HIVfird

Beta version

USER GUIDE

TABLE OF CONTENTS

About this tutorial

HIVfird is an online tool designed to assist health professionals and researchers to detect HIV-1 mutations that confer resistance to antiretroviral drugs of the fusion inhibitor class. His conception was based on the idea of simplifying as much as possible the handling and being as informative as possible while maintaining the same reliability of a robust literature search.

Organization

This tutorial contains these sections:

- Getting started: How to use HIVfird, data submission and how to read the results
- Under the hood: Principles of HIVfird, properties and functions.
- Troubleshooting: Errors and how to solve.

Assumptions

This tutorial makes some assumptions about you before using HIVfird:

- You must have prior knowledge on molecular biology. Knowing about DNA, RNA, proteins and transcription and translation processes are basic prerequisites before using this tool. If you do not have proper skill about these topics, the help of a specialist will be required.
- It is necessary that you have basic skills in bioinformatics, particularly regarding the use and handling of FASTA files.
- This tool will have a better performance when used on a computer with internet access and an updated browser. Although the site is compatible with mobile phones, difficulties may occur during the handling of the tool during submissions of large datasets.

Getting Started

Section 1 – Home page

The home page is basically split in two side-by-side frames. The frame on the left shows the current stage of system development and general instructions on the use of HIVfird. The table on the right contains the data submission form, with the following items:

- <u>FASTA input.</u> In this field you must enter your HIV DNA sequence in an FASTA format. It is necessary that the file is correctly formatted and that the sequence has at least 36 base pairs. You can either enter a single sequence or multiple sequences.
- <u>Alignment platform.</u> On HIVfird execution, your query sequence is aligned with our reference sequence, and you can choose one of the available platforms.
- <u>Articles mode.</u> You can use two data sources in the search for resistances. By keeping this option selected, HIVfird will search only at previously cataloged papers, otherwise it will search in some websites related to HIV resistance. You can learn more about this option on page 8.
- <u>Reset button.</u> Resets all form values.
- <u>Submit FASTA sequence button.</u> Send the sequence to analysis.

Section 2 – Multiple sequence results page
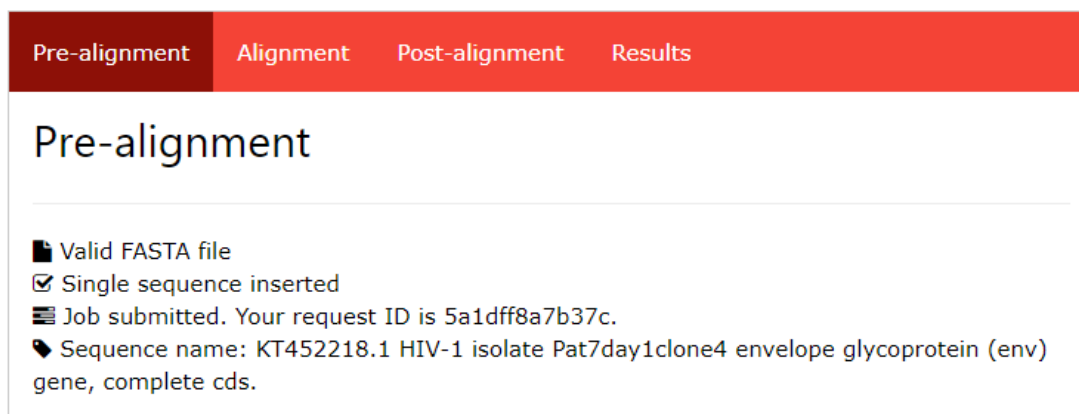
Preparation

- Valid FASTA file
- 7 sequences inserted
- Job submitted successfully

Legend

- ✔ No resistance detected
- ❗ Resistance found
- ❓ Ambiguity codes on sequence
- ✖ Error while analyzing sequence

Statistics

- ✔ 2 sequences found
- ❗ 3 sequences found
- ❓ 1 sequences found
- ✖ 1 sequences found

Download table as CSV

Request ID ▲▼	Sequence name ▲▼	Result ▲▼	Mutation ▲▼	Details
5bfc8ae5b8d96	K03455.1 Human immunodeficiency virus type 1 (HXB2), complete genome; HIV1/HTLV-III/LAV reference genome	✔		Details
5bfc8ae5c9641	FM165628.1 Human immunodeficiency virus 1 env gene for gp160 protein, strain PIC771_15	✔		Details
5bfc8ae5cab51	U36880.1 Human immunodeficiency virus type 1 envelope glycoprotein (env) gene, complete cds	❗	N43D	Details
5bfc8ae5cc18a	KT452218.1 HIV-1 isolate Pat7day1clone4 envelope glycoprotein (env) gene, complete cds	❗	G36S, N43D	Details
5bfc8ae5cd91b	AY768658.1 HIV-1 isolate 16(24) from United Kingdom envelope glycoprotein (env) gene, partial cds	❗	G36S, V38M, N42T, L44M, G36S+L44M	Details
5bfc8ae5ced36	FJ687044.1 HIV-1 isolate BRGO3080 from Brazil envelope glycoprotein (env) gene, partial cds	❓		Details
5bfc8ae5cfc0c	DQ358802.1 HIV-1 isolate 01BR125 from Brazil, complete genome	✖		Details

This page will appear when more than one sequence is entered on the home page. It consists of four frames where the top three, Preparation, Legend and Statistics shows information about the submitted sequences, legends for the result icons and result counter. The lower table contains the results, where lines and each sample data separate each sample is divided into five columns: request identifier, sequence name according to the FASTA file, the result of the analysis according to the legend, mutations found on sequence and a "Details" button to open a new window to view the individual results of that sample.

