*International Journal of Conceptions on Computing and Information Technology*
*Vol.10, Issue 1, April' 2025; ISSN: 2345 - 9808*

# Construction of genome diagram with its different forms of benyvirus – a python approach

Prof. Dr. P K Srimani
Director, R&D Division, B.U.,
DSI, Bangalore, India.
profsrimanipk@gmail.com

Mrs. Kumudavalli M V
Dept. of MCA (BU), DSI,
Bangalore, India.
Research Scholar, SCSVMV University,
Kanchipuram, India.
kumudamanju@gmail.com

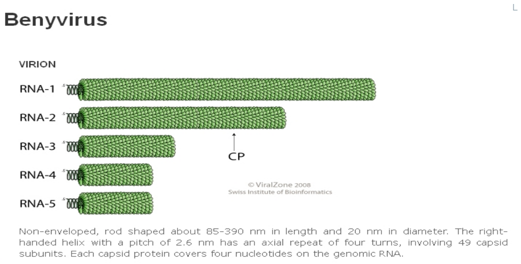
*Abstract*— **Sequencing the Genome and analyzing it is the key concept in molecular Biology and Bioinformatics these days, where the genome sequencing generally produces a huge quantity of data which is used for gene expression study and comparing the same across various complete genomes. A diagrammatic approach is more easy and symbolic for understanding and interpreting the genomic data. This paper gives the genome diagrams of Beet necrotic yellow vein virus (BNYVV), a complete genome using Python-programming language. GenomeDiagram is a module which is integrated with the freely-available Biopython bioinformatics libraries of python programming language which is being used in this paper to produce high quality scalable vector format images, in particular Adobe. The results of the present study emphasizes that python could certainly inspire the researchers to develop more programs related to Bioinformatics field.**

***Keywords***- **beny virus, Genome, Genome diagram, python, Biopython**

## I. INTRODUCTION

Biological data and the computational techniques used for the study of the same are in general known as Bioinformatics. Technology and development are the key issues in the field of life sciences, since they involve huge Biological data and varied problems related to the field. One such issue is the Genome study. In recent years, lot of research work has been carried out with regard to Genome Projects which requires the diagrammatic representation of Genome.

The complete hereditary information of an organism is its Genome. It is encoded either in DNA/RNA. But for most of the viruses it is encoded in RNA. The genome includes both the genes and the non-coding sequences of the DNA/RNA. One such RNA encoded complete genome is Beet necrotic yellow vein virus (Beny virus)[1]. It is a plant virus and the virus particles are rod-shaped, with helical symmetry and a central canal. Their diameter is about 20 nm and the modal lengths are 390, 265, 105, 90 and 80 nm, corresponding to five RNAs [2]. RNAs 1 and 2 encode "house-keeping" genes involved in replication, assembly and cell-to-cell movement, whereas RNAs 3, 4 and 5 are associated with vector-mediated infection and disease development in sugar beet roots. RNA 5 is found only in certain isolates that occur in limited areas.



Non-enveloped, rod shaped about 85-390 nm in length and 20 nm in diameter. The right-handed helix with a pitch of 2.6 nm has an axial repeat of four turns, involving 49 capsid subunits. Each capsid protein covers four nucleotides on the genomic RNA.

Fig. 1 Physical structure of Beny Virus

The disease is usually distributed as foci (patches) in the sugar beet field. The most useful leaf symptom is visible at the end of the growing season. Leaves become pale green in color, with long petioles and upright growth. Symptoms are characterized by root stunting and proliferation of lateral rootlets on the main tap roots, and yellow-brown coloration of vascular bundles.

Fig. 2 Sugar Beet infected by Beny Virus

These typical root symptoms give rise to the name "rhizomania". In early and severe infection, the plants are stunted, wilted and eventually die. In this condition, the bright yellow color followed by necrosis along veins (giving the virus name "necrotic yellow vein") is very rarely seen.

This symptom only results from movement of virus to leaves. The virus causes severe damage in sugar beets. Yield losses depend greatly on the inoculums level in the soil, the weather conditions and the time of infection. Severe infection leads to a reduction in root yield of 50 % or more, and the sugar content is greatly decreased. Unusually low levels of sugar are used as an early indicator of virus infection [3].

Fig. 3 Beet Necrotic Yellow Vein Virus infected leaf

The most important factors affecting the development of disease in infected fields are the level of inoculums, and soil temperature and moisture. The disease becomes severe where there is poor soil structure, inadequate drainage, frequent heavy rainfall or the use of (excessive) irrigation.

