HyDRA Web User Guide

National Microbiology Laboratory

Public Health Agency of Canada

2016-8-31

PROTECTING AND EMPOWERING CANADIANS
TO IMPROVE THEIR HEALTH

This application was developed through collaboration between the National Laboratory for HIV Genetics and the Bioinformatics Core at the National Microbiology Laboratory, Public Health Agency of Canada. The project was funded through the federal Genomics Research and Development Initiative.

Publication Notes:

HyDRA Web User Guide

Prepared by: Mikaela Nykoluk & Tracy Taylor

Approved by: Hezhao Ji & Eric Enns

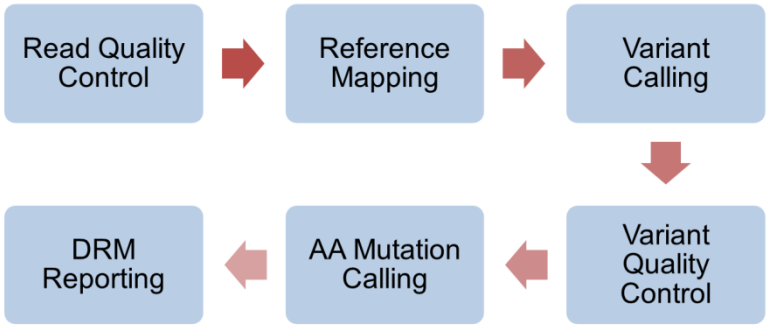
Date: Aug 31, 2016

Table of Contents

1. Introduction to HyDRA and HyDRA Web

HyDRA is an annotated reference-based bioinformatics pipeline scripted in Perl, which analyses next generation sequencing (NGS) data for genotyping HIV-1 drug resistance (HIVDR) mutations[1]. It utilizes an annotated HXB2 sequence (GenBank Accession number: K03455) for reference mapping by Bowtie2[2], and stringent data quality assurance and variant calling criteria to identify HIVDR associated mutations based on the Stanford HIV Drug Resistance Database (http://hivdb.stanford.edu/) and 2009 WHO list for Surveillance of Transmitted HIVDR. All HIVDR mutations found in the *pol* genes; *protease* (PR), *reverse transcriptase* (RT), and *integrase* (IN), are reported according to classifications outlined in the Stanford Surveillance Drug Resistance Mutation list[3,4].

HyDRA Workflow:



HyDRA Web is a freely available web portal that allows for user-friendly access to an automated HyDRA pipeline. Users worldwide can perform HyDRA analysis by simply uploading their raw NGS data to HyDRA Web, and defining analysis parameters. HyDRA Web accepts data from major NGS platforms, primarily Illumina MiSeq, as well as Roche 454, and Ion Torrent from ThermoFisher. Stringent default analysis parameters are described in this document and fully customizable to satisfy varied users' needs. For registered users, data is processed and results are stored on the HyDRA server, retrievable at any time for up to 90 days. Analysis results available for download include: an HIVDR report, consensus sequence at a user-defined threshold, complete NGS data assembly file (BAM), as well as a comprehensive variant call file and amino acid mutation report.

Requiring no advanced bioinformatics skills or high capacity computing equipment, HyDRA Web enables end-users worldwide to easily conduct large scale analysis of NGS data for HIV-1 drug resistance genotyping.

2. Terms and Conditions

THE HyDRA WEB APPLICATION IS OWNED BY HER MAJESTY THE QUEEN IN RIGHT OF CANADA, AS REPRESENTED BY THE MINISTER OF HEALTH, ACTING THROUGH THE PUBLIC HEALTH AGENCY OF CANDA ("PHAC") AND IS PROTECTED UNDER COPYRIGHT LAWS AND INTERNATIONAL COPYRIGHT TREATIES. THE HyDRA WEB APPLICATION IS PROVIDED UNDER THE FOLLOWING LICENSE WHICH DEFINES WHAT YOU MAY DO WITH THE PRODUCT AND CONTAINS LIMITATIONS ON WARRANTIES AND/OR REMEDIES.

IMPORTANT: CAREFULLY READ THIS LICENSE BEFORE USING THIS PRODUCT. USING THIS PRODUCT IN ANY WAY INDICATES YOUR ACKNOWLEDGMENT THAT YOU HAVE READ THIS LICENSE AND AGREE TO ITS TERMS.

1. LICENSE: PHAC ("we" or "us") provides the user ("you") with access to the HyDRA Web Application that includes electronic components, data, and printed materials (collectively called the "Product") and grants you a limited, non-exclusive license to use the Product in accordance with the terms of this License Agreement. The copyright and all other rights to the Product shall remain with us.

2. YOU MUST NOT: a) make copies of the Product; b) modify the Product or make derivative works; c) translate, reverse engineer, decompile, or disassemble the Product; d) rent, lease loan, sub-license, distribute, assign, or transfer the Product or any code for the product; or e) modify the Product or merge all or any part of the Product with another program.

3. TERM: This License Agreement shall continue for as long as you use the Product. However, this License Agreement will terminate automatically without notice from PHAC if you fail to comply with any of its terms or conditions. You agree, upon termination, to destroy the Product and any copies that have been produced in violation of this License Agreement. The Limitations of Warranties and Liability set out below shall continue in force even after any termination.

4. PRIVACY: You agree not to record, place, or otherwise include any personal identifiers or other information that could identify an individual in the raw sequencing data or metadata, on the HyDRA Web Application. We agree to refrain from accessing your records, unless required to do so for the purpose of maintaining the Application. You agree that we may remove any data from the Application where said data has remained on the Application for a period of 90 days.

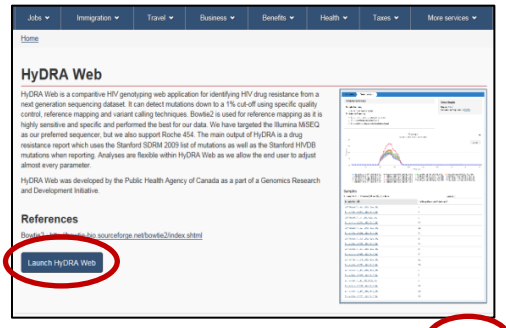
5. LIMITATION OF WARRANTIES AND LIABILITY: THE PRODUCT IS PROVIDED ON AN "AS IS" BASIS, WITHOUT ANY WARRANTIES OR CONDITIONS, EXPRESSED OR IMPLIED, INCLUDING BUT NOT LIMITED TO, ANY IMPLIED WARRANTY OF MERCHANTABILITY OR FITNESS FOR A PARTICLULAR PURPOSE, OR ANY WARRANTY ARISING BY LAW, STATUTE, USAGE OF TRADE, OR COURSE OF DEALING. YOU ASSUME ALL RESPONSIBILITY AND RISK FOR THE USE OF THIS PRODUCT AND PHAC DISCLAIMS ALL LIABILITY FOR ANY LOSS, INJURY OR DAMAGE RESULTING FROM USE OF THIS PRODUCT, WHETHER DIRECT OR INDIRECT, AND WHETHER OR NOT PHAC HAS BEEN ADVISED OF OR HAS KNOWLEDGE OF THE POSSIBILITY OF SUCH LOSS, INJURY OR DAMAGE . IN NO EVENT SHALL PHAC HAVE ANY LIABILITY TO YOU OR ANY OTHER PERSON OR ENTITY FOR ANY DIRECT, COMPENSATORY, INDIRECT, INCIDENTAL, SPECIAL, CONSEQUENTIAL OR EXEMPLARY DAMAGES WHATSOEVER, INCLUDING, BUT NOT LIMITED TO, LOSS OF REVENUE OR PROFIT, LOST OR DAMAGED DATA OR OTHER COMMERICAL OR ECONOMIC LOSS, EVEN IF WE HAVE BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES, OR THEY ARE FORESEEABLE. WE ARE ALSO NOT RESPONSIBLE FOR CLAIMS BY A THIRD PARTY. OUR MAXIMUM AGGREGATE LIABILITY TO YOU SHALL NOT EXCEED THE AMOUNT PAID BY YOU FOR THE PRODUCT. THE LIMITATIONS IN THIS SECTION SHALL APPLY WHETHER OR NOT THE ALLEGED BREACH OR DEFAULT IS A BREACH OF A FUNDAMENTAL CONDITION OR TERM OR A FUNDAMENTAL BREACH.

