



## Module I :NCBI Genome database:

### Genome:

In modern molecular biology and genetics, the genome is the genetic material of an organism. It consists of DNA (or RNA in RNA viruses). The genome includes both the genes, (the coding regions), the noncoding DNA (Genome is complete set of genes).

### NCBI Genome database:

URL:<http://www.ncbi.nlm.nih.gov/nuccore/?term>

Fig: NCBI Genome database homepage

### Exercise:

1. Search for HADV9 and HADV37 and retrieve its nucleotide sequence.
  2. Go to the NCBI taxonomy database and Search for viruses
- ==>

Taxonomy  [Create alert](#) [Limits](#) [Advanced](#)

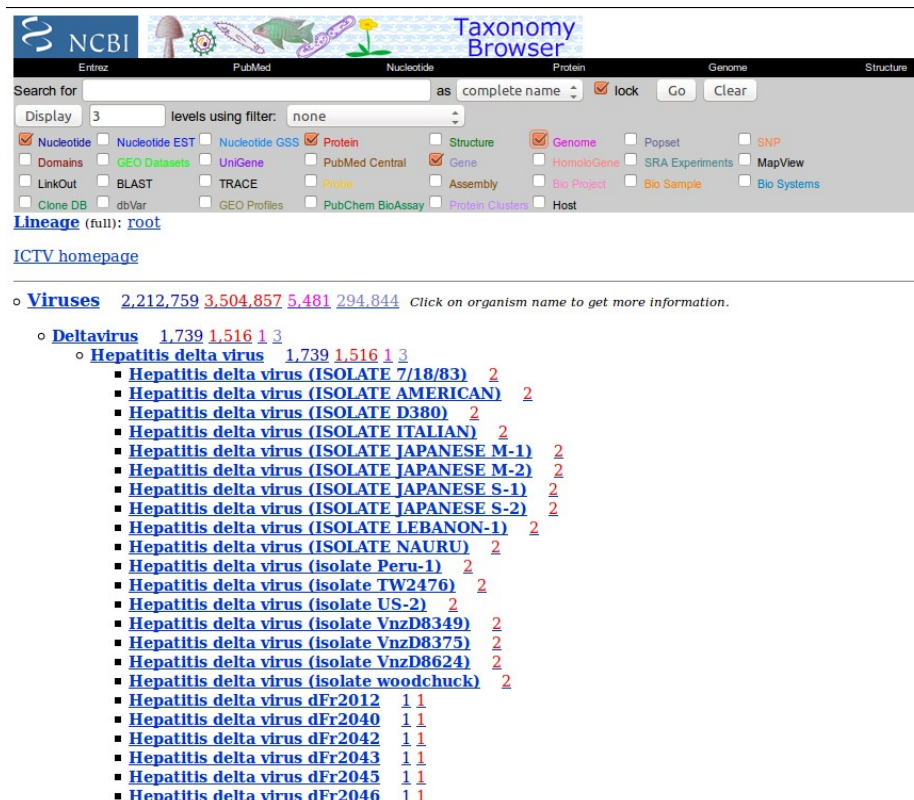
Display Settings: [Summary](#)

[Viruses](#)

superkingdom, **viruses**

[Nucleotide](#) [Protein](#)

--click on "Viruses"



NCBI Taxonomy Browser

Search for:  as: complete name ☒ lock [Go](#) [Clear](#)

Display: 3 levels using filter: none

☒ Nucleotide ☐ Nucleotide EST ☐ Nucleotide GSS ☒ Protein ☐ Structure ☒ Genome ☐ Popset ☐ SNP

☐ Domains ☐ GEO Datasets ☐ UniGene ☐ PubMed Central ☒ Gene ☐ HomoloGene ☐ SRA Experiments ☐ MapView

☐ LinkOut ☐ BLAST ☐ TRACE ☐ Probe ☐ Assembly ☐ Bio Project ☐ Bio Sample ☐ Bio Systems

☐ Clone DB ☐ dbVar ☐ GEO Profiles ☐ PubChem BioAssay ☐ Protein Clusters ☐ Host

[Lineage](#) (full): [root](#)

