settings'
Starting development server at http://127.0.0.1:8000/
Quit the server with CTRL-BREAK.
```

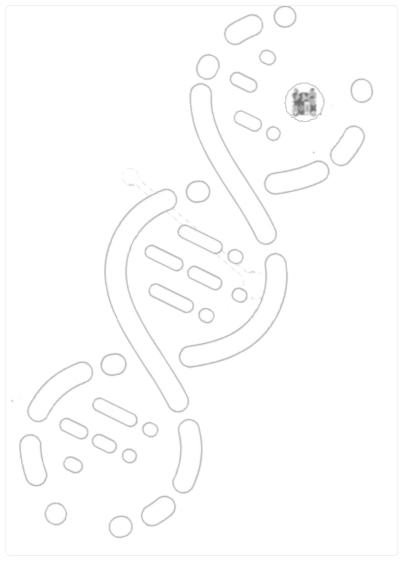
Console output.

10. Afterwards you can use the application by accessing the address from your browser:

```
1 127.0.0.1:8000
```

After executing the commands you are ready to run Bio-Suite-T.

Software overview



Bio-Suite-T logo.

This tool is a web application planned to gather a series of tools for the analysis of proteomic and genomic sequences, mainly focused on Algal and Fungi, however, it is not limited only to these, but it is able to carry out the analysis of different sequences, it has 16 tools among which are:

- 1. Analysis of DNA or nucleotide sequence properties.
- 2. Analysis of protein sequence properties.
- 3. Sequence transcription.
- 4. Reverse transcription of sequences.
- 5. Translation of sequences.
- 6. Sequence alignment.
- 7. BLAST.
- 8. PDB file viewer.
- 9. PDB file analyzer.
- 10. Calculation and generation of filogenetic trees.

- 11. MOTIFS sequence analysis.
- 12. Search for conserved regions at the genomic and proteomic level of Algal and Fungi by regular expressions.
- 13. Fungi protein database information viewer (you need to have installed MongoDB, MongoAtlas and load the databases available in JGI or NCBI, among others).
- 14. Fungi genome database information viewer (you need to have installed MongoDB, MongoAtlas and load the databases available in JGI or NCBI, among others).
- 15. Algal protein database information viewer (you need to have installed MongoDB, MongoAtlas and load the databases available in JGI or NCBI, among others).
- 16. Algal genome database information viewer (you need to have installed MongoDB, MongoAtlas and load the databases available in JGI or NCBI, among others).

This version uses works with DJANGO as a development framework for front and back end development together with JS.

Requirements to run the software (Recommended) «



The software has been developed and tested on a computer with the following features:

- 8 GB RAM
- Cent OS 7, Windows 10 and MAC OS as operating systems.
 - O Note: MAC OS has incompatibilities with the BLAST module, because therefore you may not be able to run this module from this operating system.
- AMD Ryzen 5 5600H processor. About

the software features:

- Python 3.11.
- MongoDBCompass.
- Chrome as recommended browser.

The project contains all the necessary elements to be deployed on a server if required, to start the application the following command must be executed.

python .manage.py runserver #from the root folder of the project

Initial configuration

As mentioned, this software can deploy this application either locally or on a server if necessary, this is up to the user of the application and their needs depending on the computational resources available.

Take into consideration that if multiple users connect at the same time, tasks will be interrupted one to another, therefore the application is for use by one user at a time, if the other user uploads tasks to the application while another user is executing other tasks, priority will be given to the last user's tasks and the previous user's tasks will be deleted.

Local Server/Local Computer/External Server

In case you deploy the application using a local server (which can be another computer). Follow the instructions below:

- 1. Obtain the IP address of the computer.
 - a. For Windows follow the instructions below:
 - i. https://support.microsoft.com/en-us/windows/find-your-ip-address-in-windows-f21a9bbc-c582-55cd-35e0-73431160a1b9
 - b. For Linux:
 - i. Ubuntu: https://help.ubuntu.com/stable/ubuntu-help/net-findip.html.en
 - ii. Cent OS / Red Hat: https://access.redhat.com/documentation/es-en/red_hat_enterprise_linux/7/html/migration_planning_guide/ch04s04s07
 - iii. For other distributions look for oficial information.
- 2. Navigate through the file system until you find the root folder containing the project.
- 3. Execute the application with its respective command.

