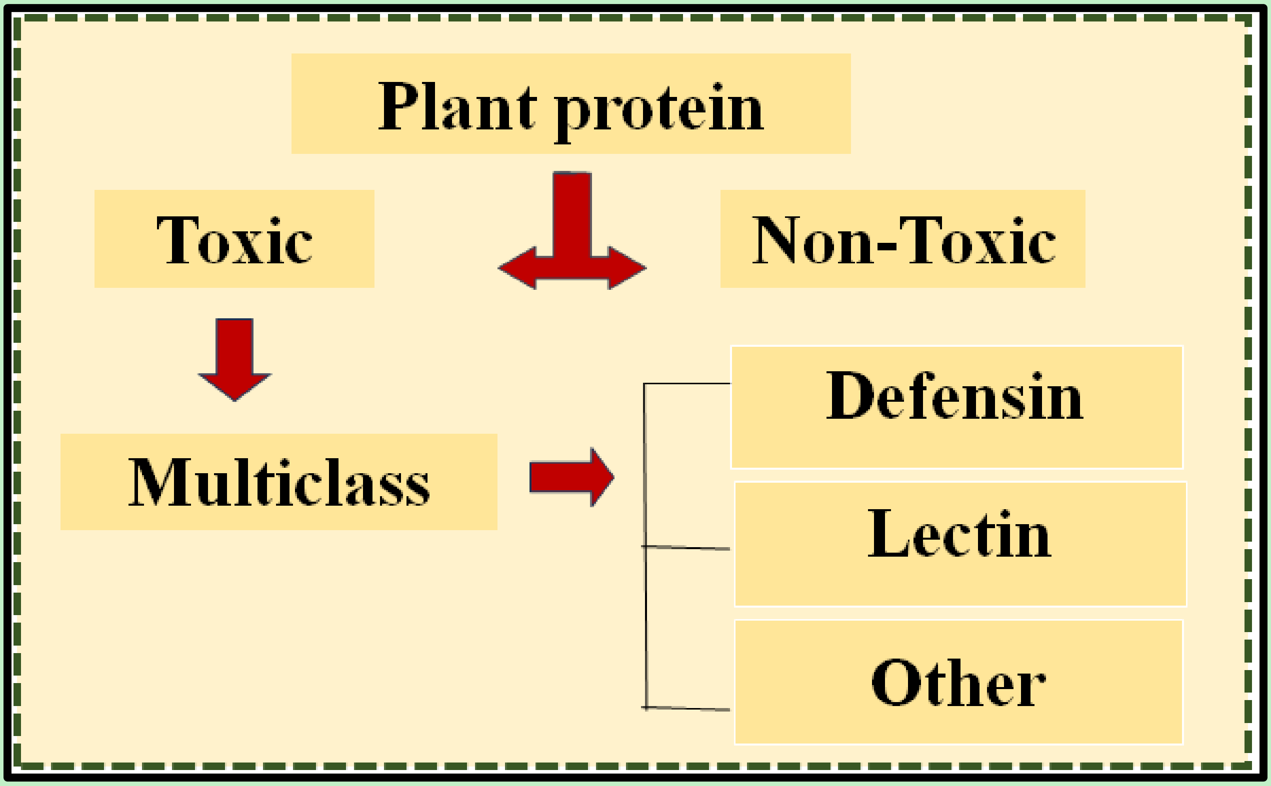
Manual

for

Toxic Protein Predictor for Plantae Prediction Server

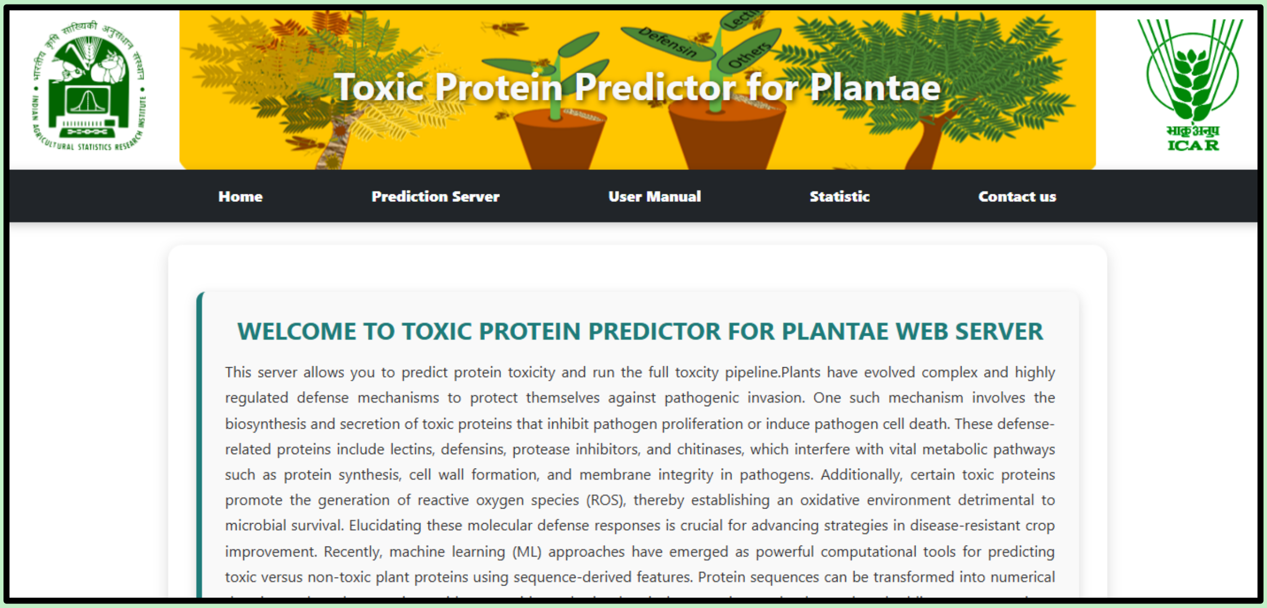


INTRODUCTION

Plants have developed the ability to produce a wide range of harmful compounds as a survival strategy under adverse environmental conditions. Among these, many plant proteins exhibit toxic properties and serve as key components in defense mechanisms against insect herbivores and microbial pathogens. The rapid advancement of high-throughput technologies has led to the accumulation of extensive plant protein sequence data; however, assigning accurate functional annotations to these proteins remains a complex task. Laboratory-based validation of protein function is both time-consuming and expensive. Moreover, conventional computational tools such as homology-driven BLAST searches are often constrained by the scope and completeness of existing databases. Machine learning offers a promising alternative by uncovering hidden patterns in data to predict toxic and non-toxic plant proteins. The **Toxic Protein Predictor for Plantae**, a machine learning-based tool designed to identify and classify toxic proteins in plant species.

Overview

This guide explains how to use the Prediction Server to get screening of potential plant toxin from your protein sequences. You only need a FASTA file — no coding or installation required.



