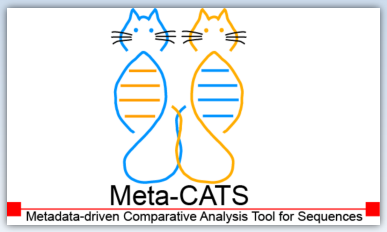
ViPR
Virus Pathogen Resource

http://www.viprbrc.org/

Freely available | *Integrated datasets* | *Bioinformatics tool suite*

Metadata-driven Comparative Analysis Tool for Sequences (Meta-CATS)



What is Meta-CATS?

A unique comparative genomics analysis tool in ViPR to identify nucleotide/amino acid positions that significantly differ between two or more groups of virus sequences.

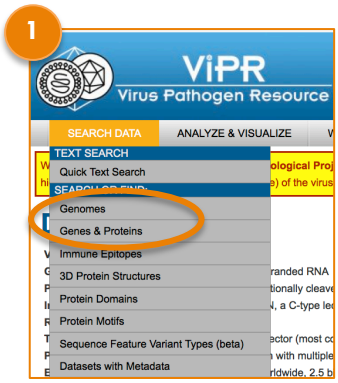
How does Meta-CATS work?

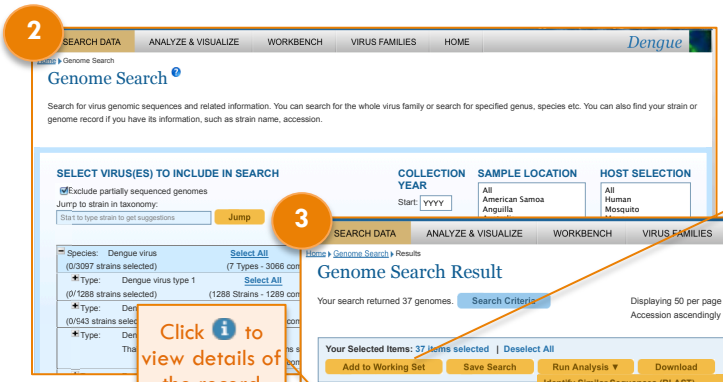
It runs a multiple sequence alignment, a chi-square test to identify positions that significantly differ from the random distribution of residues between all metadata groups, and a Pearson's chi-square test to identify the specific pairs of groups that contribute to the observed statistical difference.

How to use Meta-CATS?

- Input nucleotide/amino acid sequences to Meta-CATS
- Divide sequences into two or more groups based on metadata (phenotype)
- Run Meta-CATS

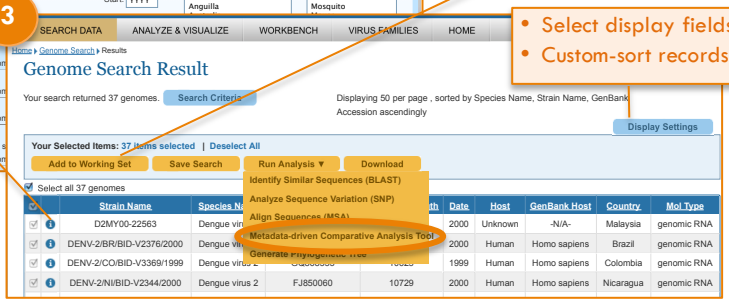
Option 1: Search for sequences and then input sequences to Meta-CATS





Select sequences and add them to a working set for future analysis. You'll need to register for a Workbench account to use this feature.

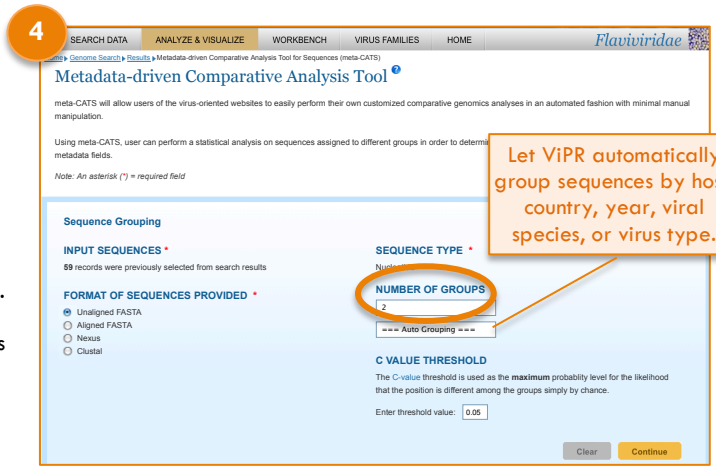
Click ⓘ to view details of the record



- Select display fields
- Custom-sort records

On the ViPR homepage, choose a virus family to start.