```
1 python .manage.py runserver
```

Own computer

If you are running it locally on your computer the IP does not change, therefore you must enter http://127.0.0.1:8000/. The configuration of the firewall depends on each user and for this application if it is running locally it is not necessary to open ports to external users.

Functions

Analysis of DNA properties

In this functionality, you must enter a nucleotide sequence as input. From this, the tool will generate and show you a series of relevant DNA properties of the entered sequence.

Among these properties, you will find:

- Sequence complement.
- Inverse complementary sequence.
 - In its decimal and percentage representations.
- Distribution of amino acids present in the sequence.
- A visual and informative pie chart illustrating the distribution of (A) adenine, (C) cytosine, (G) guanine and (T) thymine in the sequence.

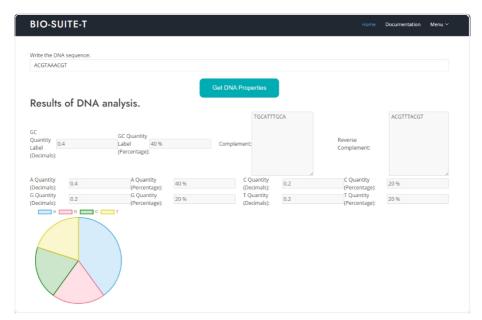
Summary

The entries are:

- User provides the DNA sequence. Process:
- Software Analyzes DNA sequence.

Output:

- Sequence complement.
- Inverse complementary sequence.
 - o In its decimal and percentage representations.
- Distribution of amino acids present in the sequence.
- A visual and informative pie chart illustrating the distribution of (A) adenine, (C) cytosine, (G) guanine and (T) thymine in the sequence.



Example output.

Function 2. Protein properties analysis

In this functionality, the application allows you to enter a protein sequence. In addition, you have the option to modificate the pH of the sequence. Based on this data, the system will automatically generate a number of important parameters related to the proteome entered, such as:

• Proteome size. •

Molecular weight.

- Aromaticity.
- Instability index. •

Isoelectric point.

- Secondary structure.
- Molecular extinction coeficient.
- Coeficient quenching with reduced cysteines.

Amount of disulfide bridges.

- Hydropaticity index (Gravy).
- Electrical charge at a specific pH. •

flexibility data.

• Distribution of elements in the sequence. •

Hydrophobicity K&D.

In addition, to facilitate the interpretation of this data, the tool includes representative graphs that provide you with a way to facilitate the visualization of the information.

Summary

The entries are:

User provides protein sequence.
 Load at pH.

Process:

Software analyzes the protein sequence.

Output:

• Proteome size. •

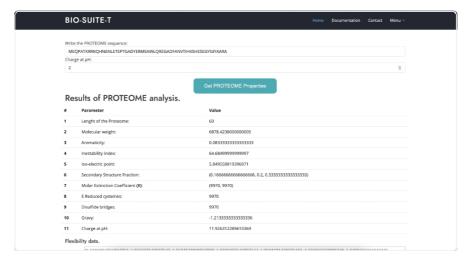
Molecular weight.

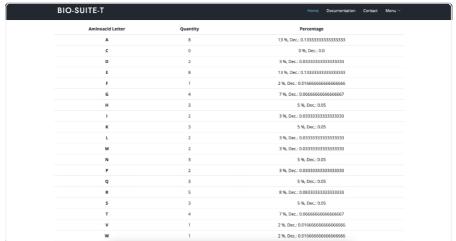
- Aromaticity.
- Instability index. •

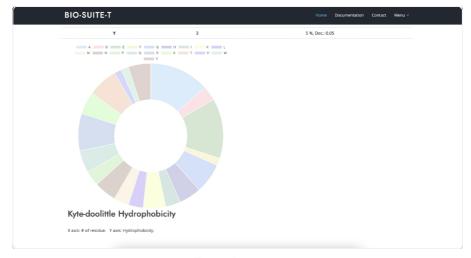
Isoelectric point.

Secondary structure.

- Molecular extinction coeficient.
- · Coeficient quenching with reduced cysteines.
- · Amount of disulfide bridges.
- Degree of hydrophobicity
- (Gravy). Electrical charge at a
- specific pH. flexibility data.
- Distribution of elements in the sequence. KB/KV
- ratios.







Example output

Function 3. Transcription to mRNA

From a nucleotide or DNA sequence, the tool is able to generate the corresponding mRNA transcript as well as provide the template DNA. In addition, it facilitates the visualization of the template DNA strands in both directions, from 5' to 3' and from 3' to 5'.

