The Immune Epitope Database Analysis Resource

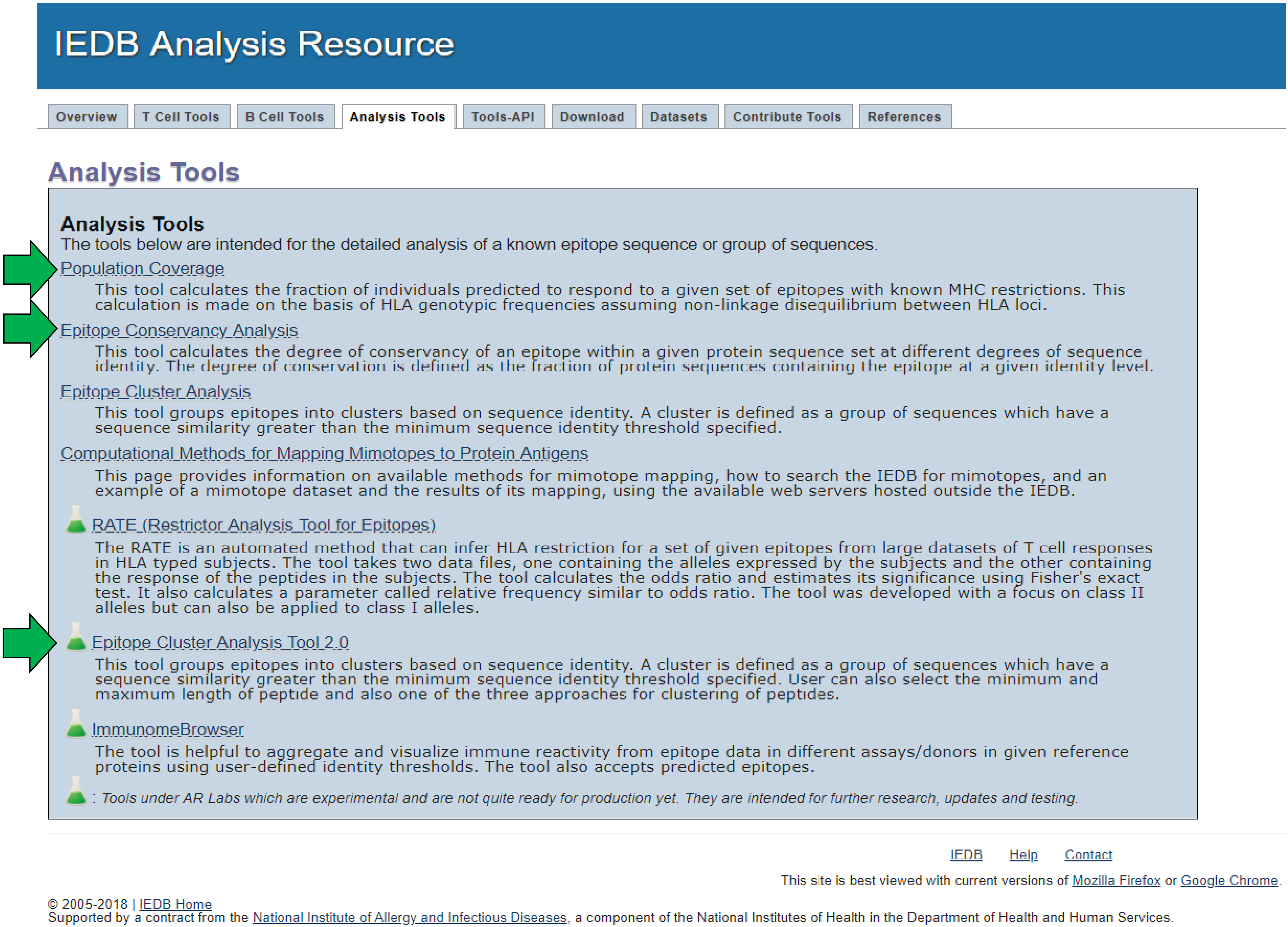
Epitope Analysis Tools:

Adding value to epitope datasets

Bjoern Peters

IEDB User Workshop

October 23, 2018



IEDB Analysis Resource

Analysis Tools

Analysis Tools

The tools below are intended for the detailed analysis of a known epitope sequence or group of sequences.

Population Coverage

This tool calculates the fraction of individuals predicted to respond to a given set of epitopes with known MHC restrictions. This calculation is made on the basis of HLA genotypic frequencies assuming non-linkage disequilibrium between HLA loci.

Epitope Conservancy Analysis

This tool calculates the degree of conservancy of an epitope within a given protein sequence set at different degrees of sequence identity. The degree of conservation is defined as the fraction of protein sequences containing the epitope at a given identity level.

Epitope Cluster Analysis

This tool groups epitopes into clusters based on sequence identity. A cluster is defined as a group of sequences which have a sequence similarity greater than the minimum sequence identity threshold specified.

Computational Methods for Mapping Mimotopes to Protein Antigens

This page provides information on available methods for mimotope mapping, how to search the IEDB for mimotopes, and an example of a mimotope dataset and the results of its mapping, using the available web servers hosted outside the IEDB.

🧪 RATE (Restrictor Analysis Tool for Epitopes)

The RATE is an automated method that can infer HLA restriction for a set of given epitopes from large datasets of T cell responses in HLA typed subjects. The tool takes two data files, one containing the alleles expressed by the subjects and the other containing the response of the peptides in the subjects. The tool calculates the odds ratio and estimates its significance using Fisher's exact test. It also calculates a parameter called relative frequency similar to odds ratio. The tool was developed with a focus on class II alleles but can also be applied to class I alleles.

🧪 Epitope Cluster Analysis Tool 2.0

This tool groups epitopes into clusters based on sequence identity. A cluster is defined as a group of sequences which have a sequence similarity greater than the minimum sequence identity threshold specified. User can also select the minimum and maximum length of peptide and also one of the three approaches for clustering of peptides.

🧪 ImmunomeBrowser

The tool is helpful to aggregate and visualize immune reactivity from epitope data in different assays/donors in given reference proteins using user-defined identity thresholds. The tool also accepts predicted epitopes.

🧪 : *Tools under AR Labs which are experimental and are not quite ready for production yet. They are intended for further research, updates and testing.*

