Section C. Sequence Feature Variant Type Analysis

I. Analyze Sequence Feature Variant Types

At the end of this exercise you will be able to identify regions of viral proteins with known structural, functional and immune epitope properties, and will be able to assess the level of sequence variation in these regions.

Sequence Feature Variant Type (SFVT) can help you identify sequence variations that may correlate with phenotypic characteristics, e.g., drug sensitivity/resistance, virulence, transmissibility, etc.

- **Sequence Features (SFs):** specific regions defined based on functional properties, structural properties, and immune epitope locations. Obtained from literature, public archives, or direct submissions and validated by domain experts.
- **Variant Types (VTs):** polymorphisms in each Sequence Feature are identified as Variant Types of the Sequence Feature.
- The SFVT analysis is available for *influenza A virus* in the Influenza Research Database and *Dengue virus, Hepatitis C subtype 1a* and *Vaccinia viruses* in the Virus Pathogen Resource.

For this exercise, you will use the Influenza Research Database to examine a B-cell epitope located on the HA protein and the sequence polymorphism in this region.

a. Go to the IRD homepage ([http://www.fludb.org](http://www.fludb.org)). From the gray navigation bar, mouse over “Search Data” and click “Sequence Feature Variant Types”.

b. The SFVT landing page will be loaded. Here you can search for Sequence Features or click “Go to Sequence Feature List” to browse all Sequence Features of influenza A proteins.
c. Now search for immune epitope Sequence Features of HA H1 subtype. Type in “H1” in the Subtype for HA and NA box, select “A HA” from the segments list and “Epitope” from the Sequence Feature Type list. Then click “Search”.

d. The Sequence Feature search results will be displayed in a table.

![Sequence Feature search results table]

i. How many epitope Sequence Features are there for the H1 protein?

ii. How many amino acid residues does this Sequence Feature cover?

e. Click “View” for the Sa epitope Sequence Feature to view its details.

i. What is the reference strain defining Variant Type-1 for this Sequence Feature?

ii. What is the source of information that supports its epitope classification?

![Sequence Feature definition]

iii. How many Variant Types exist for this Sequence Feature in IRD?

iv. Click the number in the “Strain Count” column for VT-1 to view the metadata associated with the strains harboring this Variant Type. Sort the list by Subtype, Host, and Collection Date by clicking the corresponding column header. To get a better sense of the strains containing this VT, you can download this table to your computer by ticking the checkbox above the table and then clicking the “Download” button. Then open the downloaded file in Excel and try various filters. What is the subtype and host for the majority of the strains? Sort the records by Collection Date. Write down the flu seasons for these strains.
v. Return to the previous page by clicking the breadcrumb. Follow the instructions in step iv. to examine the metadata associated with the strains harboring VT-4. Write down the host, subtype, and flu season for the majority of the strains.

vi. Now search for Variant Types with Q at position 180. Click the blue “Find a VT(s)” button to expand the VT search panel. Click “Fill wildcards” to have “?” populated in all positions and then change position 180 to “Q” to search for variant types harboring 180Q. Click “Search”.

vii. How many variant types did you find?

viii. Examine VT-9 and write down the subtype, host, and collection date for the majority of the strains.
ix. Compare your notes from steps iv. (VT-1), v. (VT-4), and viii. (VT-9). Does VT-9 have the same flu season range as VT-1 and VT-4? Considering this Sequence Feature as a B-cell epitope and the metadata associated with VT-1, VT-4, VT-9, respectively, what hypothesis could you generate?

II. Submit new Sequence Features to IRD/ViPR

In an effort to keep the Sequence Feature (SF) list up-to-date and to facilitate greater community involvement in annotation efforts, we encourage you to submit new SFs to IRD/ViPR. The Submit Sequence Feature page is accessible from the “Submit Data” tab.

- This tool provides a user-friendly interface for online submission of new SFs. Drop-down menus are made available for fields such as virus type, protein, SF category, etc. to maximize the efficiency.

- Newly submitted SFs are first manually inspected and then validated using appropriate Quality Control measures to ensure accuracy and authenticity, and to avoid duplicates.

- Following Quality Control process, the new SFs are either accepted or rejected and the contributors are notified of the decisions. Submitters of accepted SFs are acknowledged on the IRD site.

Reference: