

# User Manual for PhEVER

Leonor Palmeira, Simon Penel

mlpalmeira@ulg.ac.be

simon.penel@univ-lyon1.fr

June 29, 2010

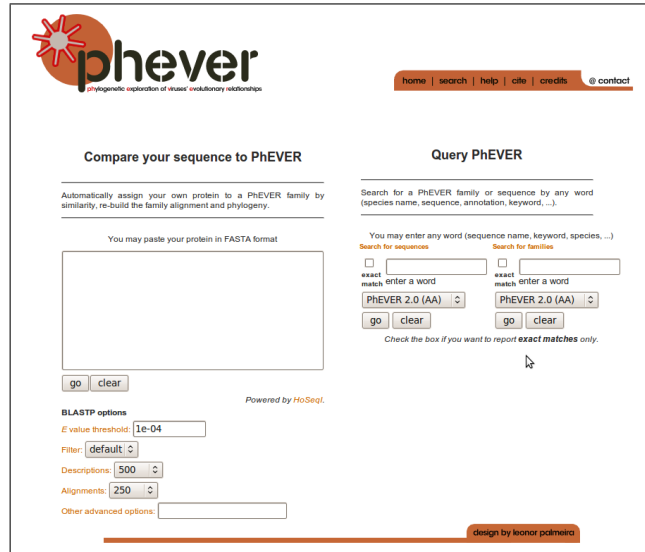


## Contents

<b>1</b>	<b>Overview</b>	<b>2</b>
<b>2</b>	<b>Using the PhEVER interface</b>	<b>2</b>
1	Using the Quick-Search tool . . . . .	2
1.1	Searching for a sequence ( <b>A</b> ) . . . . .	3
1.2	Searching for a family ( <b>B</b> ) . . . . .	4
1.3	Retrieve the species of a family . . . . .	5
1.4	Retrieve the sequences of a family . . . . .	5
1.5	Retrieve and manipulate a family alignment . . . . .	5
1.6	Retrieve and anipulate a family phylogeny . . . . .	6
2	Using the HoSeqI tool . . . . .	6
<b>3</b>	<b>Other tools to query PhEVER</b>	<b>7</b>

# 1 Overview

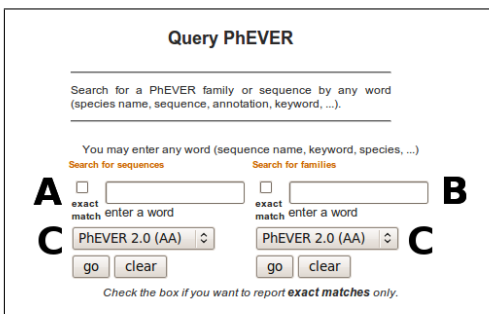
When accessing the PhEVER website (<http://pbil.univ-lyon1.fr/databases/phever>), the main page presents both an HoSeqI interface on the left and a Quick-Search tool on the right. These tools are the simplest ways to access the PhEVER database. More complex web queries can also be built using the 'WWW-Query' tool to search for sequences or families by combining several criteria and the 'Cross-Taxa Search' tool to query for families containing several taxa. All these web-interfaced tools and many others, including standalone software, are available from the PBIL website and explained more in detail at <http://pbil.univ-lyon1.fr/databases/phever/search.html>.



## 2 Using the PhEVER interface

### 1 Using the Quick-Search tool

The Quick-Search tool allows to query the PhEVER database in a very simple and straightforward way.



In the Quick-Search tool, you can either wish to look for (i) a sequence or (ii) a family corresponding to your query. If you wish to query a sequence, enter the query on the left side

of the tool (A), if you wish to query a family, enter the query on the right side of the tool (B). As PhEVER is actually composed of two distinct databases which are the nucleic acid database and the amino acid database, you must specify which database to query. This can be chosen in the drop-down menu (C). Note that the nucleic acid database was built using the genomic flat files and their annotations. The amino acid database was built by translating all CDS and mature peptides contained in the genomic flat files.

## 1.1 Searching for a sequence (A)

Let us search for all the sequences of Human herpesvirus 8 (*a.k.a* Kaposi Sarcoma-associated Herpesvirus). We may build a query in (A) using sequence names, species names or keywords, so for our example we can use: **human herpesvirus 8** or **kaposi** or **NC\_009333**. We choose PhEVER 2.0 (aa) from the drop down menu to look for families in the protein database.