6. GENERAL: This License Agreement is the entire agreement between us and you, superseding any other agreement or discussions, oral or written, and may not be changed except by a signed agreement. This License Agreement shall be governed by and construed in accordance with the laws of the Province of Ontario, Canada. If any provision of this License Agreement is declared by a court of competent jurisdiction to be invalid, illegal, or unenforceable, such a provision shall be severed from the License Agreement and the other provisions shall remain in full force and effect.

3. Launching HyDRA Web

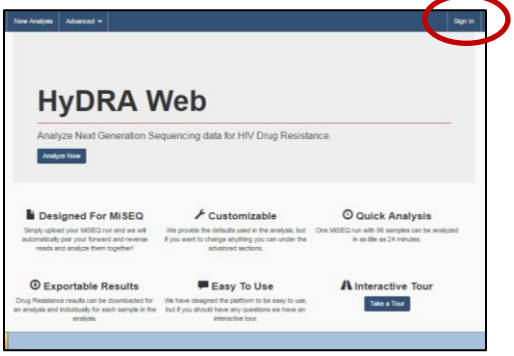
- In your browser, go to http://hydra.canada.ca

- HyDRA Web is a Government of Canada website available in both official languages. Select "**English**" or "**French**" to proceed.

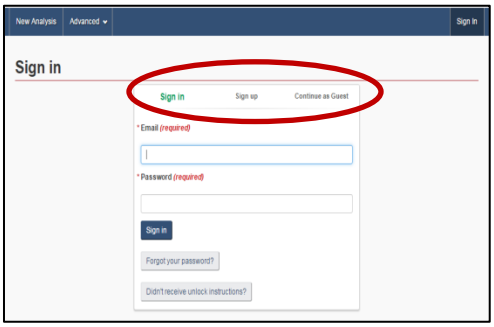


- Click "**Launch HyDRA Web**" to be directed to the HyDRA Web home page.

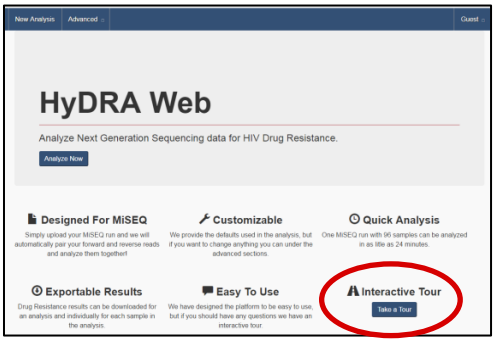
- To use HyDRA Web, a user must **Sign in**, **Sign up**, or **Continue as Guest**. Click "**Sign in**" in the menu bar to open the **Sign in** window. Attempting to access any HyDRA Web applications will also prompt this window to open.



- "**Continue as Guest**" allows the user full access to HyDRA Web; however, the analyses and derived results will NOT be retrievable upon exiting the program, unless an email notification is requested, which will then allow results to be accessible for up to 90 days.



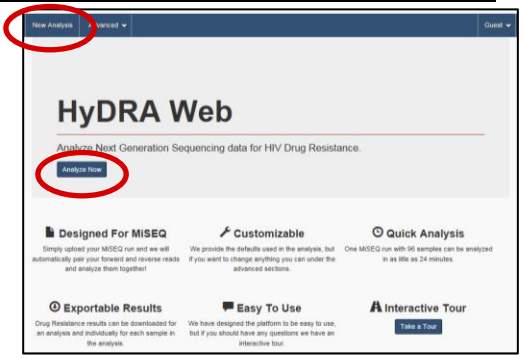
- It is recommended that new user select the "**Sign up**" option to create a new personal user account, which requires only a valid email address and user-defined password.
  - *A registered user is able to store and retrieve previous analyses (for at least 90 days) when signed in.*
  - *Your email will not be distributed.*

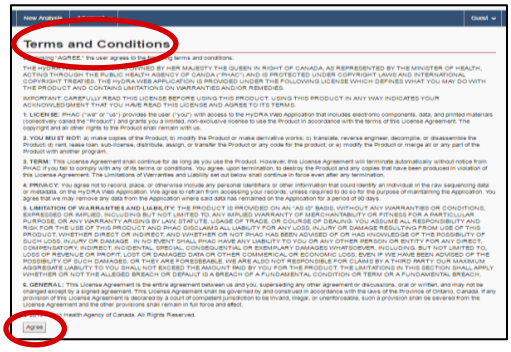


- If you already have a user account, proceed directly to "**Sign In**". Click "**Forgot your password?**" to receive instructions by email for re-setting a new password.

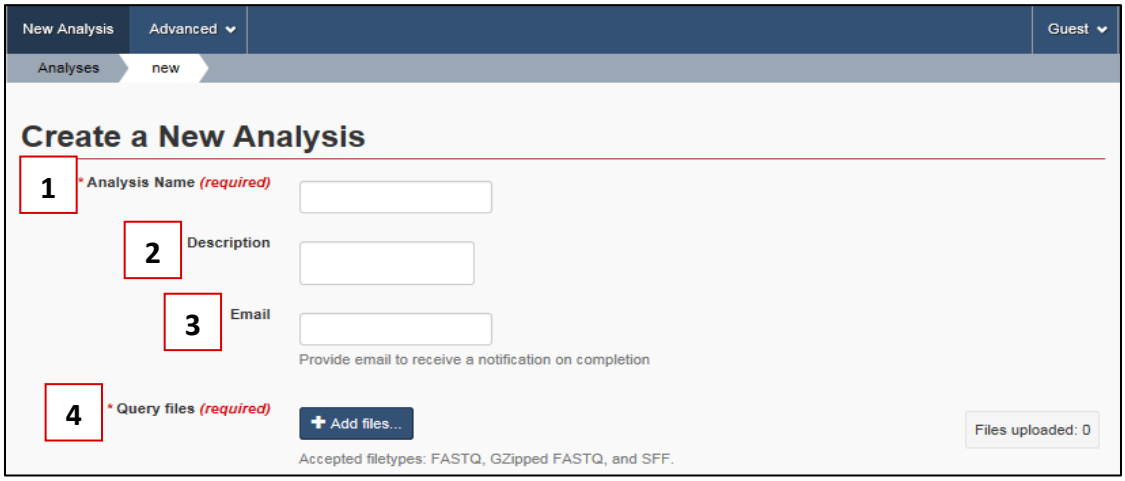
- For a brief introduction and walk-through of the analysis features of HyDRA Web, click "**Take a Tour**" at the bottom of the home page. A pop-up window will navigate step-by-step through a HyDRA Web analysis.

4. Initiating a new HyDRA Web analysis





- To initiate a new analysis, click "**New Analysis**" on the menu taskbar or "**Analyze Now**" on the homepage.
- Note, you will always be asked to agree to the **Terms and Conditions** prior to starting a new analysis. Please read and understand the terms before clicking "**Agree**" at the bottom-left corner to proceed.
- You will then be prompted to a "**Create a New Analysis**" window shown as below.



1 **Analysis Name:** Provide a unique name to identify the analysis. It will be saved in the "**Analyses**" tab under this name.