[ICTV homepage](#)

- Viruses 2,212,759 3,504,857 5,481 294,844 [Click on organism name to get more information.](#)
  - Deltavirus 1,739 1,516 1 3
    - Hepatitis delta virus 1,739 1,516 1 3
      - Hepatitis delta virus (ISOLATE 7/18/83) 2
      - Hepatitis delta virus (ISOLATE AMERICAN) 2
      - Hepatitis delta virus (ISOLATE D380) 2
      - Hepatitis delta virus (ISOLATE ITALIAN) 2
      - Hepatitis delta virus (ISOLATE JAPANESE M-1) 2
      - Hepatitis delta virus (ISOLATE JAPANESE M-2) 2
      - Hepatitis delta virus (ISOLATE JAPANESE S-1) 2
      - Hepatitis delta virus (ISOLATE JAPANESE S-2) 2
      - Hepatitis delta virus (ISOLATE LEBANON-1) 2
      - Hepatitis delta virus (ISOLATE NAURU) 2
      - Hepatitis delta virus (isolate Peru-1) 2
      - Hepatitis delta virus (isolate TW2476) 2
      - Hepatitis delta virus (isolate US-2) 2
      - Hepatitis delta virus (isolate VnzD8349) 2
      - Hepatitis delta virus (isolate VnzD8375) 2
      - Hepatitis delta virus (isolate VnzD8624) 2
      - Hepatitis delta virus (isolate woodchuck) 2
      - Hepatitis delta virus dFr2012 1 1
      - Hepatitis delta virus dFr2040 1 1
      - Hepatitis delta virus dFr2042 1 1
      - Hepatitis delta virus dFr2043 1 1
      - Hepatitis delta virus dFr2045 1 1
      - Hepatitis delta virus dFr2046 1 1

**Fig:** result page of taxanomy database.

--select Nucleotide, Gene, Genome and Protein.

4. Select any one virus with minimum 5-10 genes and make the table wich will have following information

- Gene start, Gene end (or stop), Strand, Product

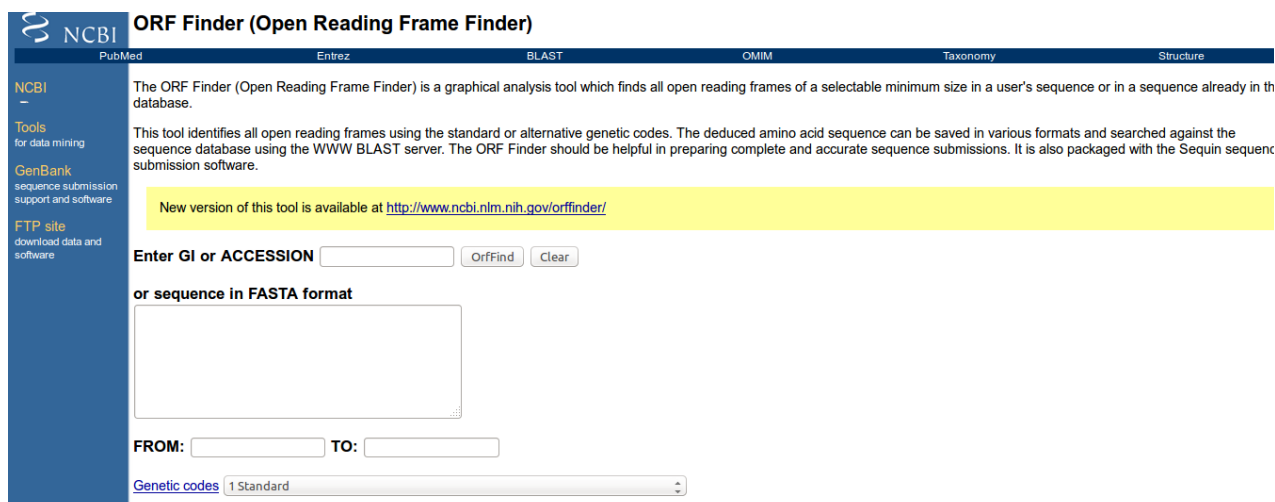
| Gene Start | Gene End | Strand | Product | ORF finder | GeneMark S |
|------------|----------|--------|---------|------------|------------|
| 549        | 977      | +      | E6      |            |            |
| 989        | 1264     | +      | E7      |            |            |
| 1266       | 3104     | +      | E1      |            |            |
| 4359       | 5888     | +      | L2      |            |            |
| 5885       | 7435     | +      | L1      |            |            |

