# IDENTIFICATION OF TOXIC BACTERIA

Mukesh Tiwari
Astha Pandey

**Team ATGC**

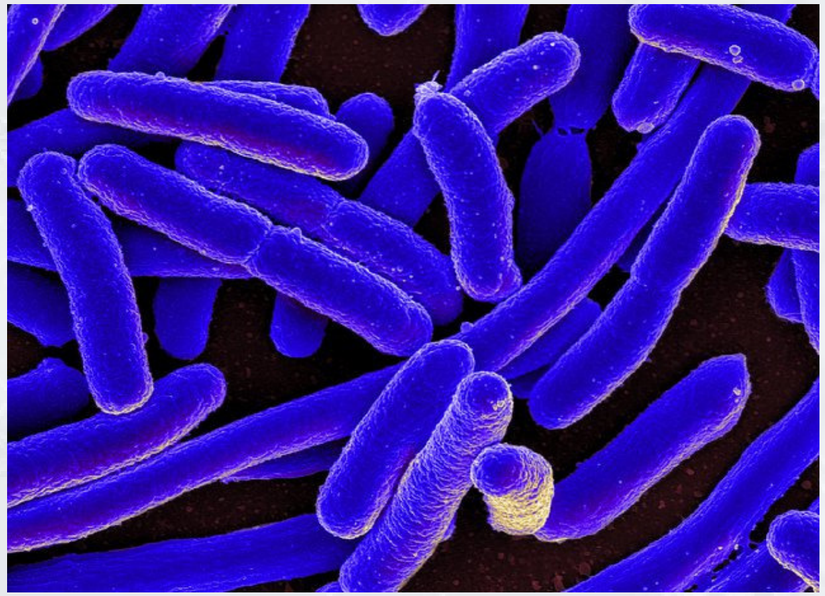
# WHAT IS THE PROBLEM?

Identification of the bacteria which is toxic to human population from the collected drinking water sample by constructing a pipeline

# BACKGROUND

Bacteria causes many diseases.

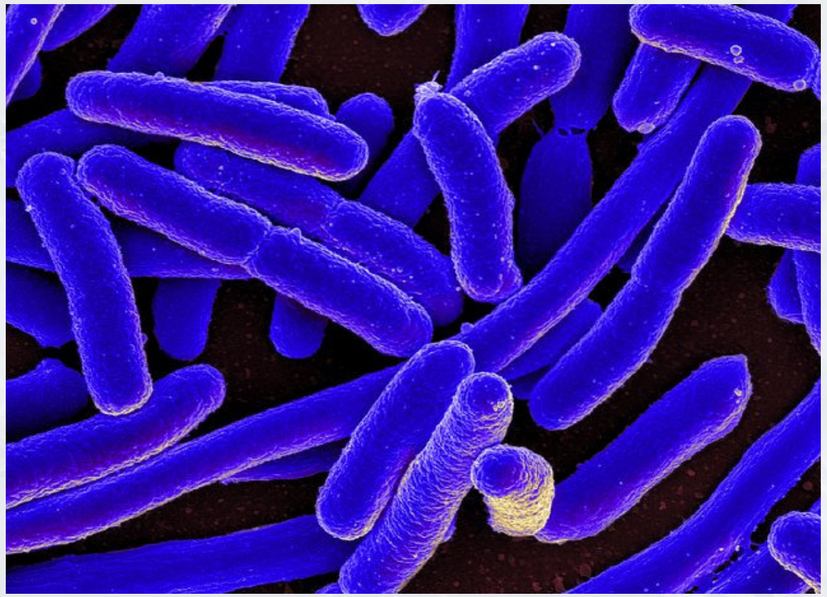
However, not all the bacterias are harmful to human.



# BACKGROUND

Bacteria causes many diseases.

However, not all the bacterias are harmful to human.



So, what makes a bacteria harmful?