2 **Description:** Provide additional information about the analysis. (optional)

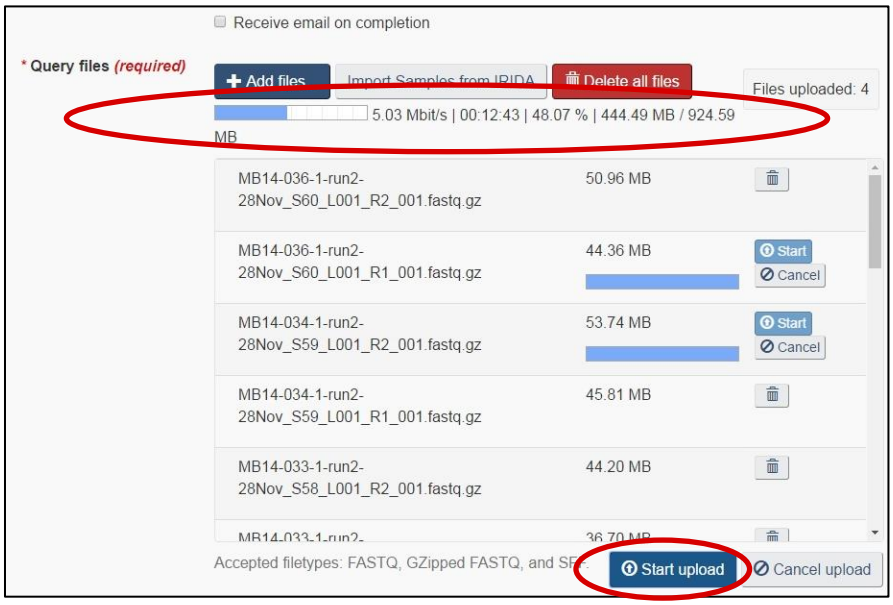
NOTE: *Do not use any personal or patient information in the analyses names or descriptions.*

3 **E-mail:** You may provide an e-mail address in order to be notified when the analysis is finished processing. (optional)

4 **Query files:** Click on "**+ Add files…**" to import sample files for analyses. The files must be in a **FASTQ**, **GZipped FASTQ**, or **SFF** format.

- **Uploading MiSeq Files:** HyDRA automatically pairs MiSeq R1 and R2 data, therefore it is recommended not to alter the naming of the sequence files as they are produced by the MiSeq platform. Click "**+Add files**" to provide sample files for the analysis.

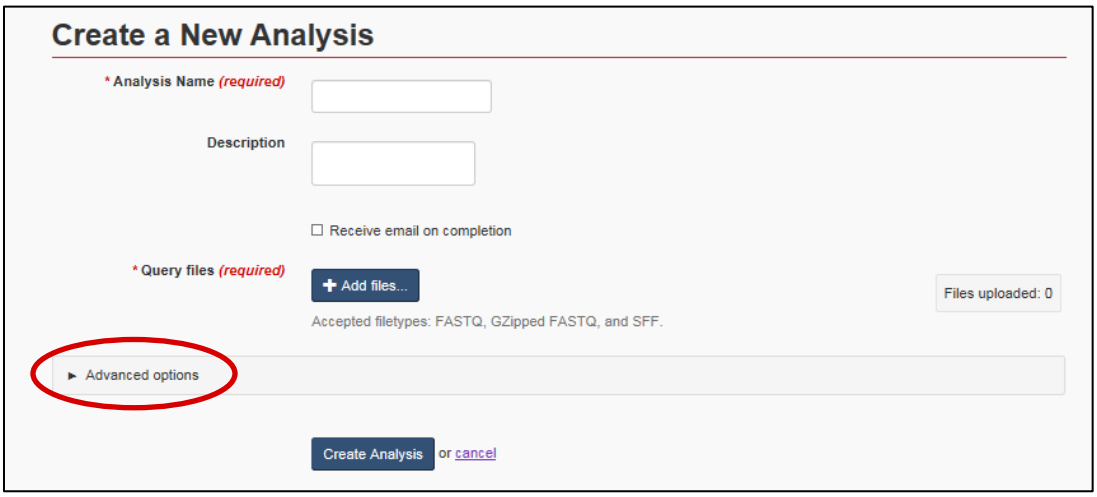
  - Files can be added or deleted at any time while on this page. There is no maximum to the amount of files that can be added – however, more files will increase processing time.
  - Once all files are selected, click "**Start Upload**" to begin uploading sample data to HyDRA. A status bar will appear showing the progress of the uploading.



- If you do not wish to adjust the default analysis settings, you may now click "**Create Analysis**" to begin processing. To alter any parameter setting, see section 5 - **Advanced Options**.

5. Advanced Options

- Click the "**Advanced Options**" bar in the "**Create a New Analysis**" page to open the drop down section for defining various advanced parameters of the analysis.

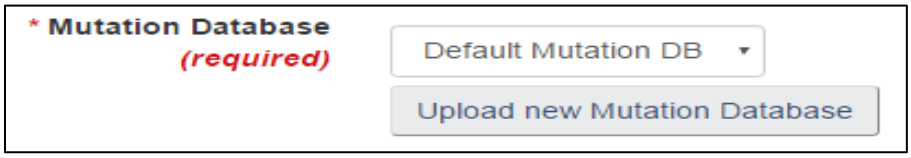


- The following parameters may be modified through "**Advanced Options**":

(1) Consensus Percent

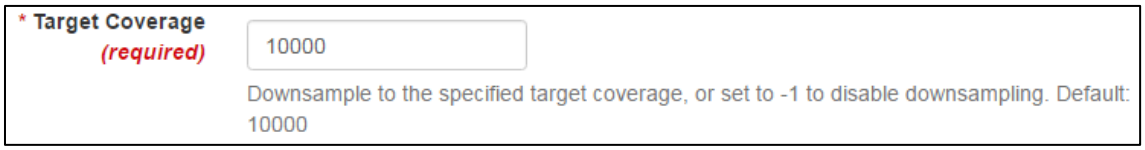
This parameter sets the minimum frequency required for a mixed base call to be included in the consensus sequence generated from HyDRA analysis. The default is set to **20%** (maximum setting) to approximate Sanger sequencing data. Any variants that are present at a lower frequency will NOT appear in the final consensus sequence.

(2) Mutation Database



The default mutation database used for reporting HIVDR associated mutations is based on the Stanford HIV Drug Resistance Database (http://hivdb.stanford.edu/) with additional annotations from the 2009 WHO SDRM list. See page 20 of this User Guide for instructions on how to upload a new mutation database.

(3) Target Coverage



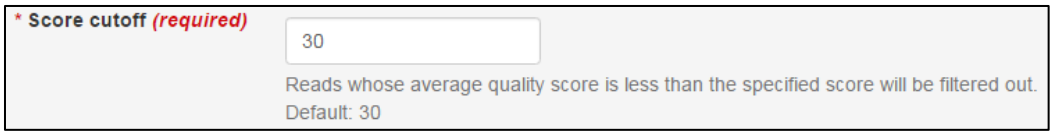
Coverage is defined as the average number of reads representing a given nucleotide in the assembled sequence alignment. The default setting is **10,000** reads. This maximizes the time-efficiency of the analysis while retaining its sensitivity for low frequency variants. The data is randomly down-sampled to the target coverage before mapping, but after read quality and length filtering. Entering the value "**-1**" will disable the down-sampling function and allow all reads to be included in the analysis.

- **Note**: HyDRA assumes you are analyzing a ~3kb of HIV pol gene (PR, RT, and IN). In the event that you are analyzing only a partial *pol* gene sequence, you will see proportionally higher coverage for that analysis. E.g. ~30,000 coverage over a 1kb sequence analysis (PR and RT only)

(4) Filtering: Options under this heading are used to filter out MiSeq reads that do not meet the specified criteria.

(a) **Reads:**

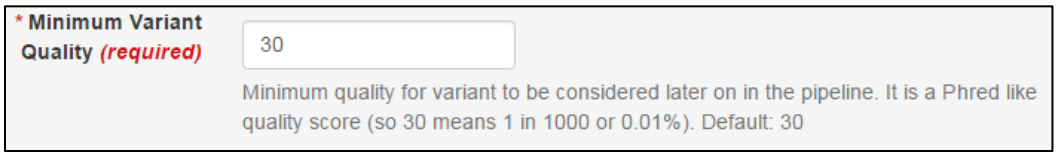
Length Cutoff: Reads which fall short of the specified length will be filtered out and excluded from the analysis. The default is set to **100** nucleotides. Short reads are generally more difficult to align, and may include artifacts such as primer dimer library sequences which can slow down processing.