**Table:** Gene information of *Castor canadensis papillomavirus 1*

## ORF Finder:

- Open ORF finder using following URL

URL: <http://www.ncbi.nlm.nih.gov/gorf/gorf.html>



**ORF Finder (Open Reading Frame Finder)**

The ORF Finder (Open Reading Frame Finder) is a graphical analysis tool which finds all open reading frames of a selectable minimum size in a user's sequence or in a sequence already in the database.

This tool identifies all open reading frames using the standard or alternative genetic codes. The deduced amino acid sequence can be saved in various formats and searched against the sequence database using the WWW BLAST server. The ORF Finder should be helpful in preparing complete and accurate sequence submissions. It is also packaged with the Sequin sequence submission software.

New version of this tool is available at <http://www.ncbi.nlm.nih.gov/orffinder/>

Enter GI or ACCESSION  OrFind Clear

or sequence in FASTA format

FROM:  TO:

Genetic codes 1 Standard

Fig: Homepage of ORF finder

- Paste genome sequence of *Castor canadensis papillomavirus 1* in provided space
- Select largest ORF and compare the results of ORF and downloaded data.

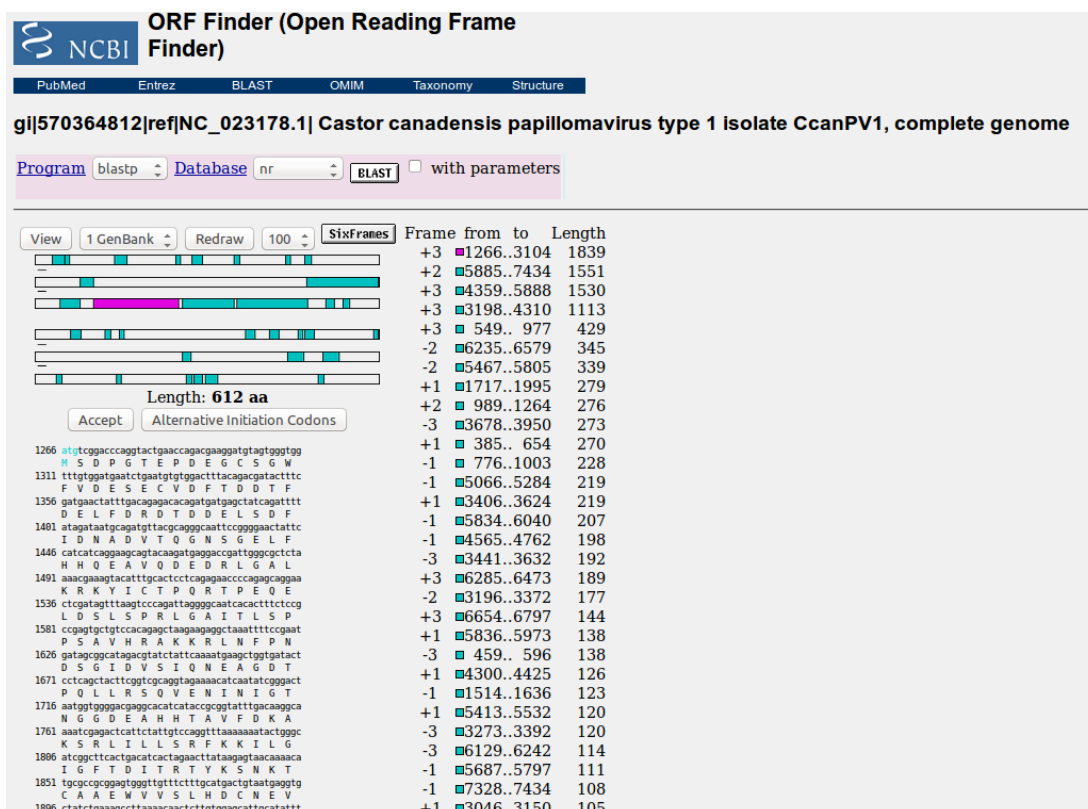


Fig: Result page of ORF finder.