Summary

The entries are:

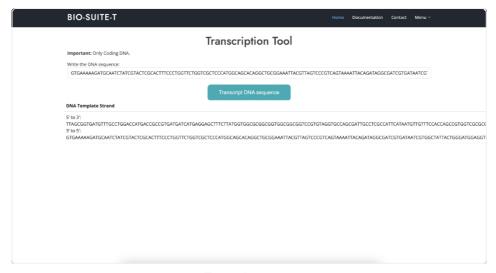
• User, provides DNA sequence or nucleotides.

Process:

• Software analyzes the sequence.

Output:

- DNA to mRNA transcription.
- DNA template strands both 5' to 3' and 3' to 5'.



Example output

Function 4. Reverse transcription

This functionality allows entering an mRNA sequence. Once entered, the system will automatically process this sequence and perform a reverse transcription, generating the corresponding DNA or nucleotide sequence.

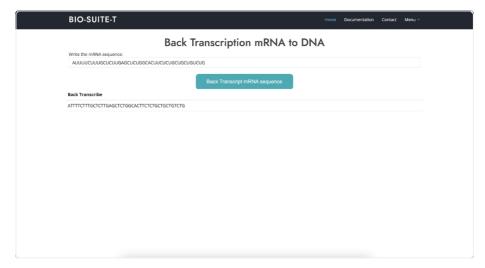
Summary

The entries are:

- User provides mRNA sequence. Process:
- Software analyzes the sequence.

Output:

Reverse transcription mRNA to DNA.



Example output

Function 5. mRNA or DNA-to-protein translation

This functionality allows you to enter an mRNA or DNA sequence. You have the option to select any of the 27 available codon tables or use the standard default table. Once you have entered the sequence and selected the codon table, the system will identify and display the sequence type. In addition, it will provide you with the corresponding mRNA to DNA translation and the details of the specific codon you chose from the table.

Summary

The entries are:

- User provides mRNA sequence.
- Select any of the 27 available codon tables or the standard.

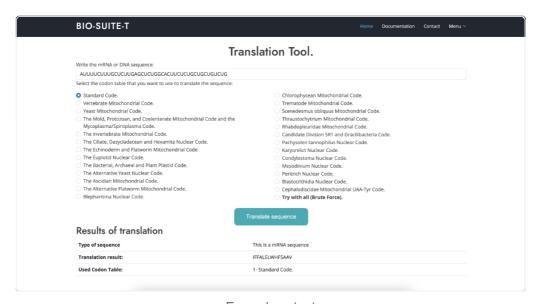
Process:

Software analyzes the sequence.

Output:

- Sequence type.
- Translation of mRNA to DNA sequence. •

Selected codon for translation.



Example output

Alignment of sequences in pairs

The pairwise sequence alignment function allows the user to compare two nucleotide or DNA sequences to identificate similarities and differences. The steps for using this function are described below:

1. Sequence Entry:

 \circ The user must enter the nucleotide or DNA sequences to be compared.

You can select the local or global alignment option according to your needs.

2. Optional Parameters:

- Percentage of Coincidence and Non-Coincidence:
 - These values are optional and are set by default. If the user wishes to customize these parameters, they can do so to suit their specific criteria.
- Penalty Matrix:
 - The penalty matrix is an optional parameter. If the user wishes to apply a custom penalty matrix, this option can be selected.

3. Results Visualization:

- o After running the alignment, the tool will provide the following results:
 - Coincidence Percentage.
 - Percentage of non-coincidence.
 - Sequence alignment.
- 4. Penalty Matrix (Optional):
 - If the penalty matrix option is selected, the tool will display the matrix used for the alignment calculation.

Summary:

Entrance:

o The user provides two nucleotide or DNA sequences for comparison.

Process:

- The software analyzes the sequences, allowing the user to choose between local or global alignment.
- Optional parameters can be set, such as match percentage, no match and a custom penalty matrix.

Outputs:

Percentage of coincidence between sequences.

Percentage of non-coincidence.

- o The visual alignment of the sequences.
- o Optional: Display of the penalty matrix used in the alignment calculation.

Function 7. BLAST

The BLAST function makes it easy for the user to scan the NCBI database in order to identify sequences similar to the sequence provided. To start the analysis, the user simply enters the nucleotide, protein or gene sequence to be studied. BLAST will perform a search of the NCBI database, identifying and presenting a list of sequences that share similarities with the one provided by the user. This result set provides a broad and detailed view of possible matches in the NCBI database.

