

1. Purpose

This document describes the process by which new concatenated reference sequences are generated for Influenza virus sequences stored in the BioHealthBase database.

2. Method Description

A BioPerl-based script is developed to string together, head to tail, and in order of the segment number, all the segment sequences for a particular Influenza virus strain.

3. Input Data Preparation

RefSeq sequences are obtained from NCBI in GenBank format.

4. Output Data, Processing and Display

Concatenated Influenza reference sequences are produced in GenBank format, with all features adjusted for the new coordinate system. They are subsequently loaded to BioHealthBase data warehouse.

The display examples relevant to the concatenated Influenza reference sequences include Gene Details page and the Genome Browser:

The screenshot shows the 'Influenza Virus' Gene Details page. The header includes the site logo, navigation links (HOME, FAQ, HELP, FEEDBACK, BIOHEALTHBASE HOME), and a shopping cart icon with a 'View GeneCart' link. A left sidebar contains a 'Database Search' menu with options like Sequence, Epitope, and Public Database Identifier, as well as 'Genome Browser' and 'Tools' sections. The main content area is titled 'Gene Details' and features a 'Download as PDF' button. It is organized into sections: 'Strain Identification' (with fields for Complete Genome Set, Organism Name, Strain Name, Subtype, Host, and NCBI Taxon ID), 'Gene Identification' (with fields for Gene Product Name, Locus Name, Entrez Gene ID, and a detailed Comment), 'Genomic Location' (with fields for Coordinates, Strand, Gene Length, and a link to View Gene Sequence), and 'Gene Features' (which includes a table of CDS features). A right sidebar contains a 'Page Contents' list and a 'References' section.

CDS	Complete CDS	Start	End	CDS Length (nt)	Protein Length(aa)	Sequence
1	Yes	10279	11643	1365	454	CDS ; Protein

Influenza A Virus, A/Puerto Rico/8/34, Reference @ BioHealthBase

Showing 13.59 kbp from concat_A_PuertoRico_8_34, positions 1 to 13,588

Instructions

Landmark or Region search functionality is currently unavailable.

Tips: If tracks are not displayed, click the browser refresh button. If you select a new Data Source item, please click on an example landmark to redisplay tracks.

Examples: [concat_A_PuertoRico_8_34.1.13588](#), [concat_A_PuertoRico_8_34](#).

[\[Hide banner\]](#) [\[Bookmark this\]](#) [\[Link to Image\]](#) [\[High-res Image\]](#) [\[Help\]](#) [\[Reset\]](#)

Search

Reports & Analysis

Annotate Restriction Sites

Scroll/Zoom: Show 13.59 kbp ☐ Flip

Data Source

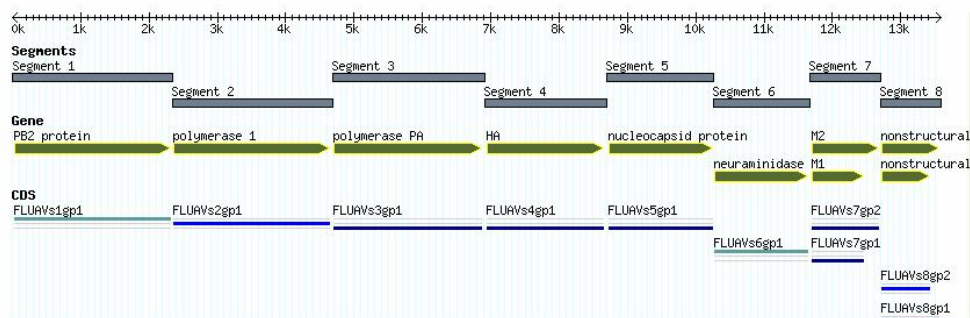
Influenza A Virus, A/Puerto Rico/8/34, Reference @ BioHealthBase

Overview

Overview of concat_A_PuertoRico_8_34



Details



5. References

1. **The Bioperl Toolkit: Perl Modules for the Life Sciences**
Stajich, J. E., etc.
Genome Research. 12(10): 1611-8. 2002
2. **GenBank.**
Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL.
Nuc Acid Res. 34: D16-20. 2006