WELCOME TO TOXIC PROTEIN PREDICTOR FOR PLANTAE WEB SERVER

This server allows you to predict protein toxicity and run the full toxcity pipeline.Plants have evolved complex and highly regulated defense mechanisms to protect themselves against pathogenic invasion. One such mechanism involves the biosynthesis and secretion of toxic proteins that inhibit pathogen proliferation or induce pathogen cell death. These defense-related proteins include lectins, defensins, protease inhibitors, and chitinases, which interfere with vital metabolic pathways such as protein synthesis, cell wall formation, and membrane integrity in pathogens. Additionally, certain toxic proteins promote the generation of reactive oxygen species (ROS), thereby establishing an oxidative environment detrimental to microbial survival. Elucidating these molecular defense responses is crucial for advancing strategies in disease-resistant crop improvement. Recently, machine learning (ML) approaches have emerged as powerful computational tools for predicting toxic versus non-toxic plant proteins using sequence-derived features. Protein sequences can be transformed into numerical

Step 1. Prepare your FASTA file

- Open any text editor (Notepad, VS Code, etc.).
- Paste your protein sequences in **FASTA format**:

  ```
  >sequence1
  MKTAYIAKQRQISFVKSHFSR
  >sequence2
  MAVPETRPNHTIYINNLNEKIKKDEL
  ```

- Save the file as `my_sequences.fasta` (or `.fa`).