Summary

Entrance:

• The user provides a nucleotide, protein or gene sequence.

Process:

o BLAST performs a search of the NCBI database to identificate similar sequences.

Outputs:

 BLAST presents a list of sequences found in the NCBI database that share similarities with the sequence provided by the user.

Function 8. PDB Viewer

The PDB Viewer is a tool that allows the user to browse and view files in PDB (Protein Data Bank) format. The following is a description of how it works:

- 1. PDB File Upload:
 - o The user must upload a file in PDB format to start viewing.

This file will contain detailed information about the three-dimensional structure.

- 2. Display Configuration:
 - Once the file is loaded, a three-dimensional structure can be displayed. The PDB Viewer offers several options to customize the display, including:
 - Visualization by lines, crosses, spheres or points.
 - Entanglement control for easy interpretation of complex structures. Atom labeling for fast identification.
 - Change of colors to highlight different components of the structure.
- 3. Exploration and Navigation:
 - The user can explore the loaded three-dimensional structure by zooming, rotating and panning to gain a complete understanding of the spatial arrangement of the elements.

Summary

Entrance:

o The user uploads a file in PDB format.

Process:

- Once the file is loaded, the PDB Viewer allows you to visualize the threedimensional structure.
- Customization options are provided for the display, such as the choice between lines, crosses, spheres or dots.
- o It offers tools for entanglement control, atom labeling and color changing to improve the interpretation of complex structures.

Outputs:

• The user can explore the loaded 3D structure by actions such as zoom, rotate and pan.

Function 9. PDB analysis

This function provides a detailed evaluation of the information contained in a PDB (Protein Data Bank) file. The user must load a PDB file, once loaded the tool will display different aspects of the entered structure, among the key aspects are general data, bibliographic information, structure composition, missing residues, additional information, physical properties, specific atoms and heteroatoms.

Summary

Entrance:

o The user uploads a PDB file.

Process:

• The tool performs a detailed evaluation of various aspects of the entered structure.

Outputs:

- o General Data:
 - Name of the structure. ■

Date of deposit.

- Launch date.
- Structure resolution in Ångströms.
- o Bibliographic information:
 - Key words of the structure.

■ Structure determination method. ■

Structure reference.

Reference to the journal where it was published.Authors.

- Composition of the Structure:
 - Chemical compound.
 - Origin of the structure.
- Missing Residues:
 - Presence of missing residues (Boolean).
 List of missing residues.
- Additional Information:
 - Glycosylation information.
 List of models present.
 List of chains.
 - Waste list.
 - Name and coordinates of the atoms.
- Physical Properties:
 - B-factor.
 - Distribution of elements.

Distribution of atoms.

- o Specific Atoms:
 - Detailed information about the atoms, including their name and coordinates.
- o Heteroatoms:
 - Distribution of heteroatoms in the structure.

Phylogenetic trees

The Phylogenetic Trees function allows users to analyze and visualize evolutionary relationships between biological sequences using Clustal files.

The user loads the file and the tool performs a calculation of distances between these sequences, generating a matrix that reflects the evolutionary differences. It also displays a filogenetic tree.

Summary

Entrance:

o The user uploads a file in Clustal format.

Process:

 The tool performs a calculation of distances between the loaded sequences, generating a matrix that reflects the evolutionary differences.

Outputs:

o The tool presents a filogenetic tree.

MOTIFS analysis

This function facilitates the detailed analysis of DNA sequences and the understanding of this type of sequences on biological function, helps in the identification of the most common nucleotide sequence in a set of related MOTIFS sequences. This can be useful for identificating conserved regions in promoters or other regions.

Summary

Entrance:

The user loads a MOTIFS sequence for analysis.

Process:

o The tool performs its analysis.

Outputs:

- o Consensus.
- o Degenerate consensus.
- o Reverse Cmplementary.
- Complementary Reverse Consensus.

Function 12. Regular Expressions

In the genomic and proteomic context, regular expressions are powerful tools for searching and analyzing DNA, RNA and protein sequences.

This allows the search for MOTIFS in DNA to identificate transcription factor binding sites, regulatory elements, among others. Another application that can be given is the identification of phosphorylation sites, glycosylation, among others.