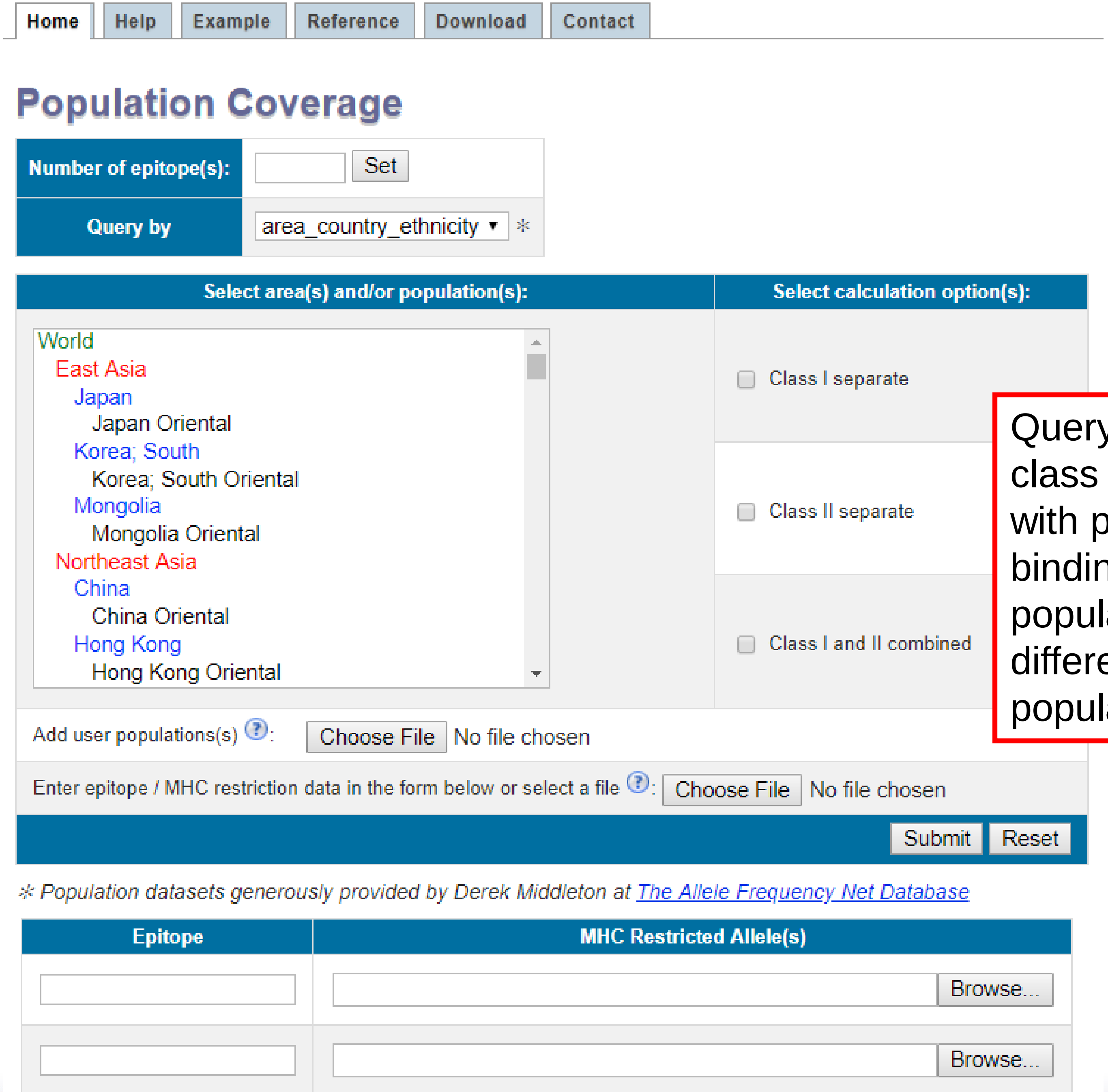
IEDB Help Contact

This site is best viewed with current versions of Mozilla Firefox or Google Chrome.

© 2005-2018 | IEDB Home
Supported by a contract from the National Institute of Allergy and Infectious Diseases, a component of the National Institutes of Health in the Department of Health and Human Services.

Population Coverage

- Calculates the fraction of individuals predicted to respond to a given set of epitopes with known MHC restrictions
- Based on HLA genotypic frequencies assuming non-linkage disequilibrium between HLA loci
- HLA genotypic frequencies have been updated with more comprehensive dataset (*http://allelefrequencies.net* rather than *dbMHC*)



Population Coverage

Number of epitope(s):	Set
Query by	area_country_ethnicity ▾ ✻

Select area(s) and/or population(s):	**Select calculation option(s):**
World East Asia Japan Japan Oriental Korea; South Korea; South Oriental Mongolia Mongolia Oriental Northeast Asia China China Oriental Hong Kong Hong Kong Oriental	☐ Class I separate ☐ Class II separate ☐ Class I and II combined

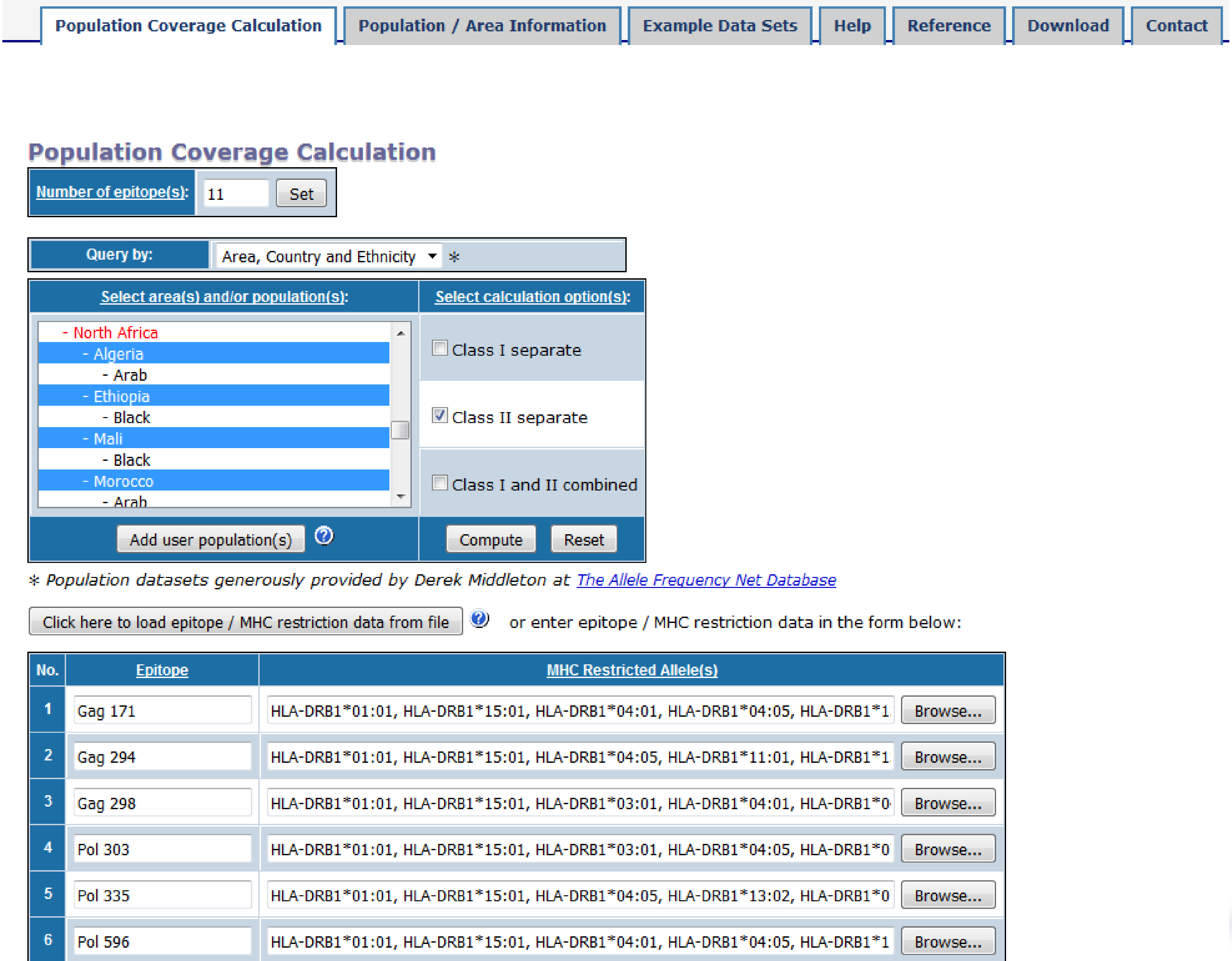
Add user populations(s) ?: Choose File No file chosen

Enter epitope / MHC restriction data in the form below or select a file ?: Choose File No file chosen

✻ *Population datasets generously provided by Derek Middleton at [The Allele Frequency Net Database]()*

Epitope	MHC Restricted Allele(s)
	Browse...
	Browse...

Query: For a set of 11 MHC class II restricted epitopes with promiscuous HLA binding, what is the population coverage in different North African populations?