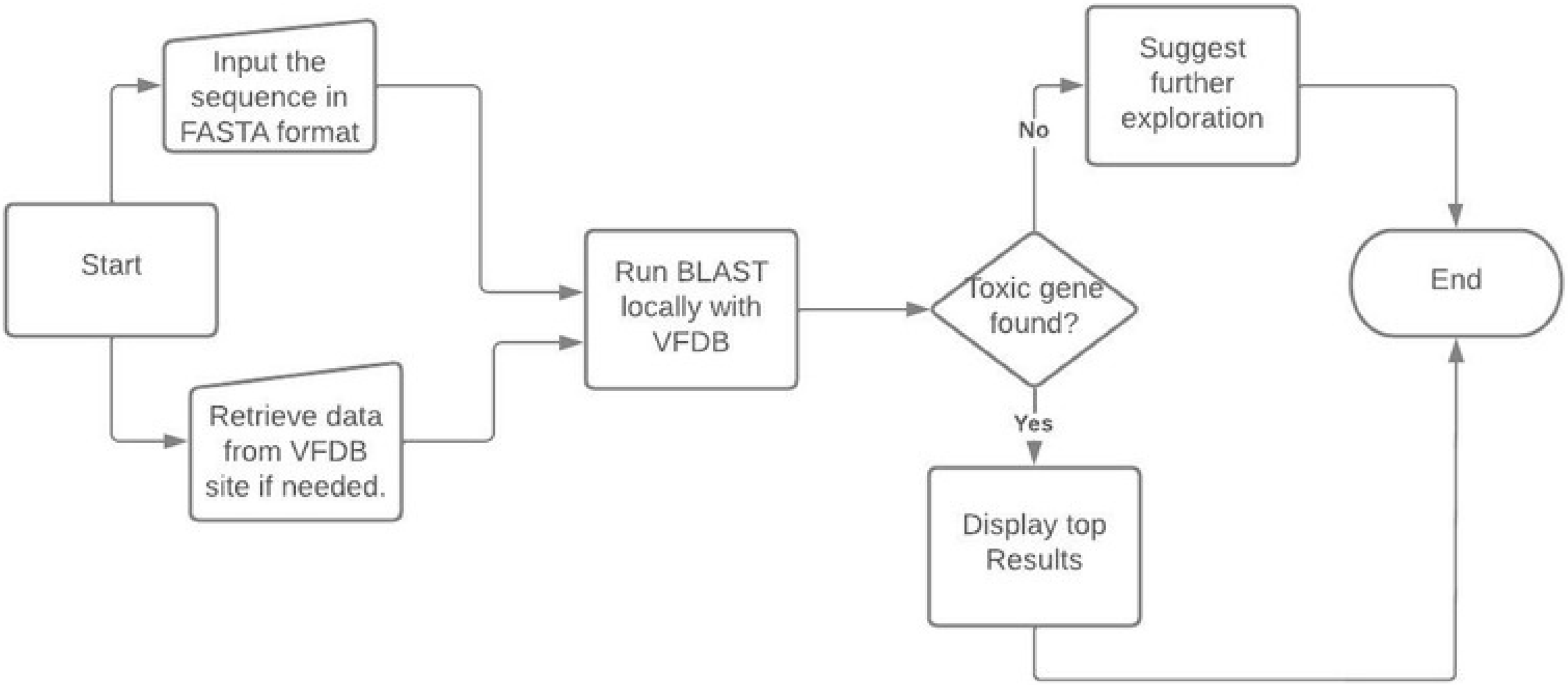
# VIRULENCE FACTOR

The property of bacteria which increases the **pathogenecity.**

Determined by the portion of the gene that encodes proteins which can cause harm in various ways.

**VFDB:** http://www.mgc.ac.cn/cgi-bin/VFs/v5/main.cgi

# Flowchart of Pipeline



# PROJECT DEMONSTRATION

# COMPLEXITY

Let N be number of entries in database.

Let E be size of entry.

Let Q be size of query.

Then , Time Complexity = O(N*E*Q)

Space Complexity = O(Q + E)

**a few seconds**

RUN-TIME

**About 100 mb**

MEMORY

# IMPACTS

Offline identification is possible

useful for field sequence analysis in remote locations.

Easy for Layman

# FUTURE POTENTIALS

- Expansion to also consider protein sequences.
- Expansion to be more specialized for nepal
- Disease source identification using location data

# RESOURCES USED

- **https://www.ncbi.nlm.nih.gov/blast+**
- **http://www.mgc.ac.cn/cgi-bin/VFs/v5/main.cgi**
- **Python and tkinter**