Summary

Entrance:

• The user types a regular expression to search either at the genomic or proteomic level.

Process:

• The tool performs the regular expression search in the sequence.

Outputs:

List of organisms, sequences and identificators of matches.
 Number of matches within the sequence.

About Regular Expressions

If you need to use any of the special characters literally (actually, searching for a "*", for example), you must escape it by placing a backslash in front of it. For example, to search for "a" followed by "b", you would use /a*b/ - the backslash "escapes" the "*", making it literal instead of special.

For more information about regular expressions, see:

https://docs.python.org/3/library/re.html

Just remember that you don't have to write / at the beginning and at the final, just place the amino acids you want to search for.

```
Characters
   X
          The character x
    \\
           The backslash character
    \On The character with octal value On (0 \le n \le 7)
 4
           The character with octal value 0nn (0 \le n \le 7)
    \0nn
    \backslash 0mnn The character with octal value 0mnn (0 <= m <= 3, 0 <=
    <= 7)
    \xhh
            The character with hexadecimal value 0xhh
    \uhhhh The character with hexadecimal value 0xhhhh
            The tab character ('\u0009')
    \t
            The newline (line feed) character ('\u000A')
    \n
    \r
            The carriage-return character ('\u000D')
            The form-feed character ('\u000C')
    \f
            The alert (bell) character ('\u0007')
    \a
            The escape character ('\u001B')
    \e
14
```

```
15 \cx The control character corresponding to
 x 16
 17
     Character classes
     [abc] a, b, or c (simple class)
 18
      [^abc] Any character except a, b, or c (negation)
                    a through z or A through Z, inclusive (range)
     [a-zA-Z]
                   a through d, or m through p: [a-dm-p] (union)
     [a-d[m-p]]
     [a-z&&&[def]] d, e, or f (intersection)
     [a-z&&[^bc]] a through z, except for b and c: [ad-
     z] (subtraction)
 24
     [a-z&&[^m-p]] a through z, and not m through p: [a-lq-
      z] (subtraction)
     Predefined character classes
             Any character (may or may not match line terminators)
            A digit: [0-9]
      \d
            A non-digit: [^0-9]
      \D
     \s
            A whitespace character: [ \a whitespace character].
     \S
            A non-whitespace character: [^\s].
            A word character: [a-zA-Z 0-9].
     \w
     / W
            A non-word character: [^{w}].
     Boundary matchers
             The beginning of a line
             The end of a line
             \bA word boundary
           A non-word boundary
     \B
            The beginning of the input
 39 \A
            The end of the previous match
     \G
 41
     \setminus Z
             The end of the input but for the final terminator, if any
 42 \z
             The end of the
 input 43
     Greedy quantifiers
            X, once or not at all
     X?
             X, zero or more times
 46 X*
     X+
            X, one or more times
 47
            X, exactly n times
 48 X{n}
    X\{n,\} X, at least n times
 49
     X\{n,m\} X, at least n but not more than m
 times 51
 52 Reluctant quantifiers
```

```
X??
              X, once or not at
              all
     X*?
             X, zero or more
    X+?
              times X, one or more
56 X\{n\}?
             times X, exactly n
57 X\{n_{i}\}?
    times X\{n,m\}? X, at least n but not more than m times X, at least n times
     Possessive quantifiers
     X?+
            X, once or not at
             all
     X*+
            X, zero or more
    X++
            times X, one or more
64 X\{n\}+
65 X\{n,\}+ times X, exactly n
     X\{n,m\}+X, at least n but not more than m times X, at least n times
    Logical operators
            X followed by
    XY
    X \mid Y
            Either X or Y
```

User's guide 🐣

Before running the application, make sure that you meet the prerequisites and requirements for running the software, as well as whether you are going to deploy the application on your local computer or run it from an external server or a local server.

Note: Remember to consider the remarks defined for the operating system: MAC OS. To stop

the application press Ctrl+C in the console you have opened.

It is important to see the output of each console because it will indicate if there is a problem in the execution of the application.

Note: If you want to deploy the application on a server you have to edit the settings file located at: /biosuite/settings.py

Then edit the DEVELOPMENT_MODE section and configure parameters according to your configuration.