The present study gives an insight about the utilization of a python module which is an integral part of Biopython called as GenomeDiagram to construct various forms of genome diagram with respect to the sequences or the complete genome under study. The program presented in this paper helps in construction of genome diagram for beny virus which aids in the interpretation, exploration and visualization of the genome data. This is a helpful program for diagrammatic approach towards the analysis of Biological sequences of various organisms in the field of biotechnology and molecular biology.

This paper is organized as follows: Related works; Need and importance of the problem; methodology; Experimental results and discussion and conclusion.

## II. RELATED WORKS

Genome projects have gained popularity in Biological fields in the recent years. Many genome projects have given fruitful results, which in turn has helped the biologists and scientist to continue with their further research and study pertaining to various organisms [4]. The complete genome study of the Beet necrotic yellow vein virus has helped the farmers and the biologists in preventing and drug designing of the virus. The thorough survey of the genome sequence analysis has led the researchers in a competitive path in the field of medicine. A pictorial representation of the genome is of importance where the data set is huge. Some tools and programs are available for construction of diagrams pertaining to genome [5][6]. One such programmatic approach for genome diagram construction is python [7]. A thorough survey of the literature pertaining to the subject reveals that very sparse literature is available in this direction. Absolutely no work is available with regard to the present work. Hence, the present investigation is carried out.

## III. NEED AND IMPORTANCE OF THE PROBLEM

*Beet necrotic yellow vein virus* is a plant virus transmitted by the protozoan *polymyxa betae*. Polymyxa forms highly resistant spores that can rest in soils for more than two decades. The virus can rest in these spores, and when the spores germinate into mobile Zoospores, they transmit the virus to the plant. The virus has a narrow host range and is transmitted by the protozoan *Polymyxa betae* and by mechanical inoculation. It occurs wherever sugar beets are grown and causes a disease of economic importance. The virus causes "rhizomania" disease of sugar beet. In nature, it infects sugar beet, fodder beet, Swiss chard and spinach. With the availability of genome sequence pertaining to beny virus it is more significant to have a diagrammatic approach in analyzing these sequences in order to detect and cure the diseases caused by the BNYVV. Since not much work in this area has been done, the present investigation is carried out to throw some light on the qualitative as well as the quantitative aspects of the problem.

## IV. MATERIALS AND METHODS

Complete genome sequences of Beet necrotic yellow vein virus (BNYVV), which is plant virus that mostly affects the sugar beet, spinach e.t.c are considered for the current study.

Beet necrotic yellow vein virus (BNYVV)

Genus: Benyvirus

Species: Yellow Vein Virus

Acronym: BYNVV

Project Data Type: RefSeq Genome

TABLE I. DATA SET

| Name | RefSeq | GenBank |
|---|---|---|
| RNA 1 | NC_003514.1 | D84410.1 |
| RNA 2 | NC_003515.1 | D84411.1 |
| RNA 3 | NC_003516.1 | D84412.1 |
| RNA 4 | NC_003517.1 | D84413.1 |
| RNA 5 | NC_003513.1 | D63936.1 |

Visualization of large-scale genomic, comparative genomic and other data with reference to a single chromosome or other biological sequences is the main objectives of the GenomeDiagram, a module/package of Python programming Language.

### A. *Python*

Python is an easy to learn computer language that has gained traction among Scientists. This is likely because it is easy to use, yet powerful enough to accomplish most programming goals. Python makes the user start doing real programming very quickly. Python is an Object-oriented language, yet it allows the user to program in a classic procedural way, which is also a reason to call Python as "Multi-paradigm language." Python has gained a lot of popularity in the field of Life Sciences. It is widely being used for molecular visualization, genomic annotation, data manipulation and countless other applications.

Python being a programming language has many features which makes it more popular viz Readability, Built-in Features, Multi paradigm, Program Portability, Support libraries, Extensibility, Open Source, Cross Platform e.t.c.

### *B. Eric*

Eric is an Integrated Development Environment for Python. Its most prominent features are: • Advanced project management facilities • Source code auto completion • Error highlighting • Integrated Python debugger • Integrated class browser • Integrated version control interface for VCS and Subversion repositories

### *C. Biopython*

Bioinformatics is a fast growing field of Biology, with various unique applications being developed quite frequently. Still there are some tasks that are repeated, many programs and standard files are being shared. This situation made the developers to come with a new package called as Biopython. Biopython is a set of libraries to provide the ability to deal with “things” of interest to biologists working on the computer. The Biopython Project is an international association of developers of freely available Python [8] tools for computational molecular biology.