Score Cutoff: Reads whose average quality score across all loci is less than the specified score will be filtered out. The default cutoff is **30**. This is a Phred-like quality score, i.e. 30 is defined as a 1 in 1,000 (0.1%) chance of an error in base call.

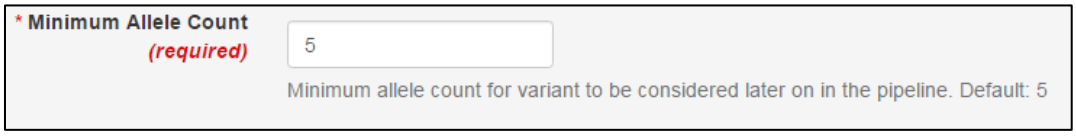
(b) **Variants:** Options under this heading are used as quality control for variant calls.

Error Rate: The error rate is set for a specific sequencing platform. The default is set to the Illumina MiSeq platform error rate of **0.0021** (**0.21%**) as determined at the NLHG[6]. This value is used as input in the Poisson distribution, which is subsequently used to assign a quality score to each variant identified.



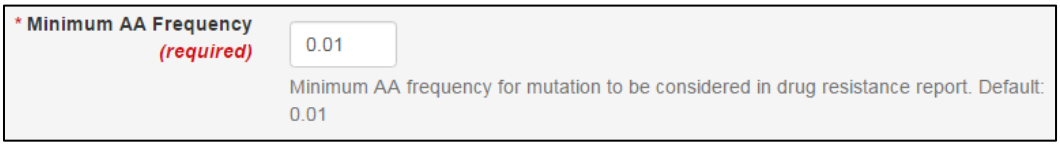
Minimum Variant Quality: This setting is the minimum quality needed for a variant to be considered later on in the pipeline. Default setting is **30**. This is a Phred-like quality score assigned to each variant base call in the read to determine the likeliness that the call was correct.

Minimum Read: This setting is the minimum read depth at a specific locus to be considered later on in the pipeline. Default is set to **100** reads. Increased read depth lends stronger support to the variant being a true mutation and not an error in variant calling.



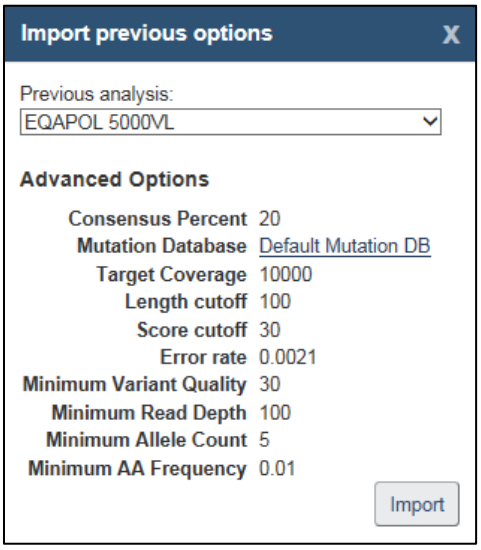
Minimum Allele Count: The minimum allele count needed for a variant to be considered later on in the pipeline. Default is set to **5**, meaning the number of reads with the variant at a specific locus has to be at least 5 in order for the variant to be called. It follows that the minimum variant frequency that can be called for a site within a region of only 100 reads would be 5%.

(c) **Mutations:** Reporting threshold for amino acid substitutions.



Minimum AA Frequency: The minimum amino acid (AA) frequency needed for a mutation to be considered in the drug resistance report. The default is set to **0.01** (**1%**). Amino acid mutations lower than the set frequency will not be shown in the final HyDRA report.

- You can reset all the **Advanced Options** to default by clicking "**Reset to default**".
- If there is any problem or missing information required to "**Create a New Analysis**", HyDRA will prompt you to check the sample files and confirm that all required fields are filled out.
- If you wish to view and/or import the options chosen for previous analyses, click "**Import Previous Options**". You will remain on the same page and a pop-up will appear allowing you to scroll through old analyses and view the parameters.



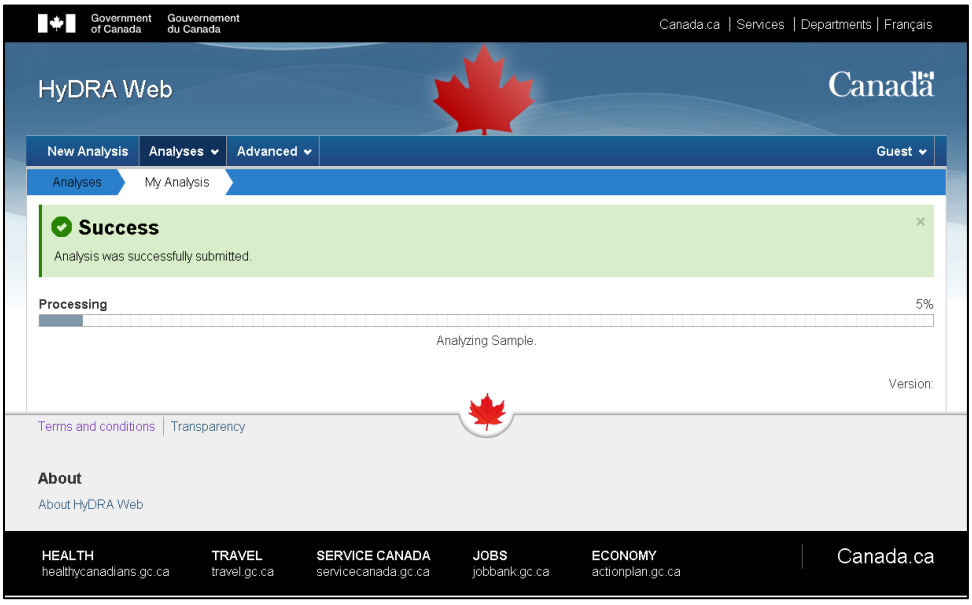
Note: The option to import previous analysis settings is only available to registered users.

- Once the parameters are set and the files are uploaded, click "**Create Analysis**" to begin processing the analysis with the new settings.

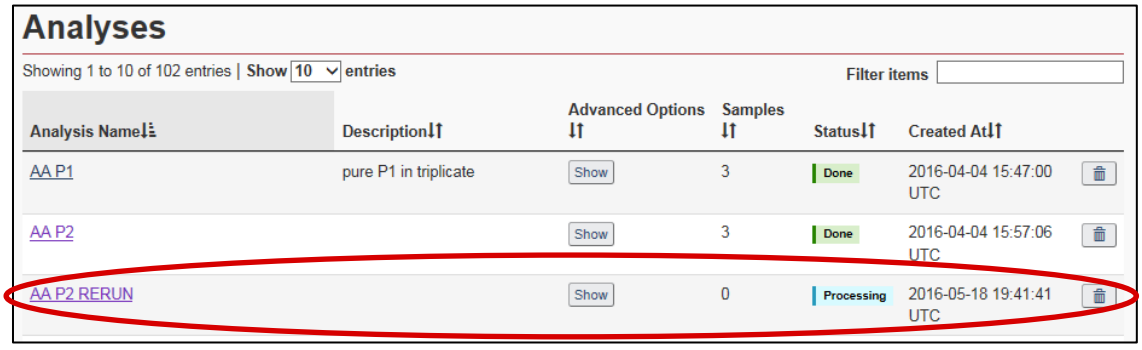
- Clicking "**Cancel**" will delete the uploaded files and advanced option settings and bring you back to the "**Analyses List**" page.

6. Processing

You will be re-directed to an "**Analysis Progress**" page where a progress bar will appear tracking the percentage of progress that has been made in the analysis. Depending on your connection speed and demand on the HyDRA server, the initial submission of the analysis may take some time. Once the data files and analysis parameters are successfully submitted, the analysis will proceed to run on the HyDRA server and no longer requires the user to be connected or logged in. If requested at the time of creating the analysis, an email will be sent to the user upon completion of the analysis. The progress bar will only advance from 5% once a sample has finished being analyzed, so if you are only analyzing one sample the progress bar will not move from 5% until the analysis is completed.



The progress of the run being processed can also be seen in the Analyses List table.