Population Coverage Calculation

Number of epitope(s): 11 Set

Query by: Area, Country and Ethnicity ✻

Select area(s) and/or population(s):

- North Africa
  - Algeria
    - Arab
  - Ethiopia
    - Black
  - Mali
    - Black
  - Morocco
    - Arab

Select calculation option(s):

☐ Class I separate

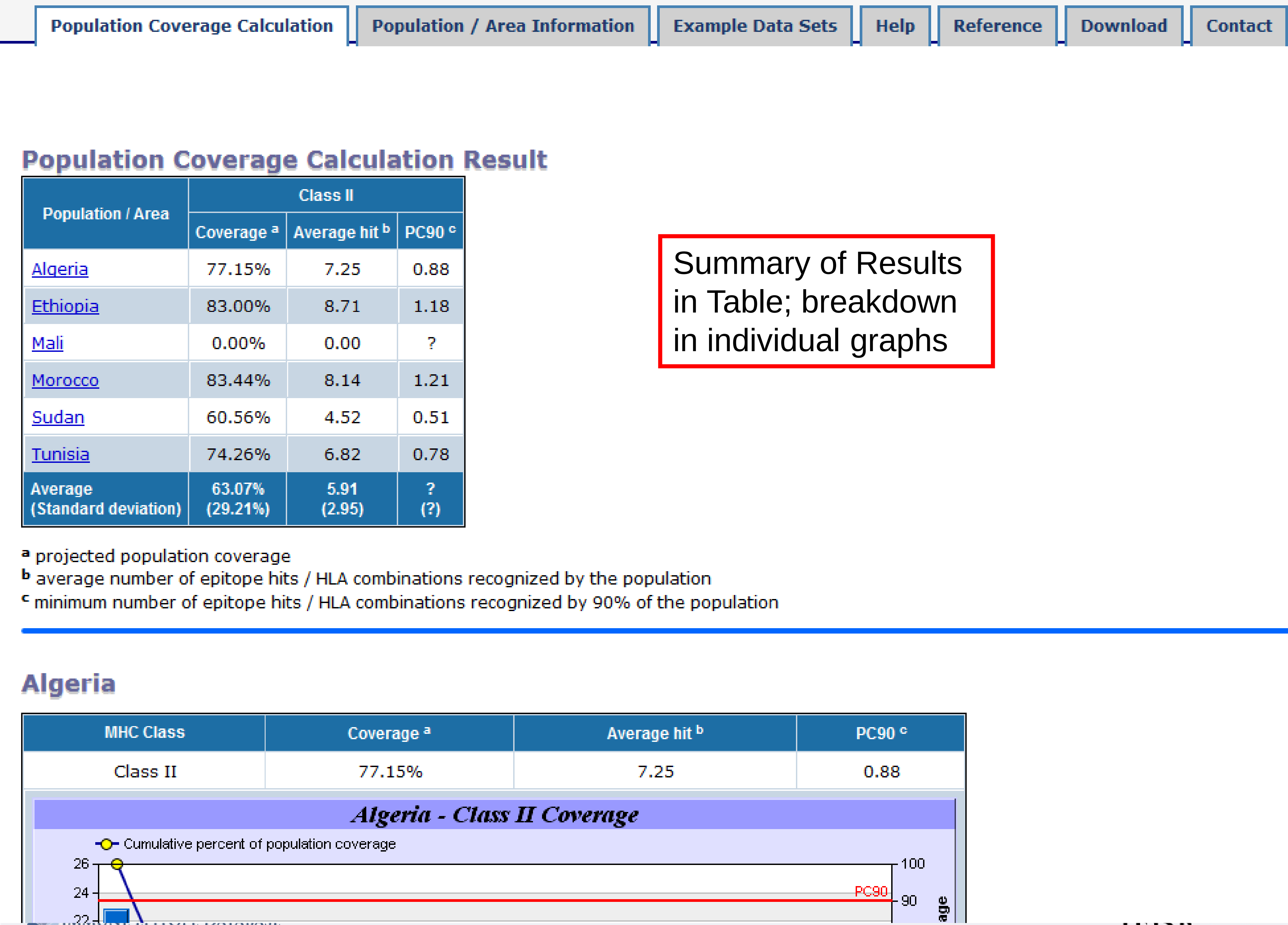
☑ Class II separate

☐ Class I and II combined

✻ *Population datasets generously provided by Derek Middleton at The Allele Frequency Net Database*

or enter epitope / MHC restriction data in the form below:

No.	Epitope	MHC Restricted Allele(s)
1	Gag 171	HLA-DRB1*01:01, HLA-DRB1*15:01, HLA-DRB1*04:01, HLA-DRB1*04:05, HLA-DRB1*1
2	Gag 294	HLA-DRB1*01:01, HLA-DRB1*15:01, HLA-DRB1*04:05, HLA-DRB1*11:01, HLA-DRB1*1
3	Gag 298	HLA-DRB1*01:01, HLA-DRB1*15:01, HLA-DRB1*03:01, HLA-DRB1*04:01, HLA-DRB1*0
4	Pol 303	HLA-DRB1*01:01, HLA-DRB1*15:01, HLA-DRB1*03:01, HLA-DRB1*04:05, HLA-DRB1*0
5	Pol 335	HLA-DRB1*01:01, HLA-DRB1*15:01, HLA-DRB1*04:05, HLA-DRB1*13:02, HLA-DRB1*0
6	Pol 596	HLA-DRB1*01:01, HLA-DRB1*15:01, HLA-DRB1*04:01, HLA-DRB1*04:05, HLA-DRB1*1



Population Coverage Calculation Result

Population / Area	Class II		
	Coverage [a]	Average hit [b]	PC90 [c]
Algeria	77.15%	7.25	0.88
Ethiopia	83.00%	8.71	1.18
Mali	0.00%	0.00	?
Morocco	83.44%	8.14	1.21
Sudan	60.56%	4.52	0.51
Tunisia	74.26%	6.82	0.78
Average (Standard deviation)	63.07% (29.21%)	5.91 (2.95)	? (?)

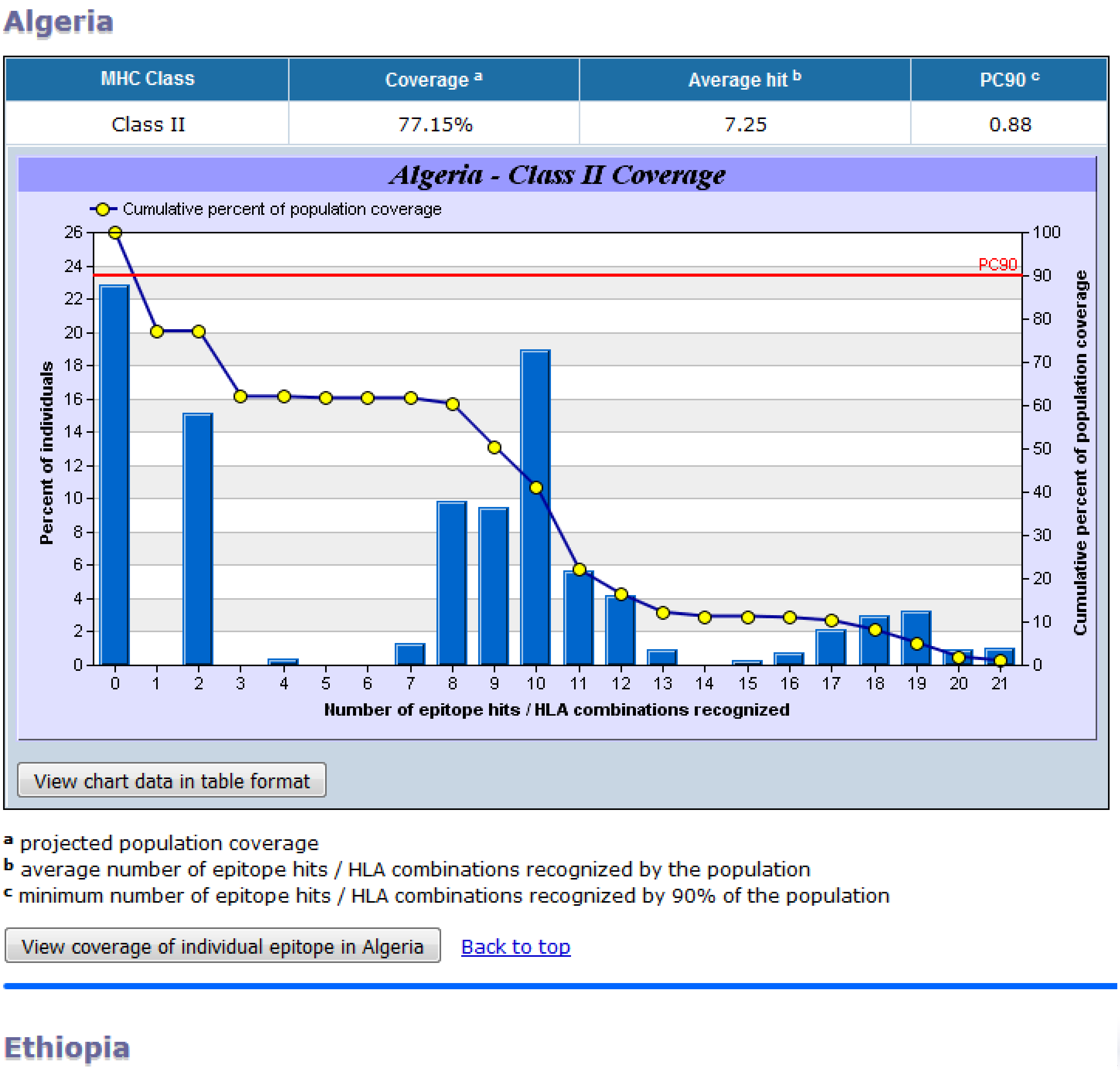
[a] projected population coverage
[b] average number of epitope hits / HLA combinations recognized by the population
[c] minimum number of epitope hits / HLA combinations recognized by 90% of the population