1. Identify sequences to include in the Meta-CATS analysis: mouse-over the "Search Data" tab and click "Genomes" or "Genes & Proteins". For this example, we will use genome sequences.
2. Select search criteria on the Genome Search page and click the "Search" button to run your query.
3. Select sequences from the search result page by clicking the checkboxes. Mouse-over the yellow "Run Analysis" button, and click "Metadata-driven Comparative Analysis Tool".
   If you want to include sequences that are not in this search result, select desired sequences and click "Add to Working Set", then add additional sequences to the same working set. Click the "Workbench" tab, find the working set you saved and click ⓘ next to it. On the working set details page, mouse-over "Run Analysis" and click "Metadata-driven Comparative Analysis Tool".
4. On the next page, choose the number of groups. If you want to group your sequences by host, country, year, viral species, or virus type, you can use ViPR's auto grouping feature by clicking the "Auto Grouping" drop-down menu and follow the prompts. Otherwise, you can group sequences manually on the next page. Adjust the C-value threshold if needed. Then click "Continue".



Let ViPR automatically group sequences by host, country, year, viral species, or virus type.

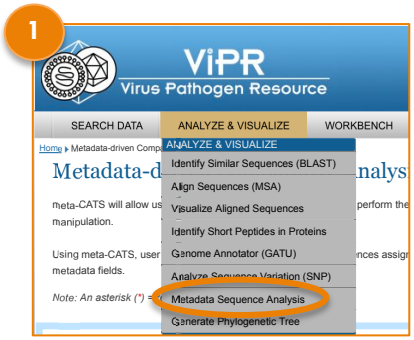
ViPR is funded by the National Institute of Allergy and Infectious Diseases (NIH / DHHS) under Contract No. HHSN272200900041C and is a collaboration between Northrop Grumman Health IT, University of Texas Southwestern Medical Center and Veena Technologies. Comments, questions, suggestions? Contact us at feedback@virusbrc.org