Biopython is a package of useful modules to develop bioinformatics applications. It caters to most of the developmental requirements for biological problems. Biopython is part of the family of open-bio projects (also known as Bio*), for which institutionally it is a member of the Open Bioinformatics Foundation [9].

The following steps were used in the analysis process:

**Step1:** The Complete genome of Beet necrotic yellow vein virus (BNYVV) was retrieved from NCBI

**Step2:** To execute the program for Diagram construction it requires; Python 2.3 or higher and recent versions of the ReportLab and BioPython packages.

**Step3:** The python program is written in the Eric IDE for compilation and execution, as follows:

```
from reportlab.lib import colors

from reportlab.lib.units import cm

from Bio.Graphics import GenomeDiagram

from Bio import SeqIO

record = SeqIO.read("NC_003514.1.gb", "genbank")

# Sequences can be changed to get different diagrams

for feature in record.features:

if feature.type != "gene":

#Exclude this feature

continue

if len(gd_feature_set) % 2 == 0:

color = colors.blue

else:

color = colors.lightblue

gd_feature_set.add_feature(feature, color=color,
label=True)
```

**Step4:** The Genome diagram for Benyvirus is in particular obtained as an Adobe PDF

```
gd_diagram.draw(format="linear",
orientation="landscape", pagesize=’A4’,

fragments=4, start=0, end=len(record))

gd_diagram.write("beet_linear.pdf", "PDF")

gd_diagram.draw(format="circular", circular=True,
pagesize=(20*cm,20*cm),start=0,

end=len(record), circle_core=0.7)

gd_diagram.write("beet_circular.pdf", "PDF")
```

**Step5:** The program codes are enhanced to get various forms and different structures of diagrams.

```
#Full height shafts, giving pointed boxes:

gd_feature_set.add_feature(feature, sigil="ARROW",
        color="brown",arrowshaft_height=1.0)

#Or, thin shafts:

#Or, very thin shafts:

        #Short arrow heads:

#Or, longer arrow heads:
```

**Step6:** Multiple tracks can be used to produce the genome diagrams with following code:

```
A_rec = SeqIO.read("NC_003514.1.gb", "gb")

B_rec = SeqIO.read("NC_003514.2.gb ", "gb")

C_rec = SeqIO.read("NC_003514.3.gb ", "gb")

D_rec = SeqIO.read("NC_003514.4.gb ", "gb")

E_rec = SeqIO.read("NC_003514.5.gb ", "gb")

from reportlab.lib.colors import red, grey, orange, green,
brown, blue, lightblue, purple

A_colors = [purple]*5 + [grey]*7 + [orange]*2 + [grey]*2
+ [orange] + [grey]*11 \ [green]*4 + [grey] + [green]*2 +
[grey, green] + [brown]*5 + [blue]*4 +\[lightblue]*5 + [grey,
lightblue] + [purple]*2 + [grey]

# similarly the color compositions are made for

# B_colors, C_colors, D_colors, E_colors.

gd_diagram.write(name + ".pdf", "PDF")
```

## V. EXPERIMENTAL RESULTS AND DISCUSSION

In this section, the results of the above python program with different constraints are presented in figures 4 to8.

Genome diagrams are generally used for drawing whole genomes, in particular prokaryotic genomes. They are drawn as linear diagrams or as circular wheel diagrams. Genome diagram uses a nested set of objects. A diagram can contain one or more tracks.

Fig 4(a) and (b) depict the whole genome of Beny virus "NC_003515.1" read from the GenBank files. It is a top down approach used to create the linear and circular diagrams and written to Adobe PDF file format.

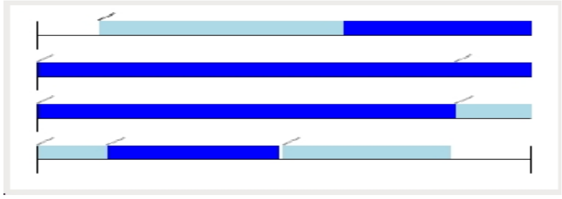


Fig. 4(a) Linear diagram of NC_003515.1 BNYVV

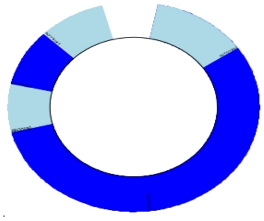


Fig. 4(b) Circular diagram of NC_003515.1 BNYVV

Fig. 5 (a) and (b) show the linear and circular arrow diagram of beny virus genome sequence NC_003513.1 and 5(c) shows the diagram with arrow shaft property being set to 0.2.