Summary of Results in Table; breakdown in individual graphs

Algeria

MHC Class	Coverage [a]	Average hit [b]	PC90 [c]
Class II	77.15%	7.25	0.88

Algeria - Class II Coverage



Algeria

MHC Class	Coverage [a]	Average hit [b]	PC90 [c]
Class II	77.15%	7.25	0.88

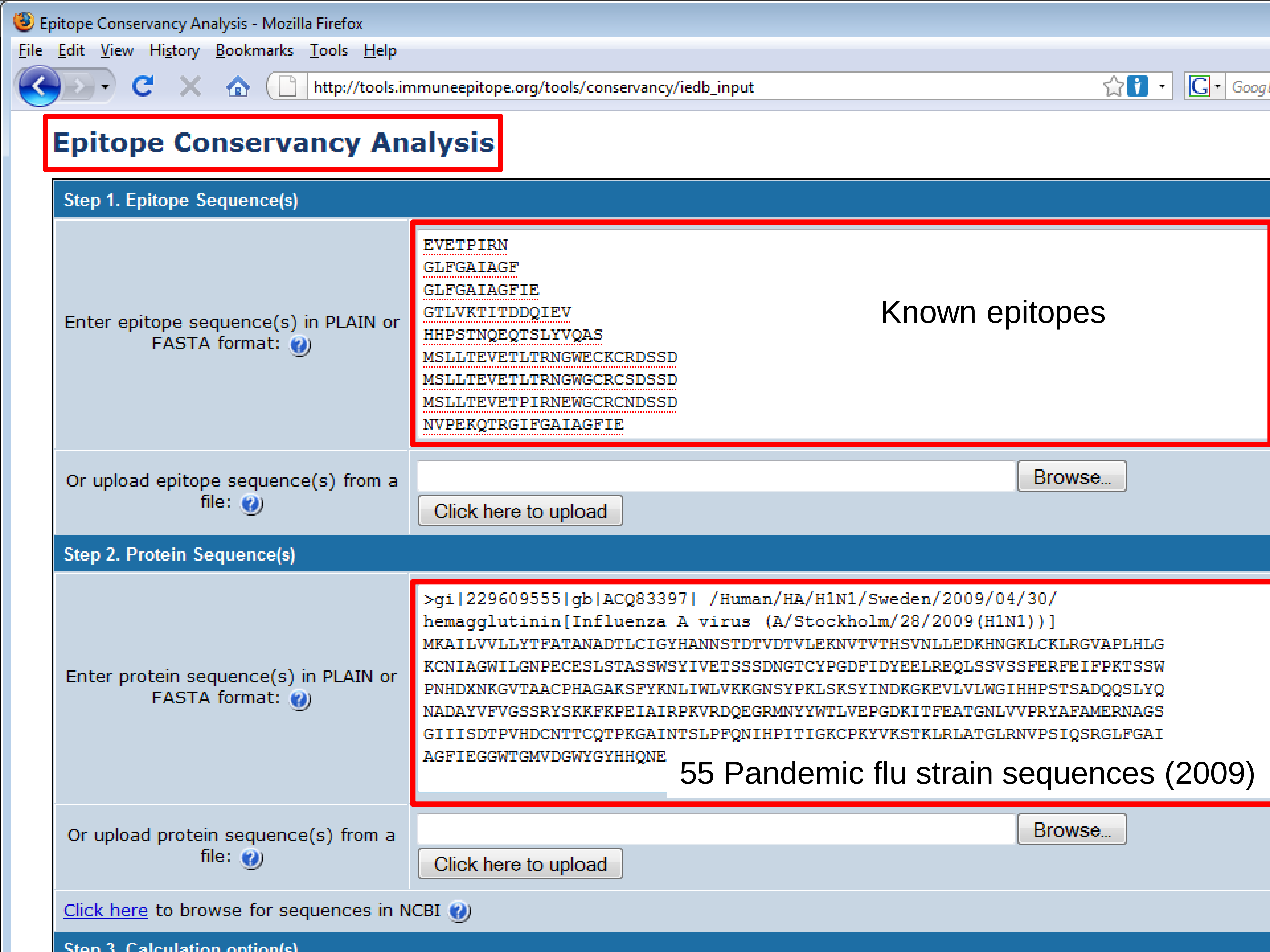
Algeria - Class II Coverage

[a] projected population coverage
[b] average number of epitope hits / HLA combinations recognized by the population
[c] minimum number of epitope hits / HLA combinations recognized by 90% of the population

Ethiopia

Epitope Conservancy Analysis

- Calculates the degree of conservancy of an epitope within a given protein sequence set at different degrees of sequence identity
- Degree of conservation is defined as the fraction of protein sequences containing the epitope at a given identity level.
- Example: How conserved are influenza A epitopes discovered prior to 2009 in the pandemic Swine flu?



Epitope Conservancy Analysis

Step 1. Epitope Sequence(s)

Enter epitope sequence(s) in PLAIN or FASTA format:

```
EVETPIRN
GLFGAIAGF
GLFGAIAGFIE
GTLVKTITDDQIEV
HHPSTNQEQTSLYVQAS
MSLLTEVETLTRNGWECKCRDSSD
MSLLTEVETLTRNGWGCRCSDSSD
MSLLTEVETPIRNEWGCRCNDSSD
NVPEKQTRGIFGAIAGFIE
```