Tips: - Only standard amino acids (A, R, N, D, C, Q, E, G, H, I, L, K, M, F, P, S, T, W, Y, V) are accepted. - Each sequence must start with a header line beginning with `>`.

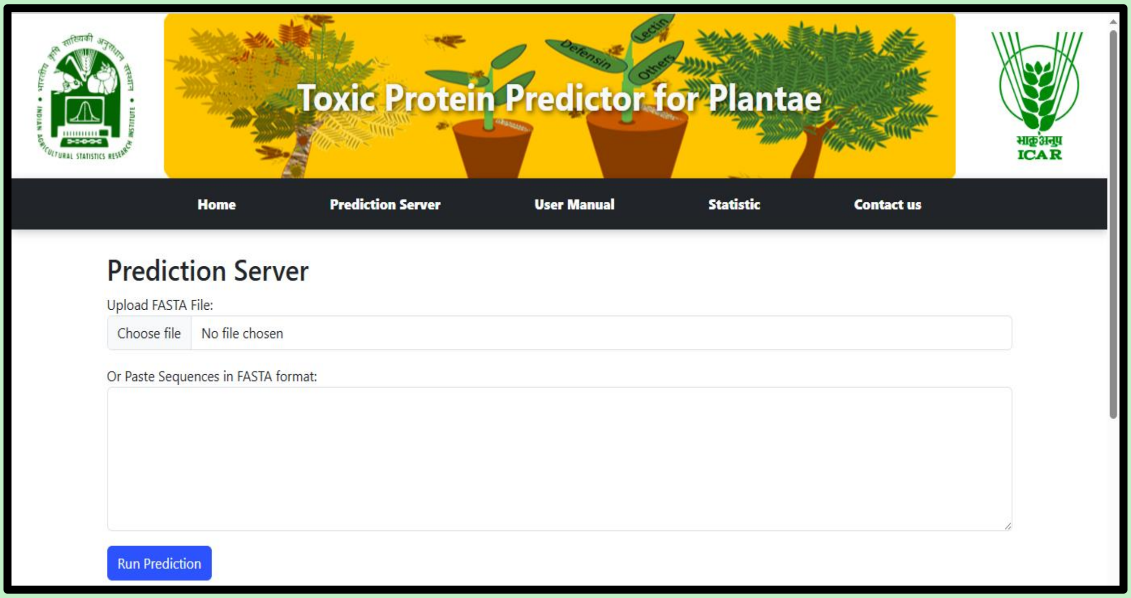
Step 2. Open the Prediction Server

1. Visit the Prediction Server website (e.g server link) highlighted in red circle.

WELCOME TO TOXIC PROTEIN PREDICTOR FOR PLANTAE WEB SERVER

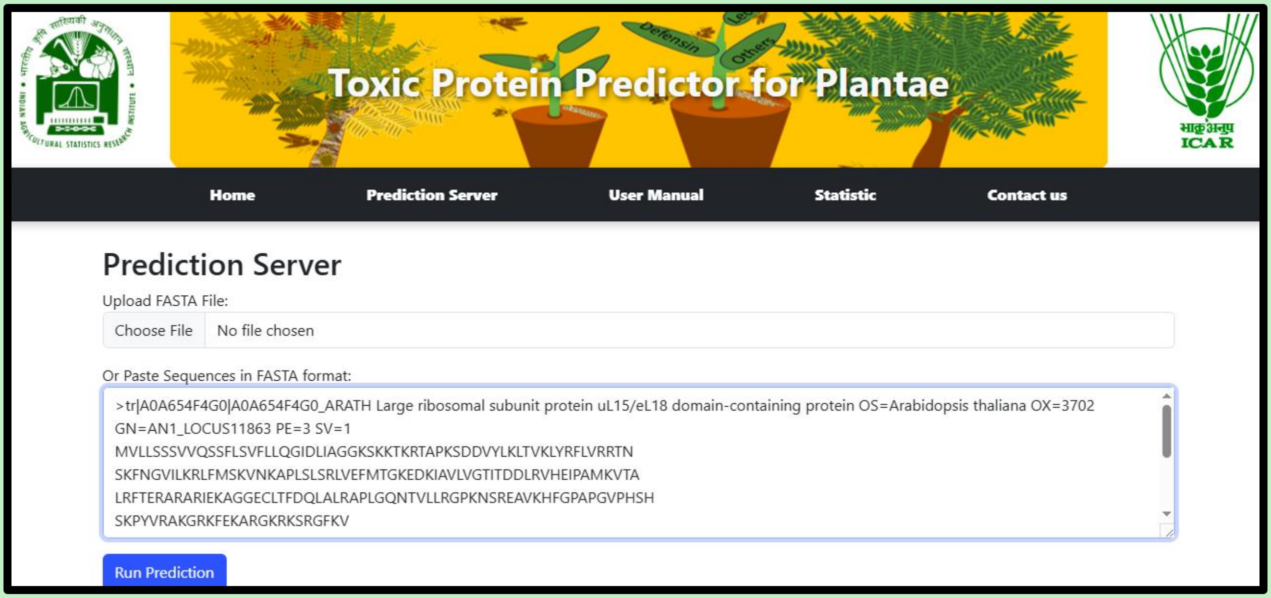
This server allows you to predict protein toxicity and run the full toxcity pipeline.Plants have evolved complex and highly regulated defense mechanisms to protect themselves against pathogenic invasion. One such mechanism involves the biosynthesis and secretion of toxic proteins that inhibit pathogen proliferation or induce pathogen cell death. These defense-related proteins include lectins, defensins, protease inhibitors, and chitinases, which interfere with vital metabolic pathways such as protein synthesis, cell wall formation, and membrane integrity in pathogens. Additionally, certain toxic proteins promote the generation of reactive oxygen species (ROS), thereby establishing an oxidative environment detrimental to

2. You’ll see a page with two options:
   - **Upload FASTA file**
   - **Paste sequence**.



Step 3. Upload your FASTA file

1. Click on **Browse button** or **Choose File**.
2. Select your saved FASTA file (`my_sequences.fasta`).
