

BINF-Practical: 05

ORF finding in viral nucleotide

Introduction: organization of Eukaryotic and Prokaryotic genomes

A genome is an organism's complete set of DNA, comprising of nuclear and mitochondrial DNA. Each genome contains all of the information needed to build and maintain that organism. A human haploid cell, consist of 23 nuclear chromosome and one mitochondrial chromosome, contains more than 3.2 billion DNA base pairs

1. Eukaryotic Genome Organization: Eukaryotic genome is linear and conforms the Watson-Crick Double Helix structural model. Embedded in Nucleosome-complex DNA & Protein (Histone) structure that pack together to form chromosomes. Eukaryotic genome have unique features of Exon - Intron organization of protein coding genes, representing coding sequence and intervening sequence that represents the functionality of RNA part inside the genome.

2. Prokaryotic Genome Organization: Each bacterial chromosome is made by a single circular DNA molecule (rarely linear). Usually each cell contain one single copy of each chromosome. The genetic material can be seen as a fairly compact clump (or series of clumps) that occupies about a third of the volume of the cell named NUCLEOID. The DNA of these loops is not found in the extended form of a free duplex, but instead is compacted by association with proteins

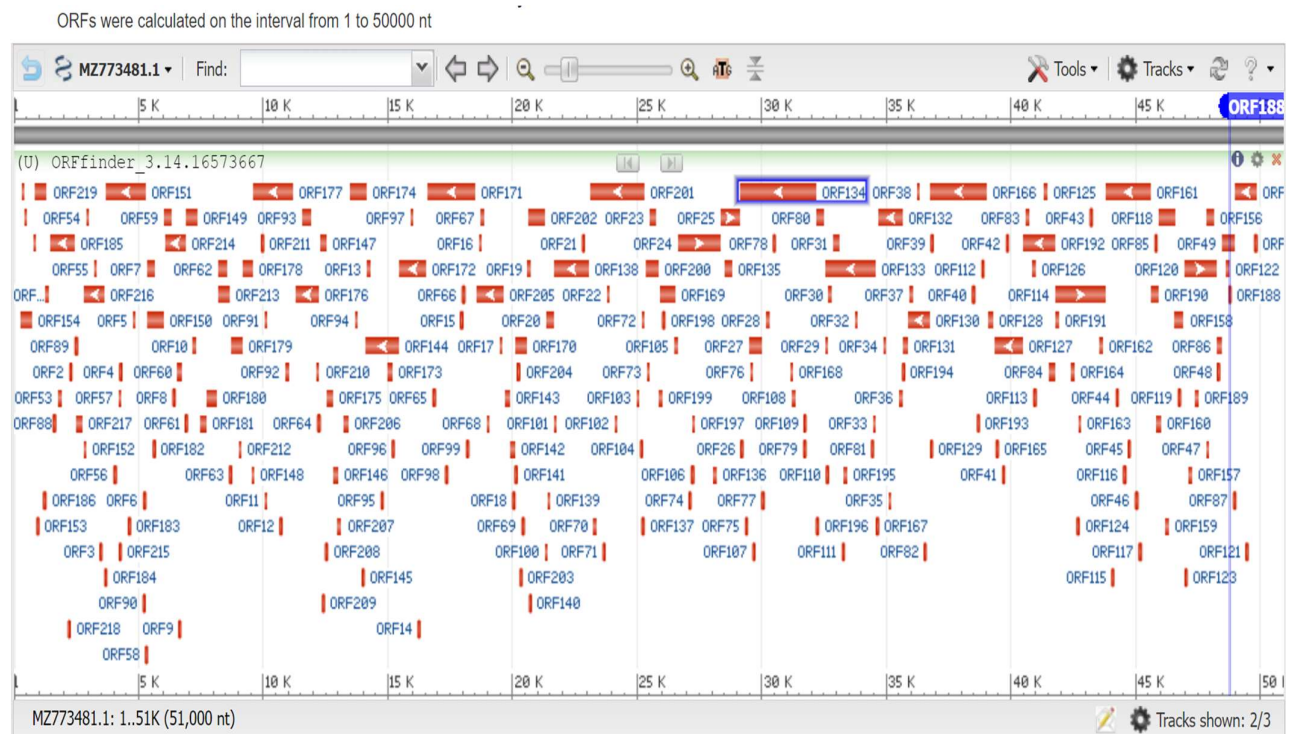
ORFS: open reading frames (ORFs) are defined as spans of DNA sequence between the start and stop codons. One common use of open reading frames (ORFs) is as one piece of evidence to assist in gene prediction. Long ORFs are often used, along with other evidence, to initially identify

candidate protein-coding regions or functional RNA -coding regions in a DNA sequence.

Methodology:

1. Viral DNA is selected first , here I've taken ***Swinepox virus*** genome with accession no. **MZ773481** from NCBI website.
2. Then accession no. is given to the ORF viewer site or we can also give the sequence in fasta format there. Along with keeping minimal ORF length (**75**), Genetic code type(**standard**) and ORF start codon to use (**ATG only**).
3. on submit the job we will get the following result telling about different ORF sites.

Results:



Label	Strand	Frame	Start	Stop	Length (nt aa)
ORF194	-	2	35736	35659	78 25
ORF44	+	1	44230	44307	78 25
ORF58	+	2	5276	5353	78 25
ORF188	-	2	48834	48757	78 25
ORF74	+	2	27086	27163	78 25
ORF78	+	2	30467	30544	78 25
ORF82	+	2	36518	36595	78 25
ORF111	+	3	33243	33320	78 25
ORF115	+	3	44046	44123	78 25
ORF129	-	3	36851	36774	78 25
ORF140	-	3	20756	20679	78 25

Some of the ORF sequences found are as below:

>|c| ORF188

MMFSKSVRCSYERNRYVCSNSHLNI

>|c| ORF44

MTTLNIDTSQRVLICYILEKHQIPI

>|c| ORF82

MNKKDIEKALILHFLYIFRCSYLV

Conclusion:

In, Swinepox virus isolate SWPV/domesticpig/GER/2019, complete genome

ORFs found: 219

Genetic code: 1

Start codon: 'ATG' only