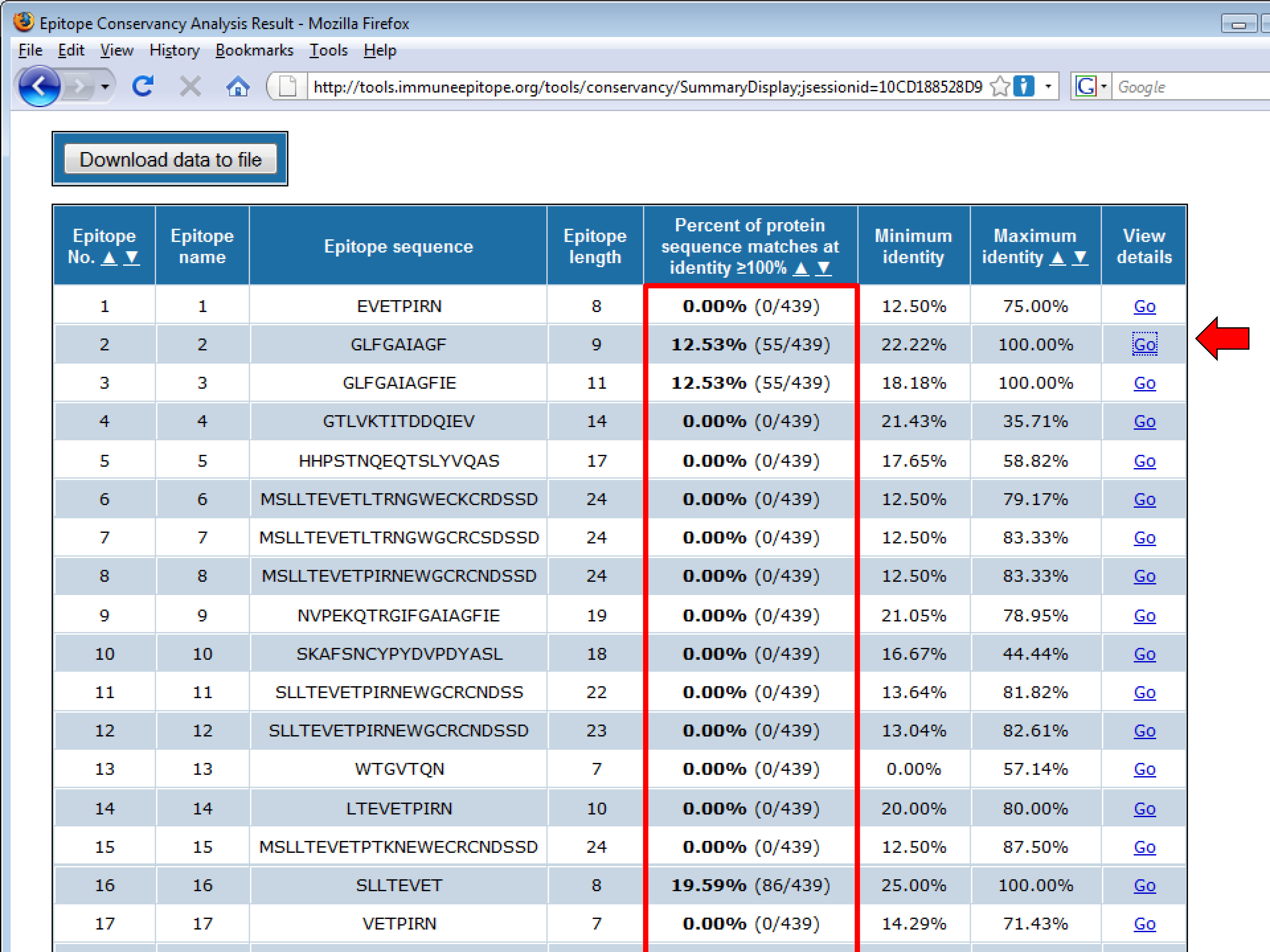
Known epitopes

Step 2. Protein Sequence(s)

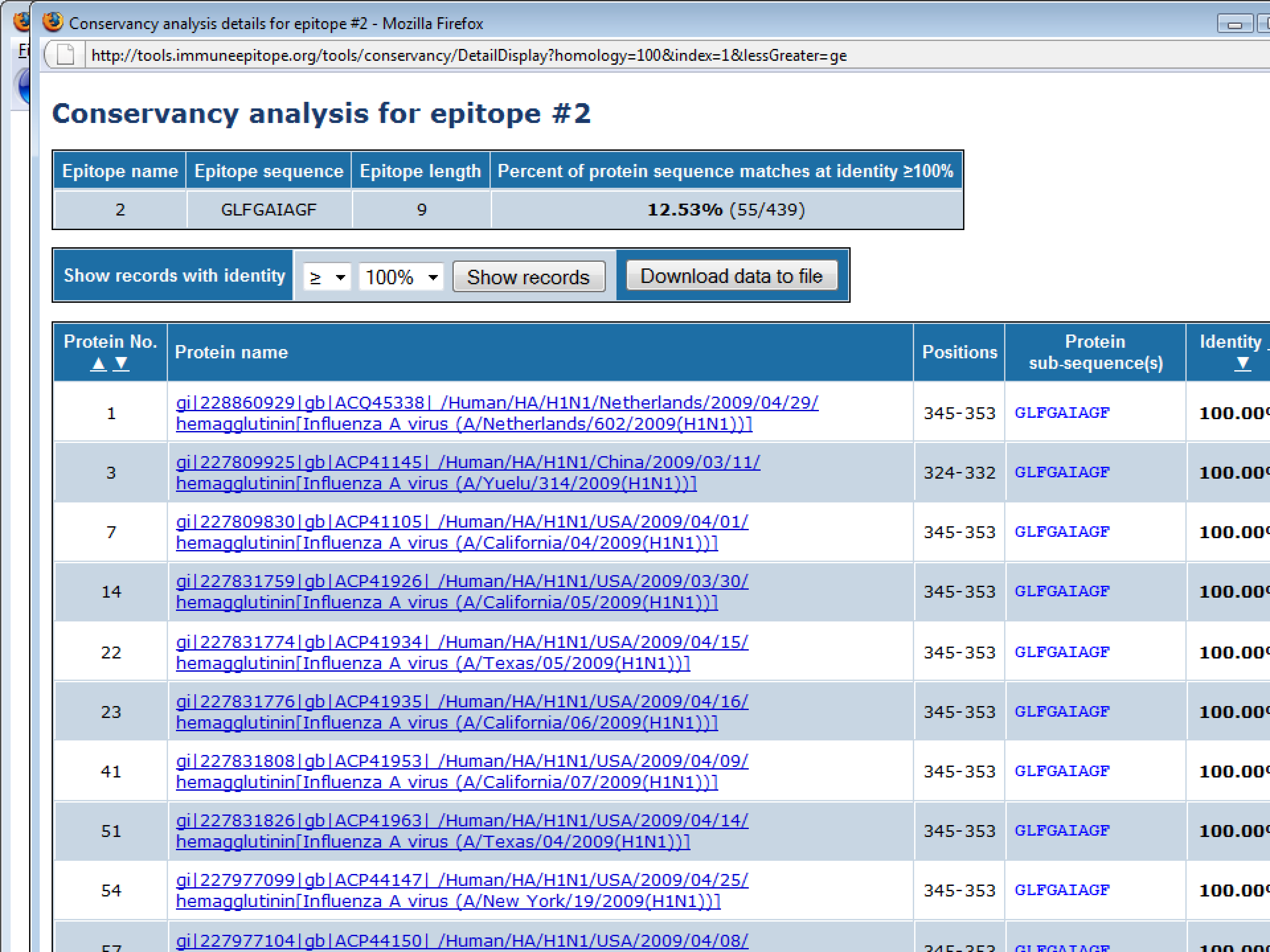
Enter protein sequence(s) in PLAIN or FASTA format:

```
>gi|229609555|gb|ACQ83397| /Human/HA/H1N1/Sweden/2009/04/30/
hemagglutinin[Influenza A virus (A/Stockholm/28/2009(H1N1))]
MKAILVVLLYTFATANADTLCIGYHANNSTDTVDTVLEKNVTVTHSVNLLEDKHNGKLCKLRGVAPLHLG
KCNIAGWILGNPECESLSTASSWSYIVETSSSDNGTCYPGDFIDYEELREQLSSVSSFERFEIFPKTSSW
PNHDXNKGVTAACPHAGAKSFYKNLIWLVKKGNSYPKLSKSYINDKGKEVLVLWGIHHPSTSADQQSLYQ
NADAYVFVGSSRYSKKFKPEIAIRPKVRDQEGRMNYYWTLVEPGDKITFEATGNLVVPRYAFAMERNAGS
GIIISDTPVHDCNTTCQTPKGAINTSLPFQNIHPITIGKCPKYVKSTKLRLATGLRNVPSIQSRGLFGAI
AGFIEGGWTGMVDGWYGYHHQNE
```

55 Pandemic flu strain sequences (2009)



Epitope No. ▲ ▼	Epitope name	Epitope sequence	Epitope length	Percent of protein sequence matches at identity ≥100% ▲ ▼	Minimum identity	Maximum identity ▲ ▼	View details
1	1	EVETPIRN	8	**0.00%** (0/439)	12.50%	75.00%	Go
2	2	GLFGAIAGF	9	**12.53%** (55/439)	22.22%	100.00%	Go
3	3	GLFGAIAGFIE	11	**12.53%** (55/439)	18.18%	100.00%	Go
4	4	GTLVKTITDDQIEV	14	**0.00%** (0/439)	21.43%	35.71%	Go
5	5	HHPSTNQEQTSLYVQAS	17	**0.00%** (0/439)	17.65%	58.82%	Go
6	6	MSLLTEVETLTRNGWECKCRDSSD	24	**0.00%** (0/439)	12.50%	79.17%	Go
7	7	MSLLTEVETLTRNGWGCRCSDSSD	24	**0.00%** (0/439)	12.50%	83.33%	Go
8	8	MSLLTEVETPIRNEWGCRCNDSSD	24	**0.00%** (0/439)	12.50%	83.33%	Go
9	9	NVPEKQTRGIFGAIAGFIE	19	**0.00%** (0/439)	21.05%	78.95%	Go
10	10	SKAFSNCYPYDVPDYASL	18	**0.00%** (0/439)	16.67%	44.44%	Go
11	11	SLLTEVETPIRNEWGCRCNDSS	22	**0.00%** (0/439)	13.64%	81.82%	Go
12	12	SLLTEVETPIRNEWGCRCNDSSD	23	**0.00%** (0/439)	13.04%	82.61%	Go
13	13	WTGVTQN	7	**0.00%** (0/439)	0.00%	57.14%	Go
14	14	LTEVETPIRN	10	**0.00%** (0/439)	20.00%	80.00%	Go
15	15	MSLLTEVETPTKNEWECRCNDSSD	24	**0.00%** (0/439)	12.50%	87.50%	Go
16	16	SLLTEVET	8	**19.59%** (86/439)	25.00%	100.00%	Go
17	17	VETPIRN	7	**0.00%** (0/439)	14.29%	71.43%	Go