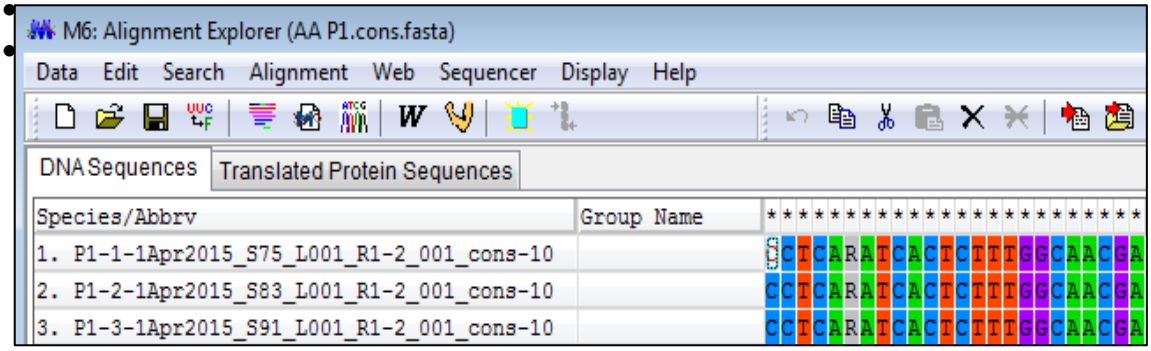
7. Analysis Results

- **Summary:** Statistical overview of total HIVDR mutations identified in the submitted samples.
- "**Resubmit Analysis**" will bring you back to the "**Create New Analysis**" page while keeping the same sample files loaded from the current run. The re-submission will require a new name, and will not save over the current run. You can adjust the options and re-run the current samples, with different analysis parameters as desired, without having to repeat the upload of raw sequence data.
- **Downloads:** Combined exportable results for all samples in the analysis.
    - A summary **Report** of HIVDR mutations identified in each sample can be downloaded in the form of a **CSV** file which can be viewed in Excel. All HIVDR mutations found at a frequency above what was set in the advanced options under **Minimum AA Frequency** (default 1%) are listed in the report.

	A	B	C	D	E	F	G	H	I
1	Sample Name	Gene	Classification	Surveillance	Wildtype	Position	Mutation	Frequency	Coverage
2	P1-1-1Apr2015_S75_L001_R1-2_001	PR	PIMajor	Yes	I	47	A	94.64	24199
3	P1-1-1Apr2015_S75_L001_R1-2_001	PR	PIMajor	Yes	I	47	V	1.3	24199
4	P1-1-1Apr2015_S75_L001_R1-2_001	PR	PIMajor	Yes	I	50	L	95.83	24199
5	P1-1-1Apr2015_S75_L001_R1-2_001	PR	PIMinor	Yes	G	73	T	1.36	24200
6	P1-1-1Apr2015_S75_L001_R1-2_001	PR	PIMajor	Yes	V	82	C	1.07	44189
7	P1-1-1Apr2015_S75_L001_R1-2_001	PR	PIMajor	Yes	V	82	F	2.32	44189
8	P1-1-1Apr2015_S75_L001_R1-2_001	RT	Other	No	I	132	L	1.13	28054

- The report includes additional information on the gene location of the mutations (PR, RT, or IN), the classification and surveillance status of the mutations, the position and identity of wild type and mutant amino acids, the variant frequency and depth of coverage at those sites.

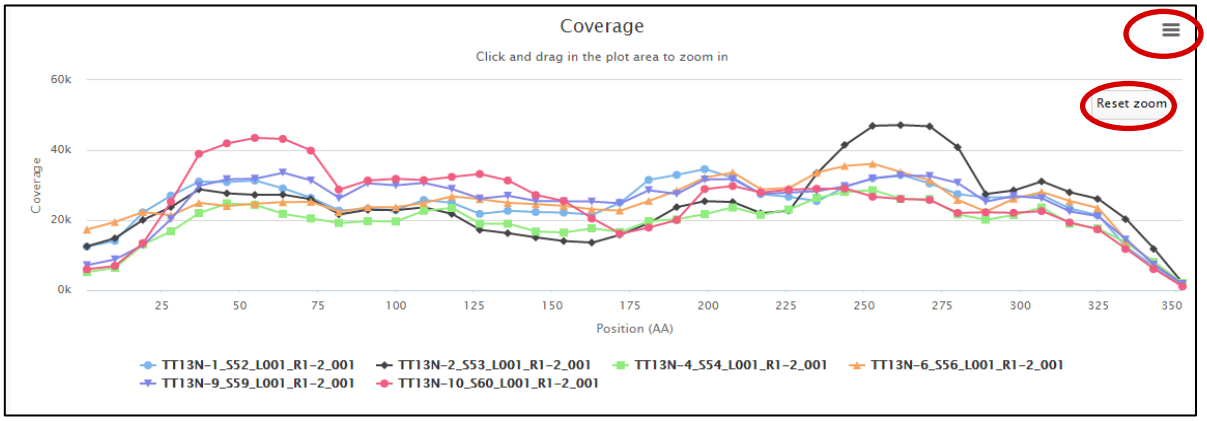
- A **FASTA** file can also be downloaded which contains a consensus sequence for each sample in the analysis.



- The consensus sequences mimic those obtained from Sanger sequencing, and will contain mixed bases present at a frequency set by the predefined "**Consensus Percent**" cutoff (see page 6).

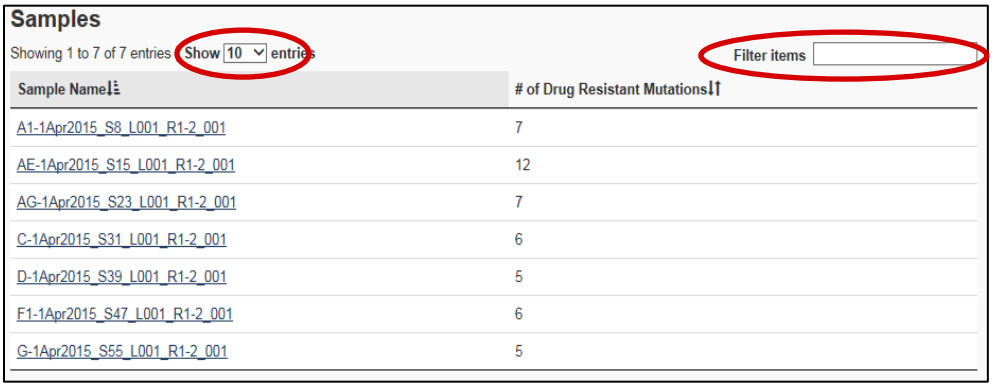
➢ **Coverage**

- The Coverage graph is a visual display of the depth of reads covering each **amino acid (AA) position** as shown on the X-axis. The Coverage graph will reflect any downsizing, set by default to 10,000 nucleotides across the 3kb reference (see **Target Coverage** page 7).