**Sequence lists retrieved by WWW-Query:**

search AC | search name | search name | search species | search keyword |

sequences matching a species: 86 [edit list](#)

---

**Page 1 - Database is PHEVER2**

**Sequence list "all":**

**Edit Modify Retrieve Analyze View History**

**Sequences 1 to 20 of 86**

Select sequences:  Display:  per page

Save selection Next Go to page 1 Clear Reset Select all:

1.  [NC\\_009333\\_PE1](#) (NC\_009333.PE1).
2.  [NC\\_009333\\_PE10](#) (NC\_009333.PE10).
3.  [NC\\_009333\\_PE11](#) (NC\_009333.PE11).
4.  [NC\\_009333\\_PE12](#) (NC\_009333.PE12).
5.  [NC\\_009333\\_PE13](#) (NC\_009333.PE13).

This will yield 86 sequences corresponding to all proteins annotated in the *Human herpesvirus 8* genome, that can then be easily downloaded in Bank (EMBL), Fasta or Mase format through the Retrieve button.

**Edit Modify Retrieve Analyze View History**

**Download the sequence names only:**

**Retrieve the sequence data:**

Sequence:  DNA  Proteins

Format:  Bank  Fasta  Mase

Mode:  Direct sending  FTP archive

List name:

## 1.2 Searching for a family (B)

In (B), we may also build a query using sequence names, species names or keywords. Let us search for all families of uracil DNA-glycosylase in PhEVER. In (B), we type `uracil` (note that the query is actually not case sensitive) to query for all families containing at least one sequence having this word in its description. We choose PhEVER 2.0 (aa) from the drop down menu to look for families in the protein database.

We then obtain the following list of families indicating that 747 sequences matched our query and that these sequences belong to 4 associated families. Each one of these families can be analysed separately by clicking on the family accession number (e.g. HBG276673) as we will now illustrate.

**Family lists retrieved by WWW-Query:**

search AC | search name | search name | search species | search keyword | keyword : **Matching sequences: 747**  
 all sequences : **Matching sequences: 747**  
 Associated families: **4**

**Page 1 - Database is PHEVER2**

**Family list "all"**

[History](#) [View](#) [Download](#)

**Families 1 to 4 of 4**

Select families  Display:  per page

[Save selection](#) [Go to](#) [page 1](#) [Clear](#) [Reset](#) [Select all:](#)

Family	Sequences	Species	Definition
1. <a href="#">HBG276673</a>	760	511	"Uracil DNA glycosylase like protein"
2. <a href="#">HBG273995</a>	25	25	"uracil DNA glycosylase" "82R protein" "CPXV121 protein" "E4R" "EVM093" "gp079R"
3. <a href="#">HBG292879</a>	1	1	"putative uracil permease"
4. <a href="#">HBG295697</a>	1	1	"uracil DNA glycosylase"

After clicking on the family accession number, you access the webpage of the given family. From this webpage, a large number of features are available from which we will present the four most widely used.

**PHEVER2 Database**

GENE FAMILY HBG276673

Number of sequences	759
Number of taxons	510
Common ancestor	Root( <a href="#">NCBI</a> )( <a href="#">ACNUC</a> )
Definition	"Uracil DNA glycosylase like protein"

[Nucleotide](#) [Sequences](#) [Retrieve](#) [Species](#) [Keywords](#) [Alignment](#) [Tree](#)

---

**Sequences selection by species**

Please select species among the family species to get the associated sequences:  
 (The number of sequences from each species is given between brackets).

Acaryochloris marina (1)

Acholeplasma laidlawii (1)

Acidovorax avenae (1)

Acidovorax ebreus tpsy (1)

Acidovorax sp. js42 (1)

Acinetobacter baumannii (6)

Acinetobacter sp. (1)

Actinobacillus pleuropneumoniae (3)

### 1.3 Retrieve the species of a family



If you want to access the species composition of a given family, just click on the **Species** button in the above menu, you will then obtain a list of each species present, followed by the number of sequences each one contains in the given family. You can also then choose to visualize all the sequences of any of these species for a subsequent analysis and possibly to download them if needed.

Species of HBG276673 in PHEVER2	
<a href="#">Acaryochloris marina</a>	1 <a href="#">(Get all sequences)</a>
<a href="#">Acholeplasma laidlawii</a>	1 <a href="#">(Get all sequences)</a>
<a href="#">Acidovorax avenae</a>	1 <a href="#">(Get all sequences)</a>
<a href="#">Acidovorax ebreus tpsy</a>	1 <a href="#">(Get all sequences)</a>
<a href="#">Acidovorax sp. js42</a>	1 <a href="#">(Get all sequences)</a>
<a href="#">Acinetobacter baumannii</a>	6 <a href="#">(Get all sequences)</a>
<a href="#">Acinetobacter sp.</a>	1 <a href="#">(Get all sequences)</a>
<a href="#">Actinobacillus pleuropneumoniae</a>	3 <a href="#">(Get all sequences)</a>
<a href="#">Actinobacillus succinogenes</a>	1 <a href="#">(Get all sequences)</a>
<a href="#">Actinosynnema mirum</a>	1 <a href="#">(Get all sequences)</a>
<a href="#">Aeromonas hydrophila</a>	1 <a href="#">(Get all sequences)</a>
<a href="#">Aeromonas salmonicida</a>	1 <a href="#">(Get all sequences)</a>

### 1.4 Retrieve the sequences of a family



If you want to retrieve the sequences of a given family or from any set of sequences, just click on the **Retrieve** button in the menu. You will then access a page where you can choose to retrieve either the DNA or the protein sequences in the format of your choice.