Conservancy analysis for epitope #2

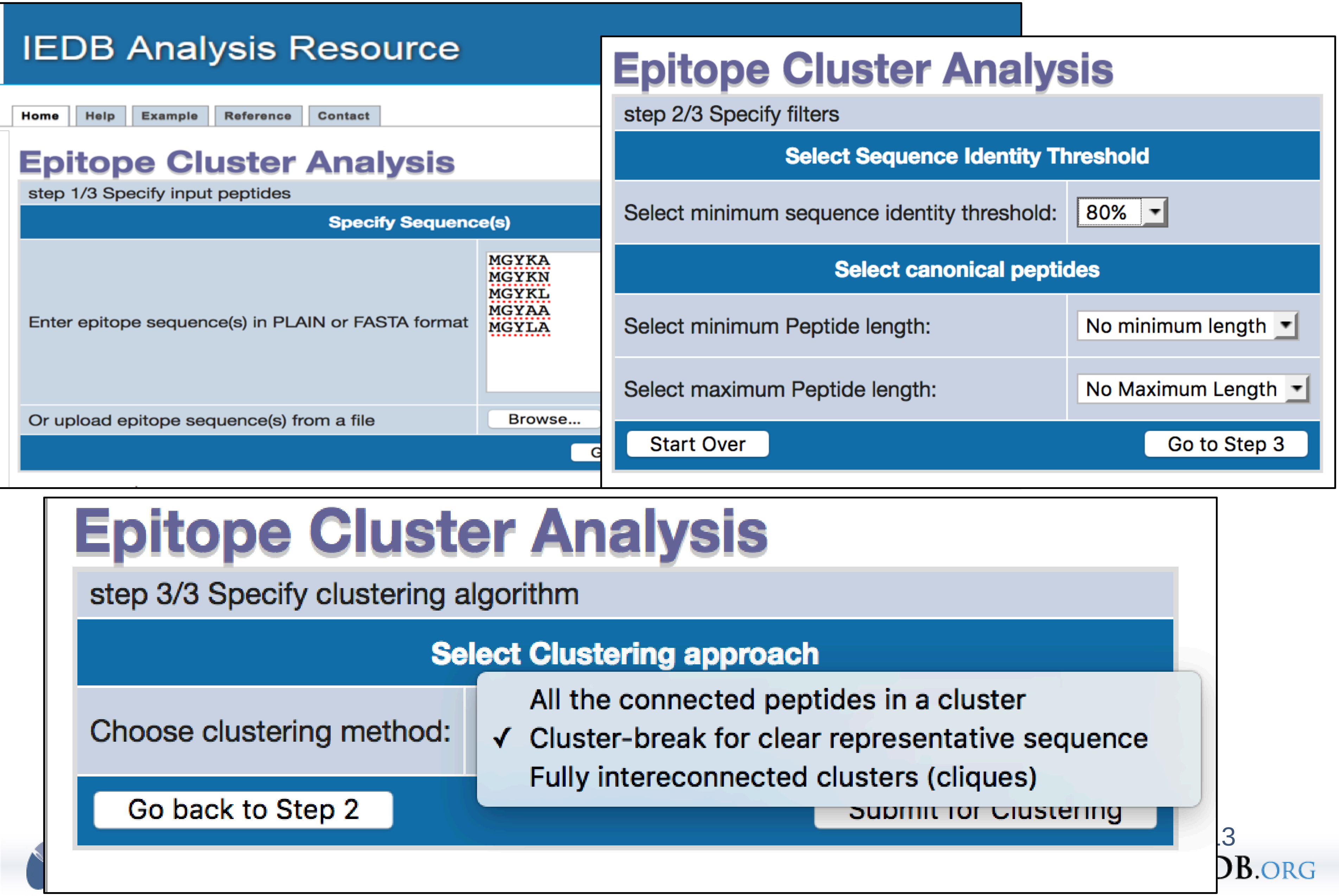
Epitope name	Epitope sequence	Epitope length	Percent of protein sequence matches at identity ≥100%
2	GLFGAIAGF	9	**12.53%** (55/439)

Protein No. ▲▼	Protein name	Positions	Protein sub-sequence(s)	Identity ▼
1	gi|228860929|gb|ACQ45338| /Human/HA/H1N1/Netherlands/2009/04/29/ hemagglutinin[Influenza A virus (A/Netherlands/602/2009(H1N1))]	345-353	GLFGAIAGF	**100.00**
3	gi|227809925|gb|ACP41145| /Human/HA/H1N1/China/2009/03/11/ hemagglutinin[Influenza A virus (A/Yuelu/314/2009(H1N1))]	324-332	GLFGAIAGF	**100.00**
7	gi|227809830|gb|ACP41105| /Human/HA/H1N1/USA/2009/04/01/ hemagglutinin[Influenza A virus (A/California/04/2009(H1N1))]	345-353	GLFGAIAGF	**100.00**
14	gi|227831759|gb|ACP41926| /Human/HA/H1N1/USA/2009/03/30/ hemagglutinin[Influenza A virus (A/California/05/2009(H1N1))]	345-353	GLFGAIAGF	**100.00**
22	gi|227831774|gb|ACP41934| /Human/HA/H1N1/USA/2009/04/15/ hemagglutinin[Influenza A virus (A/Texas/05/2009(H1N1))]	345-353	GLFGAIAGF	**100.00**
23	gi|227831776|gb|ACP41935| /Human/HA/H1N1/USA/2009/04/16/ hemagglutinin[Influenza A virus (A/California/06/2009(H1N1))]	345-353	GLFGAIAGF	**100.00**
41	gi|227831808|gb|ACP41953| /Human/HA/H1N1/USA/2009/04/09/ hemagglutinin[Influenza A virus (A/California/07/2009(H1N1))]	345-353	GLFGAIAGF	**100.00**
51	gi|227831826|gb|ACP41963| /Human/HA/H1N1/USA/2009/04/14/ hemagglutinin[Influenza A virus (A/Texas/04/2009(H1N1))]	345-353	GLFGAIAGF	**100.00**
54	gi|227977099|gb|ACP44147| /Human/HA/H1N1/USA/2009/04/25/ hemagglutinin[Influenza A virus (A/New York/19/2009(H1N1))]	345-353	GLFGAIAGF	**100.00**
57	gi|227977104|gb|ACP44150| /Human/HA/H1N1/USA/2009/04/08/	345-353	GLFGAIAGF	**100.00**

Epitope Cluster Analysis

- Groups epitopes into clusters based on having sequence identity greater than a specified threshold.
- Three different clustering approaches are implemented
- Enables diverse applications such as generating epitope pools, and understanding cross-reactivity

Clustering tool: Data input



IEDB Analysis Resource

Epitope Cluster Analysis

step 1/3 Specify input peptides

Specify Sequence(s)

Enter epitope sequence(s) in PLAIN or FASTA format

```
MGYKA
MGYKN
MGYKL
MGYAA
MGYLA
```

Or upload epitope sequence(s) from a file

Epitope Cluster Analysis

step 2/3 Specify filters

Select Sequence Identity Threshold

Select minimum sequence identity threshold: 80%

Select canonical peptides

Select minimum Peptide length: No minimum length

Select maximum Peptide length: No Maximum Length

Epitope Cluster Analysis

step 3/3 Specify clustering algorithm

Select Clustering approach

Choose clustering method:

- All the connected peptides in a cluster
- ✓ Cluster-break for clear representative sequence
- Fully intereconnected clusters (cliques)