3. Click on **Predict** button.



Prediction Server

Upload FASTA File:

No file chosen

Or Paste Sequences in FASTA format:

```
>tr|A0A654F4G0|A0A654F4G0_ARATH Large ribosomal subunit protein uL15/eL18 domain-containing protein OS=Arabidopsis thaliana OX=3702
GN=AN1_LOCUS11863 PE=3 SV=1
MVLLSSSVVQSSFLSVFLLQGIDLIAGGKSKKTKRTAPKSDDVYLKLTVKLYRFLVRRTN
SKFNGVILKRLFMSKVNKAPLSLSRLVEFMTGKEDKIAVLVGTITDDLRVHEIPAMKVTA
LRFTERARARIEKAGGECLTFDQLALRAPLGQNTVLLRGPKNSREAVKHFGPAPGVPHSH
SKPYVRAKGRKFEKARGKRKSRGFKV
```

Step 4. Wait for prediction

- The server will extract features and run your sequences through the model.
- This may take a few seconds depending on file size.

You’ll then see a **results table** with columns like: - **Sequence ID** - from your FASTA header - **Predicted Class / Label** - model’s prediction.

Output (CSV):

Prediction Results (Top 20 rows)

id	TOXIC_vs_NoNTOXIC	TOXIC_CLASS
seq1	Toxic	Defensin
seq2	Non-Toxic	NaN

Step 5. Download your results

1. Once predictions appear, click **Download CSV**.
2. A file named something like `predictions_YYYYMMDD_HHMM.csv` will be saved.
3. You can open it in Excel or any spreadsheet tool.

The CSV file includes all results for each sequence.

Step 6. Common notes

- If your FASTA has errors (e.g., invalid residues or missing headers), you’ll see a message such as *“Invalid FASTA format”*.

- For large files, upload may take a bit longer — wait for completion.