Fig. 5(a) Linear Arrow diagram of NC_003513.1

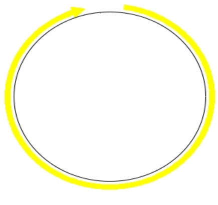


Fig.5(b) Circular arrow diagram of NC_003513.1

Fig. 5(C) Linear Arrow diagram of NC_003513.1 with 0.2 arrow shaft

Fig 6(a) and (b) show the linear arrow diagram with arrow shaft 0.1and circular diagram of NC_003514.1

Fig 6(a) Linear Arrow diagram of NC_003514.1 with 0.1 arrow shaft

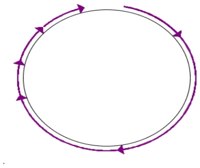


Fig. 6(b) Circular arrow diagram of NC_003514.1

Fig 7(a) and (b) shows the linear arrow diagram with arrow head 1000 and circular diagram of NC_003515.1

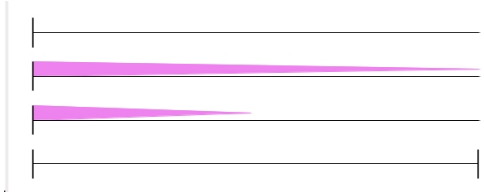


Fig. 7(a) Arrow diagram with head 1000 for NC_003515.1

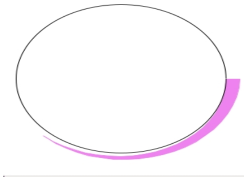


Fig. 7(b) Circular Diagram of NC_003515.1

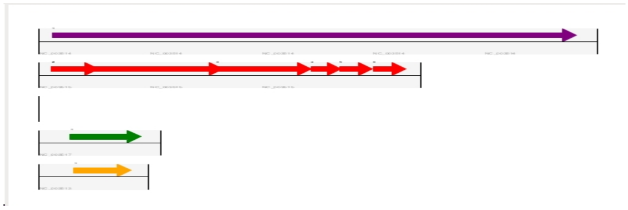


Fig.8 Multi tracks used for indicating NC_003513.1, NC_003514.1, NC_003515.1, NC_003516.1, NC_003517.1 sequences of BNYVV

Fig.8 depicts the multi tracks being used for drawing the genome diagram for complete genome of Beet Necrotic Yellow Vein Virus (NC_003513.1, NC_003514.1, NC_003515.1, NC_003516.1, NC_003517.1)

## VI. CONCLUSION

Python is an excellent programming language. It has very clear, readable syntax and strong introspection capabilities make it more popular. Python has a very user-friendly syntax, which inspires the programmers to develop various packages relating to current day requirements. In this context it is emphasized that BRIG is the only other tool available for the construction of genome diagrams that can give only the ring images of the genome under study. GenomeDiagram package helps the biologists and the medical field analysts to study the behavior of any genome diagrammatically no matter how huge the data set is. From the present study it is concluded that certainly Python inspires the programmers to develop more packages in the fields of Biology and Bioinformatics.

## REFERENCES

[1] Tamada, in *Encyclopedia of Virology,* 2nd Edn, p.154, A. Granoff & R. G. Webster, eds, London: Academic Press, 1999

[2] Tamada & Abe, *Jrnl of General Virology* 70: 3391, 1989.

[3] Asher, in *The Sugar Beet Crop: Science into practice*, p. 311, D. A. Cooke & R. K. Scott, eds, London: Chapman and Hall, 1993.

[4] Tamada, Uchino, Kusume & Saito, Phytopathology 89: 1000, 1999.

[5] Nabil-Fareed Alikhan1, Nicola K Petty1 , Nouri L Ben Zakour1 , and Scott A Beatson1 , "BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons", BMC Genomics. Vol. 12, 2011

[6] Srimani P. K, Kumudavalli M.V., "A comprehensive study of Defensin-1 Protein present in honey bees – a novel approach," Intl. Journal of Advanced Scientific and Technical Research, Vol. 6,Issue, 02, pp. 62-71, Dec, 2012.

[7] Leighton Pritchard, Jennifer A. White, Paul R.J. Birch and Ian K. Toth, "GenomeDiagram: a python package for the visualization of large-scale genomic data," Bioinformatics Applications Note, Vol. 22 ,no. 5, pp. 616–617, 2006

[8] http://www.python.org

[9] http://www.open-bio.org