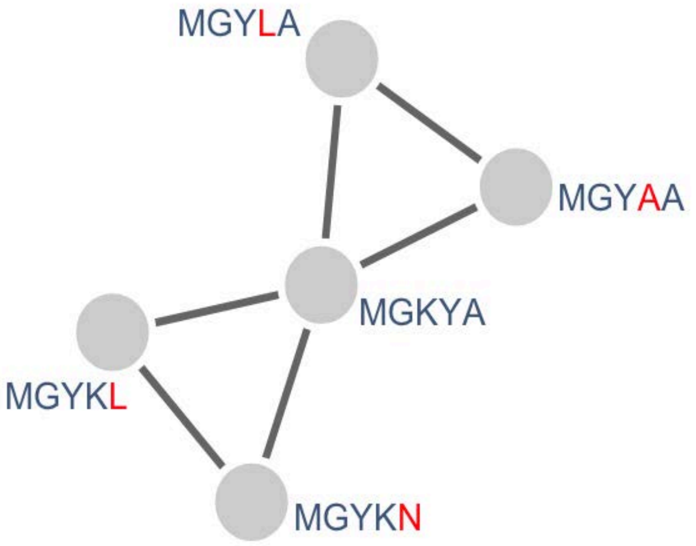
Assessing pairwise sequence identity identifies groups

Sequence identity matrix

Sequence i	Sequence j	# Residue match	Alignment length	Sequence identity (%)
MGYKA	MGYKN	**4**	**5**	**80**
MGYKA	MGYKL	**4**	**5**	**80**
MGYKA	MGYAA	**4**	**5**	**80**
MGYKA	MGYLA	**4**	**5**	**80**
MGYKN	MGYKL	**4**	**5**	**80**
MGYKN	MGYAA	3	5	60
MGYKN	MGYLA	3	5	60
MGYKL	MGYAA	3	5	60
MGYKL	MGYLA	3	5	60
MGYAA	MGYLA	**4**	**5**	**80**

Graph based representation

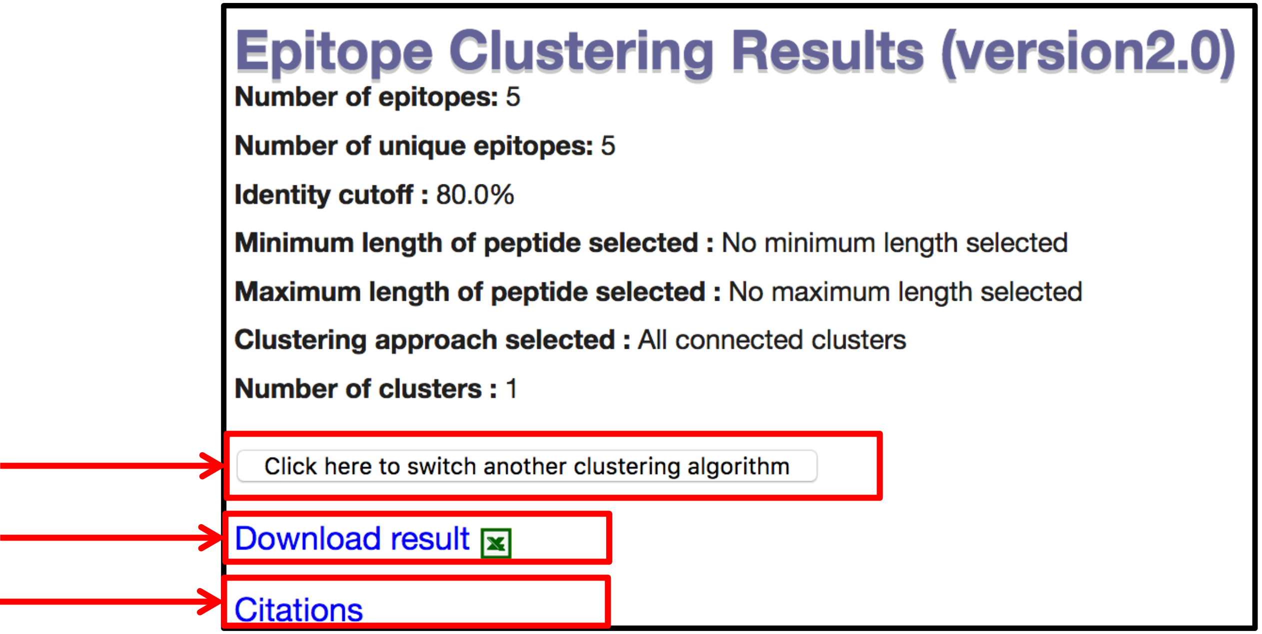
Sequence identity cutoff = 80%



Two possible grouping

Cluster: All peptides connected through a path are grouped together

Clique: Only fully connected sets of peptides are grouped together. There will be two cliques in this example.



Epitope Clustering Results (version2.0)

