Instruction for the reproduction of the experiment:

Required installed Software:
Python 3.4 or above

Required installed packages:
Biopython
Bio.Phylo
Numpy
Matplotlib
Scipy
Fuzzywuzzy
Pandas

Database Download Link:
http://www.ncbi.nlm.nih.gov/Entrez

The output returned by the Entrez Programming Utilities is typically in XML format. To parse such output, you have several options:
1. Use Bio.Entrez's parser to parse the XML output into a Python object;
2. Use the DOM (Document Object Model) parser in Python's standard library;
3. Use the SAX (Simple API for XML) parser in Python's standard library;
4. Read the XML output as raw text, and parse it by string searching and manipulation.

Genome database consists of following dataset—

Assembly	genome assembly information
BioCollections	museum, herbaria, and other biorepository collections
BioProject	biological projects providing data to NCBI
BioSample	descriptions of biological source materials
Clone	genomic and cDNA clones
dbVar	genome structural variation studies
Genome	genome sequencing projects by organism
GSS	genome survey sequences
Nucleotide	DNA and RNA sequences
Probe	sequence-based probes and primers
SNP	short genetic variations
SRA	high-throughput DNA and RNA sequence read archive
Taxonomy	taxonomic classification and nomenclature catalog

NB: You need to change the code on exact point to read various dataset from ‘Genome’ database. Also need to change parameters to get the result variation. Modular programming is used to deal with such a huge dataset. Genetic Algorithm performs the best when it’s parameters are well tuned. So, there may be some deflection in reproduction of experimental results with the actual results mention in the paper.

S M Rafiuddin Rifat
Email: *rifat11cseruet@gmail.com*
ID 0417052072

and

Most. Jannatul Ferdous
ID 0417052080