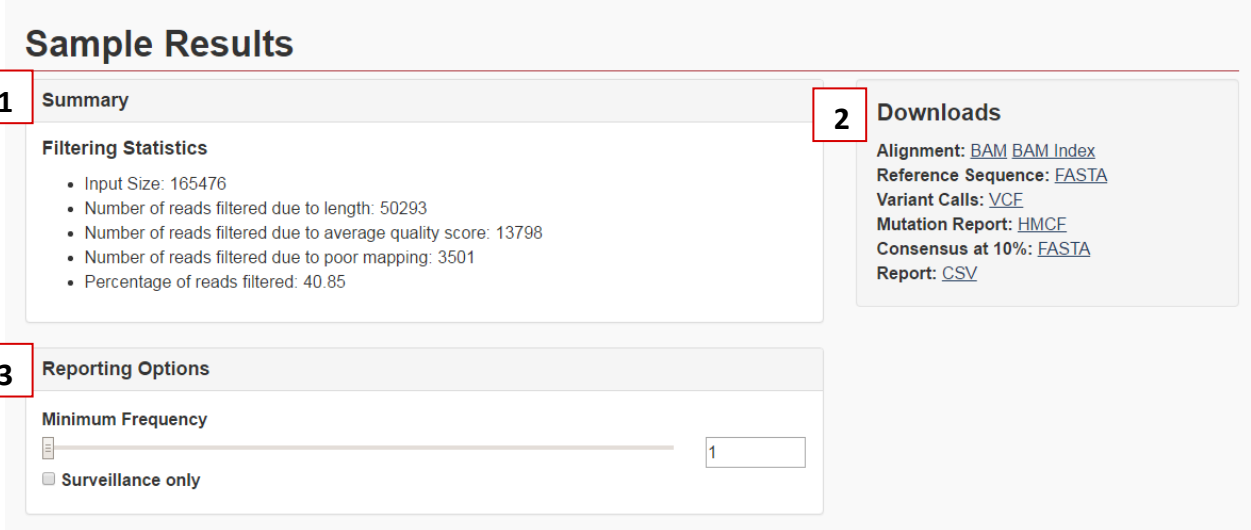


- To zoom into a specific area, click and drag the mouse over the area you would like to enlarge. Click "**Reset Zoom**" to return to the full graph.
- Click on a sample name in the legend to hide that sample from the coverage graph. The sample name in the legend will turn grey if the sample has been excluded. Click on the sample name again to add it back.
- Hovering the mouse over a plot point on the graph will display the amino acid position, sequence file name, and depth of coverage at that position.
- Click the ☰ button in the top right corner of the coverage graph for image downloading options (print graph or download in PNG, JPEG, PDF, or SVG format).

- **Samples:**
  - The table lists all samples which were run in the analysis and includes the number of drug resistant mutations found in each sample.
  - The number of samples shown per page can be adjusted to show up to 100 entries, or samples per page.
  - Entries can be filtered or searched by sample name or number of mutations by using the **Filter items** text box in the top right corner.
  - Samples can be sorted chronologically by name or by the number of drug resistant mutation by clicking on the respective heading.
  - Click on a sample to view **Sample Results**, a more detailed analysis on that sample.



- **Sample Results:** Selecting a sample from the table will bring you to a new page "**Sample Results**", with more detailed information about the analysis for that individual sample.



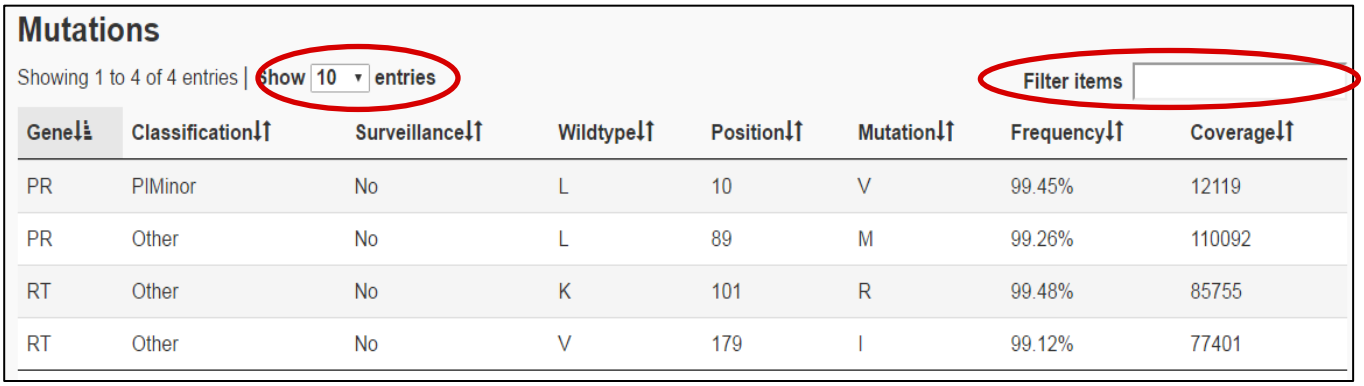
1 **Filtering Statistics:** Summarizes analytical statistics for that sample including:

- **Input Size** : number of reads originally uploaded from the NGS platform
- **Number of reads filtered due to length**: reads removed that fell below the minimum 100 nucleotide cutoff
- **Number of reads filtered due to average quality score**: reads removed due to average quality score below Q30
- **Number of reads filtered due to excess coverage**: reads removed once 10,000x coverage is achieved across the 3kb reference sequence (or ~30,000x over a 1kb amplicon)
- **Number of reads filtered out due to poor mapping**
- **Percentage of reads filtered**: overall number of reads removed from the analysis
  - *default filter settings are adjustable in the advanced options, see pages 6-8.*

2 **Downloads:** See the section "**Exportable Results for Individual Samples"** on page 16 for a complete description of each output file available for download.

3 **Reporting Options:** Click and drag the toggle to change the minimum frequency of mutations reported in the **Mutations table**. Mutations that fall below the frequency will be temporarily removed from view. Check the "**Surveillance only**" box to include only mutations listed as Stanford SDRM2009 list in the table.

- **Mutations:** The **Mutations** table displays detailed information on the HIVDR mutations identified in the individual sample.
  - The table includes the same data that is found in the downloadable **Report** (.csv); including gene location of the mutations (PR, RT, or IN), the classification and surveillance status of the mutations, the position and identity of wild type and mutant phenotypes, the variant frequency and depth of coverage of those sites.



Mutations

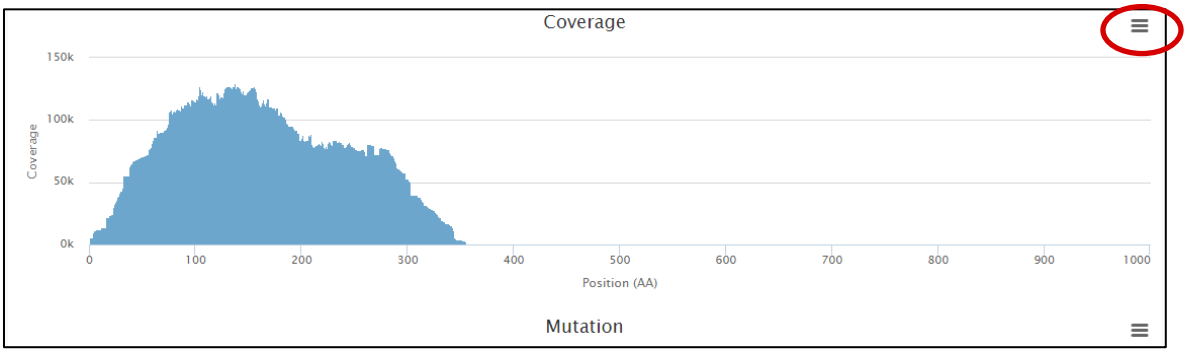
Showing 1 to 4 of 4 entries | Show 10 entries

Filter items

Gene	Classification	Surveillance	Wildtype	Position	Mutation	Frequency	Coverage
PR	PIMinor	No	L	10	V	99.45%	12119
PR	Other	No	L	89	M	99.26%	110092
RT	Other	No	K	101	R	99.48%	85755
RT	Other	No	V	179	I	99.12%	77401

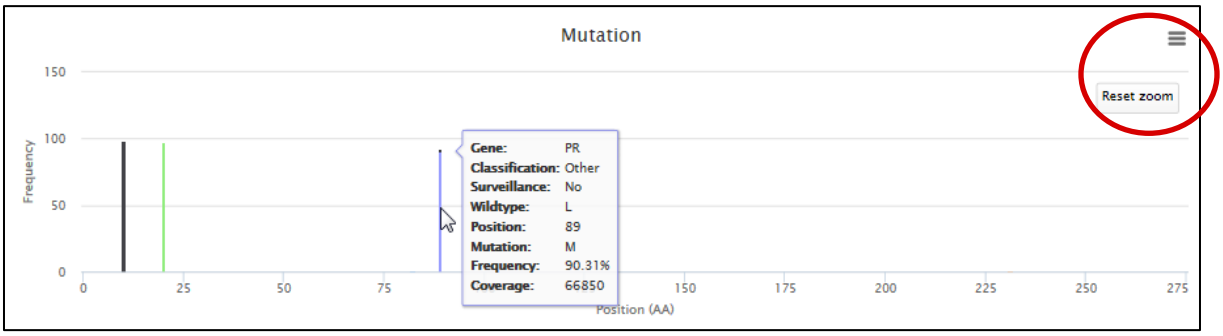
- The number of mutations listed per page can be adjusted in the **Show entries** box.
- **Filter items** box can be used to filter or search based on any criteria.
- The table can be chronologically sorted by toggling any of the headings in the desired column.