Step 1 - Unzip the application

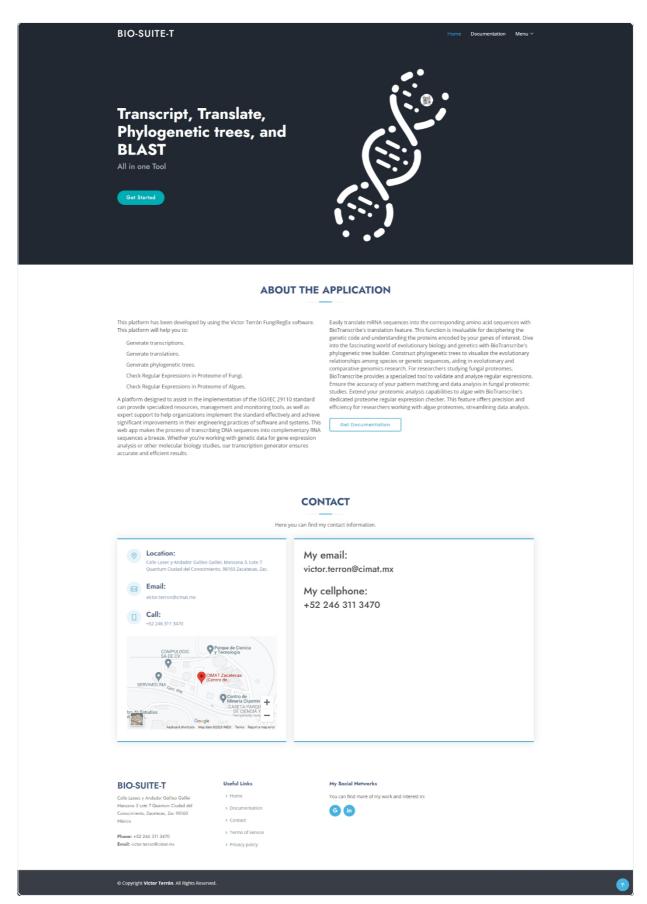
Step 2 - Install application requirements

Step 3 - Start the application

Step 4 - Open the application

After steps 1, 2 and 3 you can go to your browser, and in the address bar type 127.0.0.1:8000 or the IP address of your local or external server.

You will then see a screen like the following.



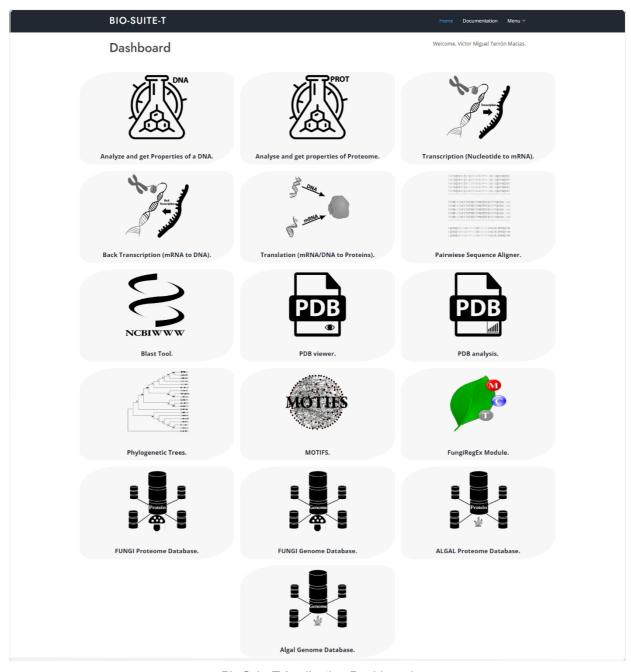
Bio-Suite-T application user interface.

Then click on Menu > Login. And enter by social account. It is a requirement that you enter from a Google account, however, it is clarified that your data are

protected in accordance with the applicable regulations in force. Likewise, the data of each user is the sole responsibility of the user, since by releasing the copy of the software, the user accepts that his/her data is his/her sole responsibility, exempting the authors of the software from any damage, loss or theft of sensitive information. The data collected are: Name and UID, among others.

Step 4 - Use one of the application modules

By logging in from your account you will be able to see the application dashboard, from which you will be able to enter any of the modules previously described in the Functions section.



Bio-Suite-T Application Dashboard

Frequently Asked Questions 🙋

Answer and document frequently asked questions here:

Approach

This application focuses on sequence analysis at different levels.

Questions, inquiries and support

For any question related to the use of the software please contact me by the following means:

✓ victor.terron@cimat.mx

Support 💬

For support or questions about the use of the software contact me at the following email:

• victor.terron@cimat.mx