## GeneMarkS:

- Open GeneMarkS using following URL.

URL: <http://exon.gatech.edu/genemark/genemarks.cgi>

**GeneMarkS**

**John Besemer, Alexandre Lomsadze and Mark Borodovsky**  
[GeneMarkS: a self-training method for prediction of gene starts in microbial genomes.](#)  
[Implications for finding sequence motifs in regulatory regions.](#)  
*Nucleic Acids Research* (2001) 29, pp 2607-2618

This webpage provides accesses to version 4.28 of gene prediction program GeneMarkS. This version combines the original 2001 prokaryotic GeneMarkS with later development, which extended the unsupervised gene prediction to intron-less eukaryotes, eukaryotic viruses, phages and EST/cDNA sequences.

[Browse GeneMarkS manual](#)

Input sequence

Enter sequence (FASTA or multi FASTA format)

or, upload file:  No file selected.

Action

Options

| Sequence type   | Output format for gene prediction                                 | Output options  | Optional: results by E-mail   |
|---|---|---|---|
| <input type="radio"/> Prokaryotic<br><input type="radio"/> Intronless eukaryotic<br><input checked="" type="radio"/> Virus<br><input type="radio"/> Phage<br><input type="radio"/> EST/cDNA | <input checked="" type="radio"/> LST<br><input type="radio"/> GFF | <input checked="" type="checkbox"/> Protein sequence<br><input type="checkbox"/> Gene nucleotide sequence<br><input type="checkbox"/> Coding potential graph (not for multi FASTA)<br><input type="checkbox"/> PDF<br><input type="checkbox"/> PostScript | E-mail<br><input type="text"/><br>Subject<br>GeneMarkS<br><input type="checkbox"/> Compress files |

**Fig:** Homepage of GeneMarkS

- Paste genome sequence of *Castor canadensis papillomavirus 1* in provided space
- Click on “Start GeneMarkS”

**GeneMarkS**

**John Besemer, Alexandre Lomsadze and Mark Borodovsky**  
[GeneMarkS: a self-training method for prediction of gene starts in microbial genomes.](#)  
[Implications for finding sequence motifs in regulatory regions.](#)  
*Nucleic Acids Research* (2001) 29, pp 2607-2618

This webpage provides accesses to version 4.28 of gene prediction program GeneMarkS. This version combines the original 2001 prokaryotic GeneMarkS with later development, which extended the unsupervised gene prediction to intron-less eukaryotes, eukaryotic viruses, phages and EST/cDNA sequences.

[Browse GeneMarkS manual](#)

GeneMarkS is running

job ID = genemarks.20160604.060323.14499 ; Estimated run time: 1 second(s)

Gene Prediction Results

| output                         | link                    |
|--------------------------------|-------------------------|
| Coordinates of predicted genes | <a href="#">gms.out</a> |

[Contact Us](#) | [Home](#)

**Fig:** Result page of GeneMarkS

--click on link “gms.out”

```
GeneMark.hmm PROKARYOTIC (Version 3.26)
Date: Sat Jun  4 06:03:23 2016
Sequence file name: seq.fna
Model file name: GeneMark_hmm_heuristic.mod
RBS: false
Model information: Heuristic_model_for_genetic_code_1_and_GC_45

FASTA definition line: gi|570364812|ref|NC_023178.1| Castor canadensis papillomavirus type 1 isolate CcanPV1, complete genome
Predicted genes
Gene  Strand  LeftEnd  RightEnd  Gene Length  Class
#
1      +       549      977       429         1
2      +       989     1264       276         1
3      +      1266     3104     1839         1
4      +      3198     4310     1113         1
5      +      4359     5888     1530         1
6      +      5885     7435     1551         1
```

**Fig:** Result page of GeneMarkS