#### Retrieve Family Sequences

Sequences data from HBG276673 in PHEVER2

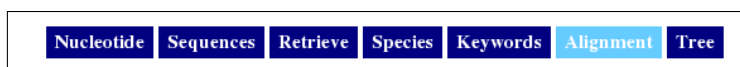
List name = HBG276673\_1st

Sequence:  DNA  Proteins

Format:  Bank  Fasta  Mase

Mode:  Direct sending  FTP archive

### 1.5 Retrieve and manipulate a family alignment



For each PhEVER family, pre-computed alignments are available through the web-interface. These are available through the **Alignment** button in the top menu of each family. As seen on the screen capture below, they can be subsequently downloaded as is, or edited using the JalView editor.

**Alignment of the gene family HBG276673 of PHEVER2**

[View alignment with JalView](#) [Download alignment](#)

Threshold for coloring = 80% of identity ( [100%](#) [90%](#) [80%](#) [70%](#) [60%](#) [50%](#) [Clustalw](#) [consensus](#) [Difference](#) )

<input type="checkbox"/>	<a href="#">NC_001993_PE208</a>	.....
<input type="checkbox"/>	<a href="#">NC_002520_PE245</a>	.....
<input type="checkbox"/>	<a href="#">HYGHE1_1_PE97</a>	.....
<input type="checkbox"/>	<a href="#">HYPHE1_1_PE235</a>	.....
<input type="checkbox"/>	<a href="#">HYGAL1_1_PE640</a>	.....
<input type="checkbox"/>	<a href="#">NC_006623_PE61</a>	.....
<input type="checkbox"/>	<a href="#">NC_005264_PE62</a>	.....MANDERVATGEDRWGRGLGVAELRIDAASVAPNDPTEDREVVLPIPGP
<input type="checkbox"/>	<a href="#">HYCAS_1_PES78</a>	.....

## 1.6 Retrieve and anipulate a family phylogeny

Nucleotide
Sequences
Retrieve
Species
Keywords
Alignment
Tree

For each PhEVER family containing at least 3 sequences, pre-computed phylogenies are available through the web-interface. These are available through the **Tree** button in the top menu of each family. The labels on the leaves can be chosen to be the **UniprotKB** accession numbers (when available), the **Species** names, the PhEVER accession numbers or the **Ensembl** accession numbers (when available). As seen on the screen capture below, there is a small menu at the top, which can lead back to the **Family** or the **Alignment** webpage. From this menu, you can also **Download** the phylogenetic tree in the parenthesized Newick format or edit the tree using the ATV Java applet.

**Phylogenetic tree of the HBG276673 gene family of PHEVER2 database**

Family Alignment Download

UniprotKB Accession number:  Species:

Display:  UniprotKB  Species

## 2 Using the HoSeqI tool

The HoSeqI interface allows you to provide your own sequence of interest to search for families of homologies. More information on the HoSeqI tool is available at:

**Compare your sequence to PhEVER**

---

Automatically assign your own protein to a PhEVER family by similarity, re-build the family alignment and phylogeny.

---

You may paste your protein in FASTA format

*Powered by HoSeqI.*

**BLASTP options**

E value threshold:

Filter:

Descriptions:

Alignments:

Other advanced options:

<http://bioinformatics.oxfordjournals.org/cgi/content/full/22/14/1786>.

Once you have provided your own sequence of interest to search for families of similar sequences, you will access a web-form where you can retrieve information on the matching family, and where you can also rebuild the alignment and the phylogeny of the family including your own sequence.

**HoSeqI**

BBE contribution to PBIL in Lyon, France

---

Query sequence

BLAST Output

Matching family: HBG292925

Multiple alignment program:  [Advanced parameters](#)

Phylogenetic tree rebuilding program:  [Advanced parameters](#)

(Pairwise distances are calculated using Kimura's correction.)

Bootstrap:  Yes  No

The alignment can then be viewed using the JalView editor, and the phylogeny can be viewed through the ATV Java Applet.

### 3 Other tools to query PhEVER

PhEVER is a database structured under ACNUC and all ACNUC tools can therefore be used to query it. These are explained in more detail on the PBIL website and more specifically at: <http://pbil.univ-lyon1.fr/databases/phever/search.html>.