➢ **Coverage Graph:** The coverage graph is a visual display of the depth of reads covering each amino acid position in the designated sample. The AA Position on the X-axis is numbered starting from AA position number 1 in the Protease gene, and follows through to include amino acid positions in the RT and IN genes.



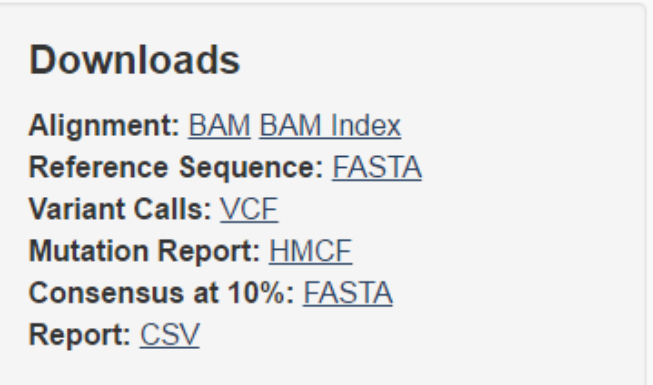
- Click the ☰ button in the top right corner of the coverage graph for image options (print graph, or download in PNG, JPEG, PDF, or SVG formats).
- Click and drag the mouse over an area to zoom into it. Click "**Reset zoom**" to zoom out to the original graph. Zooming into the coverage graph will simultaneously zoom into the same area on the **Mutation** graph.
- As the mouse is scrolled over an area on the graph, additional information will appear including the amino acid position and read coverage at that position.

➢ **Mutation Graph:** The Mutation graph is a visual display of the position and frequency of each mutation that was identified in the specified sample.



- Click the ☰ button in the top right corner of the graph for image options (print graph, or download in PNG, JPEG, PDF, or SVG format).
- Click and drag the mouse over an area to zoom into it. Click "**Reset Zoom**" to zoom out to the full graph. Zooming into the mutations graph will simultaneously zoom into the same area on the **Coverage** graph.
- When the mouse is hovered over an individual vertical bar a pop-up window will display the same information that is provided in the Mutation table for that particular position.
- If multiple mutations are present at a single site; they will appear stacked in the graph.

➢ **Exportable Results for Individual Samples (Downloads):** Various files can be downloaded from the **Sample Results** page for the analysis of an individual sample.



(1) Alignment: BAM, BAM Index

- Both files, along with the HXB2 reference sequence, are required for viewing the complete assembly and alignment of reads in a graphical viewer such as Tablet (*https://ics.hutton.ac.uk/tablet/)* or IGV (http:/software.broadinstitute.org/software/igv/).

(2) Reference Sequence: FASTA

- HXB2 *pol* gene sequence (GenBank #: K03455) used for reference mapping.

(3) Variant Calls: VCF

- Once downloaded, the *.vcf file can be opened in Excel by following steps below: Start the Excel software, click "**File**" then "**Open**", and change file types options from "**All Excel Files**" to "**All Files**" in the bottom right hand corner in the file open dialog, browse to and select the downloaded VCF file and click "**Open**". In the resulting "**Text Import Wizard**" pop-up window, under "**Choose file type that best describes your data**", click to select "**Delimited**", and then click "**Next**". In the following window, check "**Tab**" box under "**Delimiters**"", click "**Next**", under "**Column data format**", check "**General**", and then "**Finish".**

An example VCF file and the description of the file contents are as below:

	A	B	C	D	E	F	G	H	I	J
1	##fileformat=VCFv4.2									
2	##fileDate=20160608									
3	##source=HyDRA									
4	##INFO=<ID=DP,Number=1,Type=Integer,Description="Total Depth">									
5	##INFO=<ID=AC,Number=A,Type=Integer,Description="Allele Count">									
6	##INFO=<ID=AF,Number=A,Type=Float,Description="Allele Frequency">									
7	##FILTER=<ID=q30,Description="Quality below 30">									
8	##FILTER=<ID=ac5,Description="Allele count below 5">									
9	##FILTER=<ID=dp100,Description="Read depth below 100">									
10	#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO		
11	hxb2_pol	6	.	g	a	100	PASS	DP=842;AC=466;AF=0.5534		
12	hxb2_pol	7	.	g	a	100	PASS	DP=842;AC=841;AF=0.9988		
13	hxb2_pol	12	.	t	a	1	q30;ac5	DP=1800;AC=3;AF=0.0017		
14	hxb2_pol	12	.	t	g	32	PASS	DP=1800;AC=12;AF=0.0067		
15	hxb2_pol	14	.	t	a	1	q30;ac5	DP=1908;AC=3;AF=0.0016		
16	hxb2_pol	14	.	t	g	100	PASS	DP=1908;AC=32;AF=0.0168		
17	hxb2_pol	15	.	t	g	9	q30	DP=1921;AC=7;AF=0.0036		
18	hxb2_pol	17	.	g	t	2	q30;ac5	DP=1998;AC=4;AF=0.0020		
19	hxb2_pol	25	.	c	a	3	q30	DP=2118;AC=5;AF=0.0024		

- **#CHROM:** the reference sequence used for analysis
- **POS:** the position of the variant nucleotide in the sequence
- **REF:** the "wildtype" nucleotide found in the **reference** sequence at that position
- **ALT:** the "variant" nucleotide found in the **sample** sequence at that position
- **QUAL:** the quality score assigned to the variant base call
- **FILTER:** the variant nucleotide (ALT) will only **PASS** filter if it meets the requirements set in the analysis options. If the variant does **NOT** PASS filter the reason for filtering it out of the final analysis is given;
    - **q30** – variant quality is < 30
    - **ac5** – variant is present in < 5 alleles (reads)
    - **dp100** – depth of reads is < 100 at the variant's position
- **INFO:** statistics for each position
    - **DP** – read depth at that position
    - **AC** – allele count, the number of times the variant is detected in all reads
    - **AF** - allele frequency, the percentage of reads containing the variant

(4) Mutation Report: HMCF

- The *.hmcf file contains **ALL amino acid (AA) variations** identified in the analysis, including those that did **NOT** pass filter. For those variants that failed, the criterion that was **NOT** met is given.
- Similar to the *.vcf file, the *.hmcf file can be downloaded and then opened in Excel by following the same instructions as given above *(see to Page 16)*.

An example HMCF file and the description of the file contents are as below:

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	##fileformat=HMCFv1												
2	##fileDate=20151104												
3	##source=HyDRA												
4	##reference=/var/www/hydra-web/releases/20151103173244/lib/hydra/var/hxb2_pol.fas												
5	##INFO=<ID=MC,Number=.,Description="String">												
6	##INFO=<ID=MCF,Number=.,Description="String">												
7	##INFO=<ID=WC,Number=.,Description="String">												
8	##FILTER=<ID=mf0.01,Description="Mutant freq below 0.01">												
9	#GENE	CATEGORY	SURVEILLANCE	TYPE	WILDTYPE	POS	MUTANT	FILTER	MUTANT_FREQ	COVERAGE	INFO		
10	PR	.	.	mutation	L	10	A	mf0.01	0.0056	6816	WC=ctc;MC=GCT;MCF=0.0056		
11	PR	PIMinor	No	mutation	L	10	V	PASS	0.9875	6816	WC=ctc;MC=GtT,Gtc;MCF=0.9		
12	RT	.	.	mutation	P	4	L	mf0.01	0.0077	91388	WC=cct;MC=cTt;MCF=0.0077		
13	RT	.	.	mutation	V	10	I	mf0.01	0.0071	85971	WC=gta;MC=Ata;MCF=0.0071		