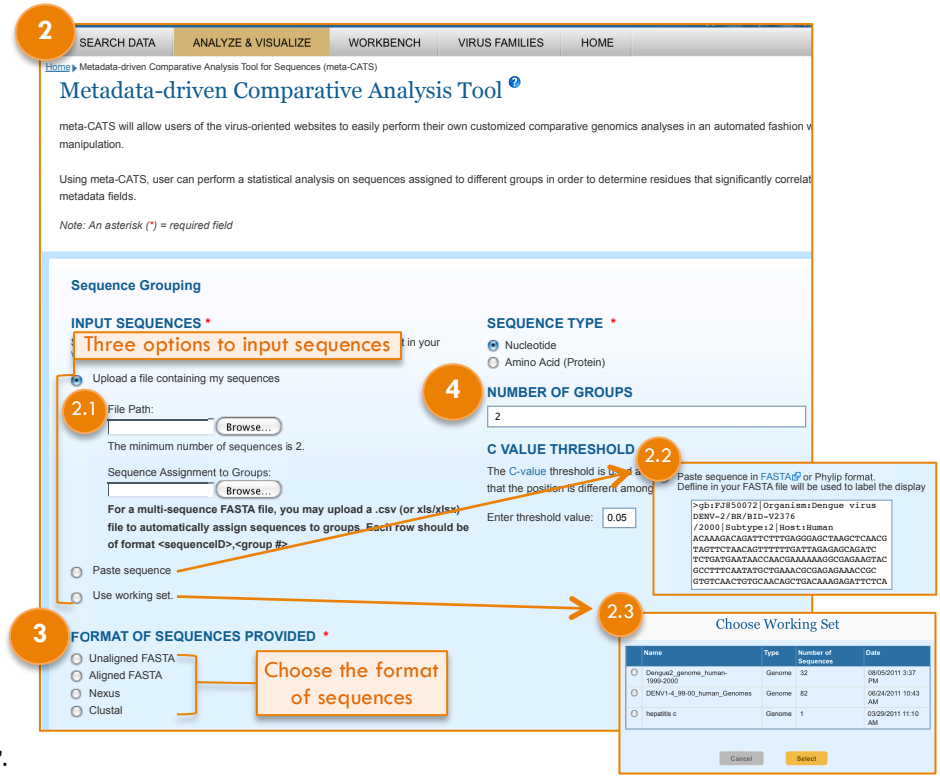
ViPR Virus Pathogen Resource

http://www.viprbrc.org/

Freely available | *Integrated datasets* | *Bioinformatics tool suite*

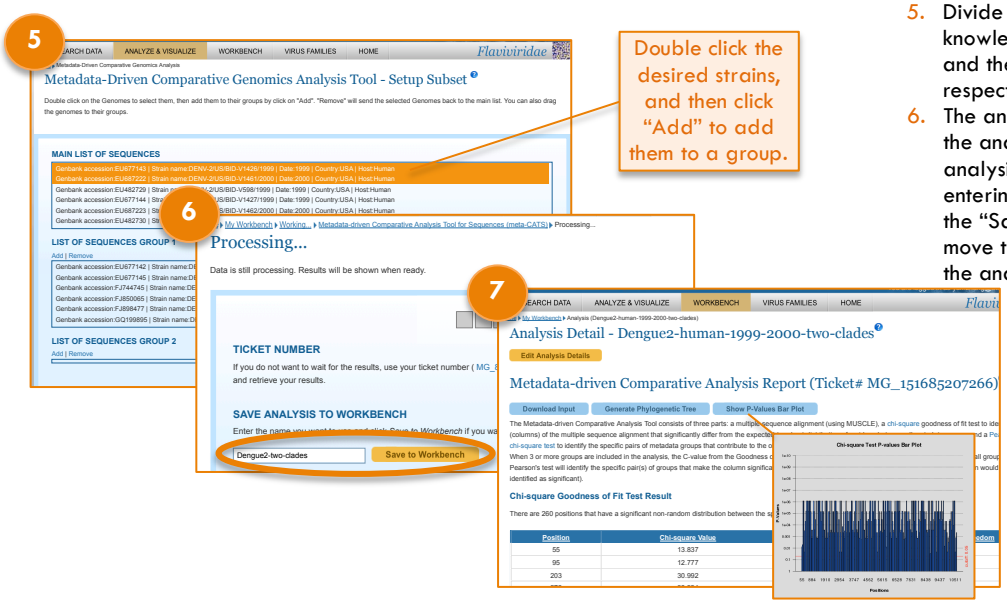
Option 2: Use a working set or upload your own sequences and then run Meta-CATS





1. Mouse-over the "Analyze & Visualize" tab and click "Metadata Sequence Analysis".
2. On the Meta-CATS tool landing page, input your sequences using one of the three options:
   - 2.1 Upload a sequence file and a sequence assignment file.
   - 2.2 Paste sequences in FASTA format.
   - 2.3 Use one or more working sets from your Workbench.
3. Choose the format of sequences provided and sequence type.
4. Choose the number of groups. Adjust the C-value threshold if needed. Then click "Continue".

Divide sequences into groups based on metadata (phenotype) and run Meta-CATS



5. Divide sequences into groups based on scientific knowledge: double click the desired sequences and then click "Add" to add them to their respective groups. When finished, click "Run".
6. The analysis may take a few minutes to run. While the analysis is running, you can choose to save the analysis (upon completion) to your Workbench by entering a name for the analysis and then clicking the "Save to Workbench" button. Then you can move to other parts of the ViPR site, and retrieve the analysis results later from your Workbench.
7. The Meta-CATS report includes:
   - Chi-square Goodness of fit test result: positions that have significant non-random distribution between the specified groups.
   - Pearson's Chi-square Pairwise Comparison Report: positions that significantly differ between the groups.
   - Save the analysis to your Workbench if needed. You can also share the analysis with collaborators.

ViPR is funded by the National Institute of Allergy and Infectious Diseases (NIH / DHHS) under Contract No. HHSN272200900041C and is a collaboration between Northrop Grumman Health IT, University of Texas Southwestern Medical Center and Vecna Technologies. Comments, questions, suggestions? Contact us at feedback@virusbrc.org