Section 3 – Single sequence results page

This page will be displayed when a single sequence is entered on the home page, or when you click on the "Details" button at any sequence of the multiple sequence results page. It consists in 4 tabs: Pre-alignment and Alignment tabs show information about the submitted sequence and the alignment's log respectively, Post-alignment and Results tabs contain the results itself, as explained below:

- Post-alignment. This tab displays the alignment results.

    o Sequence map. After clicking the "Submit" button, a new window will be open at the Los Alamos HIV sequence locator, where the page will display an image with the query sequence on red and its likely location in the HIV-1 genome. The base numbers displayed is according to the HXB2 genome.

o <u>Gaps counter.</u> This table shows the amounts of insertions and deletions of bases comparing the query sequence to the HXB2 sequence. The sum of these values must be equal to zero, or this will indicate an alignment error.

Gaps counter	
Insertions	0
Deletions	0

o <u>DNA sequence.</u> This table shows the region of the DNA responsible for the synthesis of the HIV transmembrane protein gp41, and where mutations that cause resistance to enfuvirtide are found. The first line contains the HXB2 reference base and the last line contains the query sequence. The middle line contains the base number according to the query sequence. If found, the mutations will be marked in the table as dark gray.

DNA sequence

HXB2	G	G	T	A	T	A	G	T	G	C
Base	1642	1643	1644	1645	1646	1647	1648	1649	1650	1651
Query	A	G	T	A	T	A	G	T	G	C
HXB2	A	G	C	A	G	C	A	G	A	A
Base	1652	1653	1654	1655	1656	1657	1658	1659	1660	1661
Query	A	A	C	A	G	C	A	G	A	A
HXB2	C	A	A	T	T	T	G	C	T	G
Base	1662	1663	1664	1665	1666	1667	1668	1669	1670	1671
Query	C	G	A	T	T	T	G	C	T	G

o <u>gp41 protein sequence.</u> This table is the translation of the DNA sequence from the previous table. The middle line contains the amino acid number according to the HXB2 reference. If found, the mutations will be marked in the table as dark gray.

gp41 protein sequence

HXB2	G	I	V	Q	Q	Q	N	N	L	L
Base	36	37	38	39	40	41	42	43	44	45
Query	S	I	V	Q	Q	Q	N	D	L	L

- <u>Results.</u> This tab will show the mutations present in the query sequence that confer resistance to enfuvirtide according to our database.

N43D

Thomas Melby *et al.* (2006)	Michael Mink *et al.* (2005)	Other sources
N43D is associated with 37.2 fold resistance to enfuvirtide.	N43D has associated with 18 fold resistance to enfuvirtide *in vitro*.	IAS-USA
The combination of N43D mutant and the WT variant was assosiated with 8.1 fold resistance to enfuvirtide in a single patient.	N43D has a range of 123-249 fold resistance to enfuvirtide *in vivo*.	Stanford
		HIV French Resistance

Each mutation found will be searched in the articles referring to clinical studies with enfuvirtide cataloged in our database and information regarding the number of fold resistance to enfuvirtide that a given mutation gives the HIV virus query will be displayed. A link to the source is available when the paper's title is clicked. Additionally, each mutation found will be displayed with the reference of other sources that cites it, among the websites cataloged in our database. A link to the source is available when the website's name is clicked.

Under the hood

Section 1 – Server

HIVfird is a software written in PHP. It was developed on an Orange Pi Zero model server with a 1.2 GHz quad-core processor, using the Armbian Debian Linux operating system. Now it runs on a virtual server located at the Information Technology Superintendence of the Federal University of Bahia, Brazil.

Section 2 – Alignment platforms

HIVfird has enormous flexibility in user-inserted HIV sequence size, varying from the first heptad repeat (HR1) region of the gp41 envelope gene to the entire genome of the virus. To make this analysis feasible, a prior sequence alignment with our reference sequence is necessary to properly locate the bases that confer resistance to the fusion inhibitors. For this, the software Kalign is available.

Kalign (current version: 2.04) is a free software produced by Timo Lassmann and team. This software was chosen because it is free, reliable, and extremely light compared to other alignment software. Because of this, it is HIVfird's alignment option. More information about Kalign can be found at http://msa.sbc.su.se/.

Troubleshooting

Error	Explanation and solution
No FASTA file	This error happens when the results page is accessed directly, without sending the sequences in the form of the home page. Please note that HIVfird does not store user submitted sequences, so no link is available for future access of submitted data.
Invalid FASTA file	A sequence in FASTA format begins with a single-line description, followed by lines of sequence data. The description line (defline) is distinguished from the sequence data by a greater-than (">") symbol at the beginning. So please make sure your FASTA sequence is properly formatted.
Too many sequences	The HIVfird has a limit of 300 sequences per run. Please divide your file into smaller files respecting this limitation.
Sequence too short	During alignment, the FASTA file must have at least 36 bp. Make sure your sequence has this minimum length.
Bad alignment	Sometimes during the alignment process between our reference sequence and the query sequence, the alignment software has difficulties to correctly align the sequences and usually inserts gaps between the bases. When this happens, it is not possible to define clearly where the bases that confer resistance to the fusion inhibitors are located. When this happens, you can: - Check on the DNA map if the inserted sequence belongs to the *env* gene. HIVfird only works with the HIV *env* gene. Check your electropherogram to see if there has no incorrectly bases read. It is important to note that HIVfird does not work with ambiguous bases.
Something went wrong	This error indicates a malfunction of HIVfird. Please contact the technical support with the sequence inserted and what circumstances this has occurred.