- **GENE:** location of the AA variant (PR, RT, or IN gene)
- **CATEGORY:** the drug resistance classification of the AA substitution as defined in the Stanford HIV database (blank if variant is not associated with HIVDR)
- **SURVEILLANCE:** surveillance classification (yes or no) of the AA variant, according to the Stanford SDRM 2009 list
- **TYPE:** will always read "**mutation**"
- **WILDTYPE:** the wildtype AA at that position, according to the HXB2 reference sequence translation
- **POS:** the AA position
- **MUTANT:** the AA substitution or "mutant" identified at that position
- **FILTER:** the AA variant (MUTANT) will only **PASS** filter if it meets the minimum AA frequency set in the analysis options, by default this is 1%. Other AA variants present at < 1% are indicated by **mf0.01**.
- **MUTANT_FREQ:** the frequency of the detected mutation
- **COVERAGE:** the number of reads at that amino acid position
- **INFO: WC** "wildtype codon", **MC** "mutated codon", **MCF** "mutated codon frequency"

(5) Consensus at 10%: FASTA

- This is a single FASTA consensus sequence that can be manipulated similarly to a Sanger-derived sequence. In this example, the consensus threshold for inclusion of mixed base calls was modified to 10% in the advanced options (default is 20%).

(6) Report: CSV

- The drug resistance report available here is identical to the one that can be downloaded from the **Analysis Results** view (see page 11), except that it will only contain data for the individual sample currently selected.

8. Retrieving previous analyses

To view past analyses, go to "**List Analyses**" in the drop-down menu under the "**Analyses**" tab in the menu taskbar.

By default, the 10 most recent analyses are displayed that were run under the current user logged in. Registered users can retrieve analyses for up to 90 days.

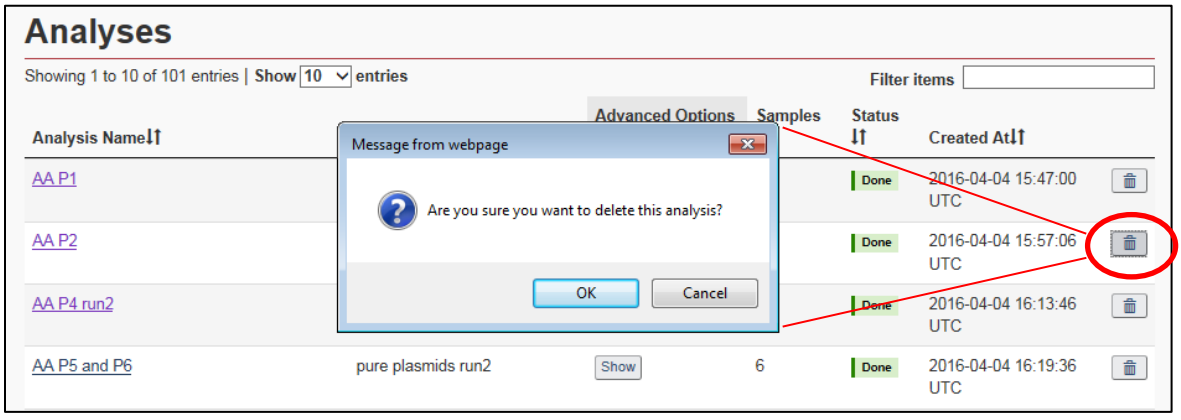
The table can be chronologically sorted according to any column by clicking on that respective heading.

Use "**Show entries**" or "**Filter items**" to view or search your analyses with ease.

In the "**Advanced Options**" column, click "**Show**" to view a quick reference of the optional parameter settings used in the corresponding analysis, including:

- Consensus Percent
- Mutation Database
- Target Coverage
- Length cutoff
- Score cutoff
- Error rate
- Minimum Variant Quality
- Minimum Read Depth
- Minimum Allele Count
- Minimum AA Frequency

To delete a completed analysis, click the **trash can** button in the analysis row. After clicking this button you will be prompted with a message window asking for confirmation to delete the analysis. Clicking "**OK**" will permanently delete it.



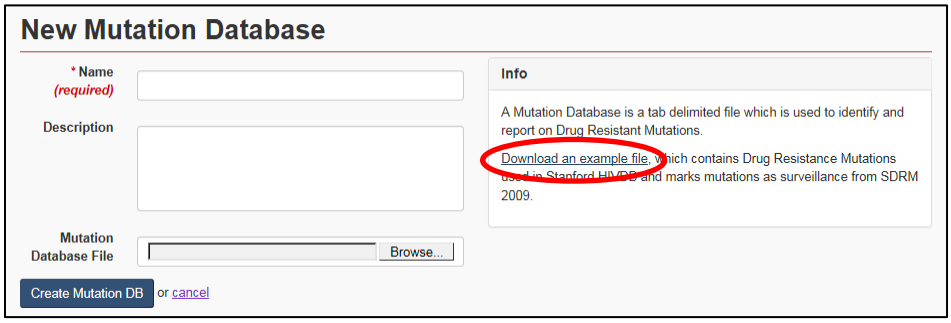
9. Mutation Database

The mutation database is used to annotate the HXB2 reference sequence and to identify HIVDR mutations in the sample data. The current default database is based on the Stanford HIV Drug Resistance Database (http://hivdb.stanford.edu/) with additional annotations from the 2009 WHO list for Surveillance of Transmitted Drug Resistance Mutations.

To view the mutation databases, click "**Advanced**" on the menu taskbar, and then "**Mutations Databases**".

The default database is listed under the "**Public**" tab. This database cannot be altered. All other databases uploaded and made public by other users of HyDRA will be listed here.

Under the "**User**" tab, click "**New Mutation Database**" to upload a new file. You will be re-directed to a new page where you can enter a name and description for the database, and choose the file (tab delimited format) for upload.



To view an example, click "**Download an example file**" (*.tsv file format), and open in Excel following the instructions provided for *.vcf files (see page 16).

10. User Account

Clicking on the username on the right side of the menu taskbar will bring up the options to either "**Sign Out**", or "**Edit Account**".

Clicking on "**Edit Account**" will open a new page where the account password can be changed or the account can be cancelled. Cancelling an account will also delete all of the analyses stored from that user.

11. Additional Resources

Galaxy: http://galaxyproject.org/

Tablet: http://ics.hutton.ac.uk/tablet/

IGV: http://www.broadinstitute.org/igv/

MEGA: http://www.megasoftware.net/

FastQC: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Bowtie2: http://bowtie-bio.sourceforge.net/bowtie2/index.shtml

Stanford HIV Drug Resistance Database:

http://hivdb.stanford.edu/

http://hivdb.stanford.edu/DR/asi/releaseNotes/index.html#hivdb_mutationcategories

References

(1) Enns, E. *et al*. HyDRA – A novel bioinformatics tool for next generation sequencing-based HIV drug resistance data analysis. Presented at: The Canadian Association for HIV Research 25th Annual Canadian Conference on HIV/AIDS Research; 2016 May 12-15; Winnipeg, MB.
(2) Langmead B, Salzberg S. Fast gapped-read alignment with Bowtie 2. Nature Methods. 2012, 9:357-359.
(3) Bennett, D. E. *et al.* Drug resistance mutations for surveillance of transmitted HIV-1 drug-resistance: 2009 update. *PLoS ONE* 4, e4724 (2009).
(4) Rhee S-Y. *et al*. Human immunodeficiency virus reverse transcriptase and protease sequence database. *Nucleic Acids Res* 31, 298-303 (2003).
(5) Shafer, R. W. Rationale and uses of a public HIV drug-resistance database. *J Infect Dis* 194, S51-80 (2006).
(6) Ji H, et al. Establishment of an Illumina MiSeq-based HIV drug resistance testing platform. Presented at: the 8th International AIDS Society Conference on HIV Pathogenesis, Treatment and Prevention (IAS 2015), July 19~22, 2015, Vancouver, Canada

Contact Us

Questions or issues arising regarding HyDRA Web can be directed to:

BioInfo-Support@phac-aspc.gc.ca