Number of epitopes: 5

Number of unique epitopes: 5

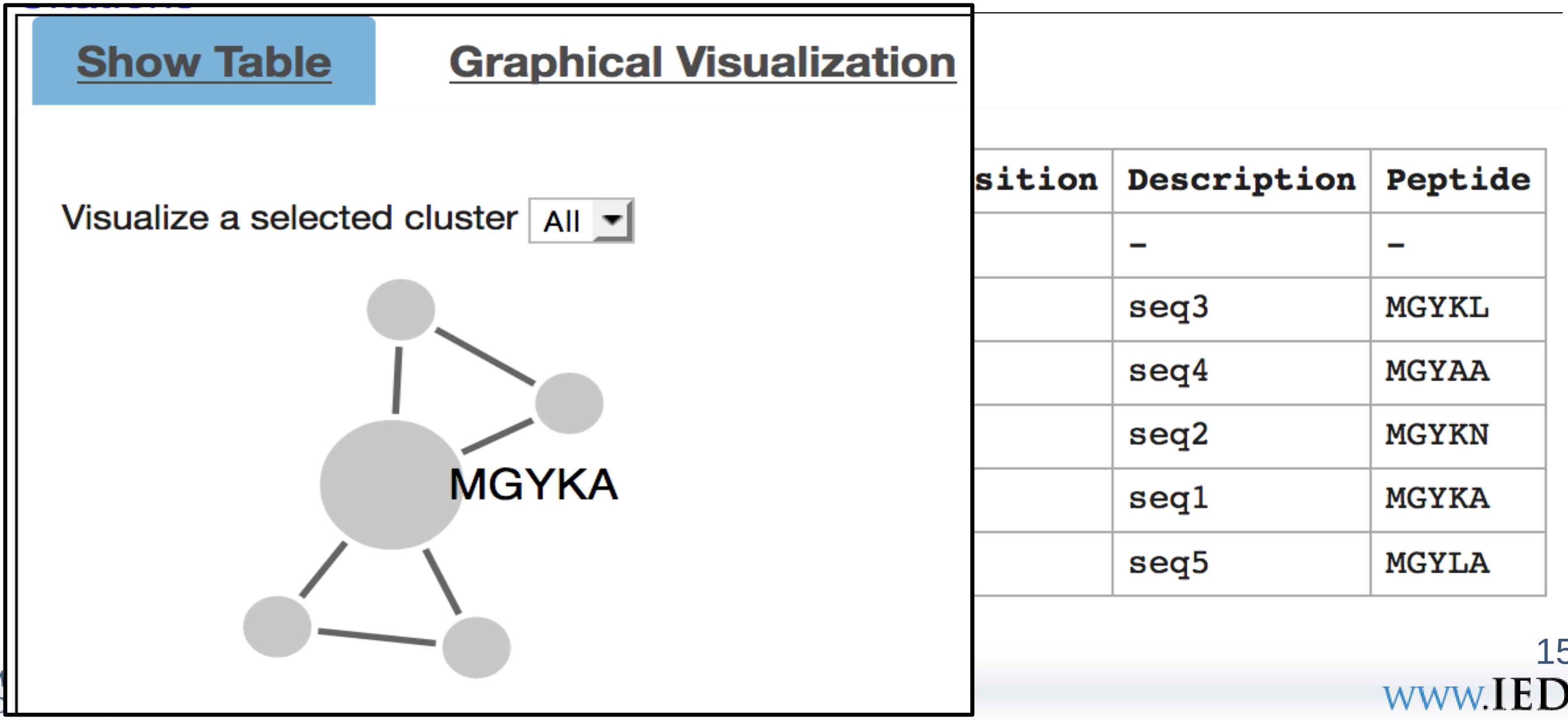
Identity cutoff : 80.0%

Minimum length of peptide selected : No minimum length selected

Maximum length of peptide selected : No maximum length selected

Clustering approach selected : All connected clusters

Number of clusters : 1



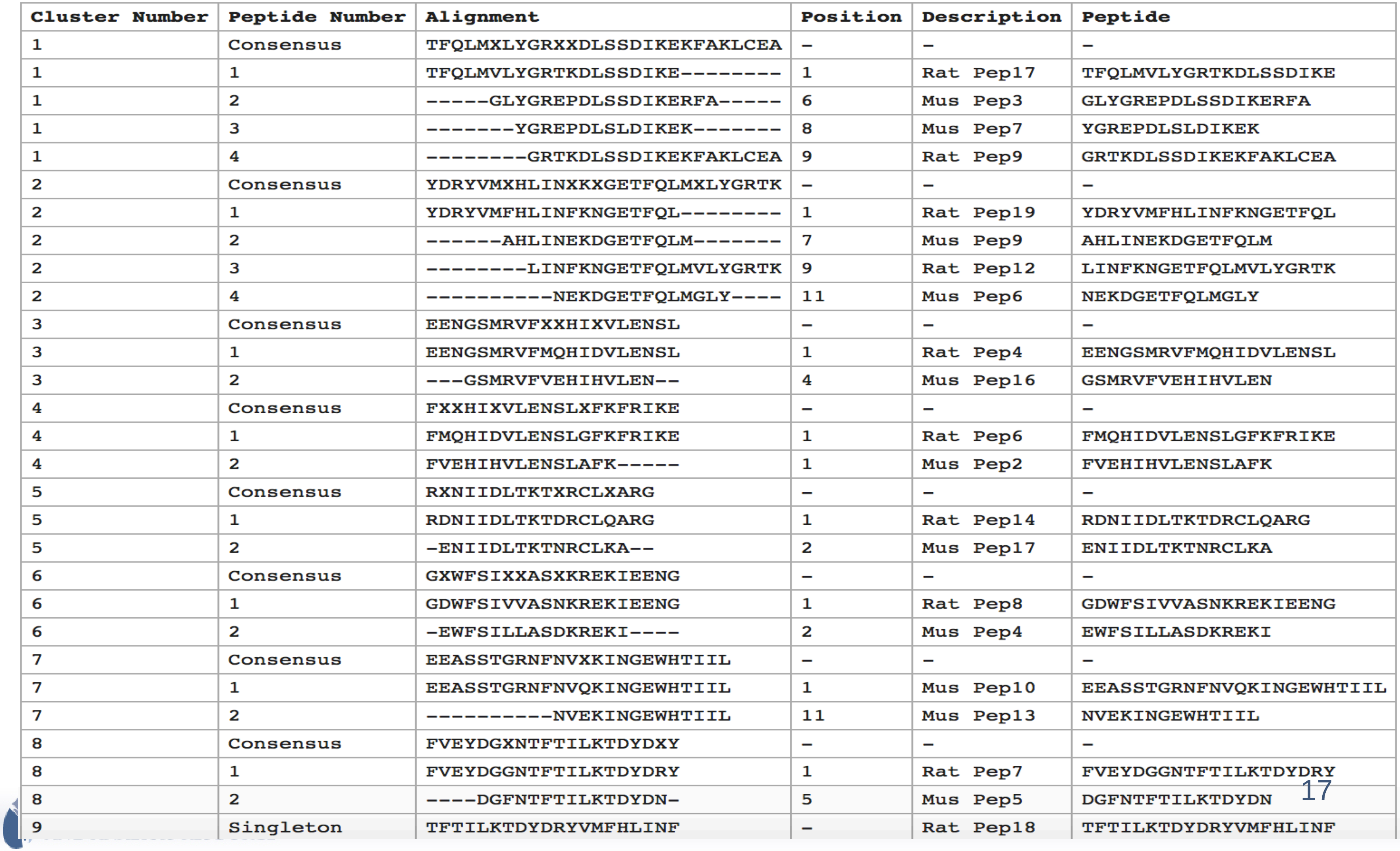
Visualize a selected cluster All

sition	Description	Peptide
	-	-
	seq3	MGYKL
	seq4	MGYAA
	seq2	MGYKN
	seq1	MGYKA
	seq5	MGYLA

Example: Cross reactivity

- Problem: We have a newly identified set of 23 T cell epitopes derived from mouse proteins recognized by mouse allergic humans
- The IEDB lists 19 T cell epitopes defined from rat proteins that are homologous to the mouse protein studied. Are the recognized epitopes related?

Clustering tool reveals groups of overlapping mouse and rat derived epitopes



Cluster Number	Peptide Number	Alignment	Position	Description	Peptide
1	Consensus	TFQLMXLYGRXXDLSSDIKEKFAKLCEA	-	-	-
1	1	TFQLMVLYGRTKDLSSDIKE--------	1	Rat Pep17	TFQLMVLYGRTKDLSSDIKE
1	2	-----GLYGREPDLSSDIKERFA-----	6	Mus Pep3	GLYGREPDLSSDIKERFA
1	3	-------YGREPDLSLDIKEK-------	8	Mus Pep7	YGREPDLSLDIKEK
1	4	--------GRTKDLSSDIKEKFAKLCEA	9	Rat Pep9	GRTKDLSSDIKEKFAKLCEA
2	Consensus	YDRYVMXHLINXKXGETFQLMXLYGRTK	-	-	-
2	1	YDRYVMFHLINFKNGETFQL--------	1	Rat Pep19	YDRYVMFHLINFKNGETFQL
2	2	------AHLINEKDGETFQLM-------	7	Mus Pep9	AHLINEKDGETFQLM
2	3	--------LINFKNGETFQLMVLYGRTK	9	Rat Pep12	LINFKNGETFQLMVLYGRTK
2	4	----------NEKDGETFQLMGLY----	11	Mus Pep6	NEKDGETFQLMGLY
3	Consensus	EENGSMRVFXXHIXVLENSL	-	-	-
3	1	EENGSMRVFMQHIDVLENSL	1	Rat Pep4	EENGSMRVFMQHIDVLENSL
3	2	---GSMRVFVEHIHVLEN--	4	Mus Pep16	GSMRVFVEHIHVLEN
4	Consensus	FXXHIXVLENSLXFKFRIKE	-	-	-
4	1	FMQHIDVLENSLGFKFRIKE	1	Rat Pep6	FMQHIDVLENSLGFKFRIKE
4	2	FVEHIHVLENSLAFK-----	1	Mus Pep2	FVEHIHVLENSLAFK
5	Consensus	RXNIIDLTKTXRCLXARG	-	-	-
5	1	RDNIIDLTKTDRCLQARG	1	Rat Pep14	RDNIIDLTKTDRCLQARG
5	2	-ENIIDLTKTNRCLKA--	2	Mus Pep17	ENIIDLTKTNRCLKA
6	Consensus	GXWFSIXXASXKREKIEENG	-	-	-
6	1	GDWFSIVVASNKREKIEENG	1	Rat Pep8	GDWFSIVVASNKREKIEENG
6	2	-EWFSILLASDKREKI----	2	Mus Pep4	EWFSILLASDKREKI
7	Consensus	EEASSTGRNFNVXKINGEWHTIIL	-	-	-
7	1	EEASSTGRNFNVQKINGEWHTIIL	1	Mus Pep10	EEASSTGRNFNVQKINGEWHTIIL
7	2	----------NVEKINGEWHTIIL	11	Mus Pep13	NVEKINGEWHTIIL
8	Consensus	FVEYDGXNTFTILKTDYDXY	-	-	-
8	1	FVEYDGGNTFTILKTDYDRY	1	Rat Pep7	FVEYDGGNTFTILKTDYDRY
8	2	----DGFNTFTILKTDYDN-	5	Mus Pep5	DGFNTFTILKTDYDN
9	Singleton	TFTILKTDYDRYVMFHLINF	-	Rat Pep18	TFTILKTDYDRYVMFHLINF

Summary of the Allergen data from mouse and rat using cluster method

- 19 peptide from Rat (Jeal et al. 2004)
- 23 peptide from Mouse (Schulten et al. 2018)
- Total clusters = 30
- Singletons=22 (10 Rat, 12 Mouse).
- Clusters with >=2 peptides = 8
- The mouse peptides that overlapped with Rat peptides were significantly more immunogenic than the singletons, suggesting that there might be a dominant immune response due to increased/repeated exposure.

RATE (Restrictor Analysis Tool for Epitopes)

http://iedb-rate.liai.org

Publications

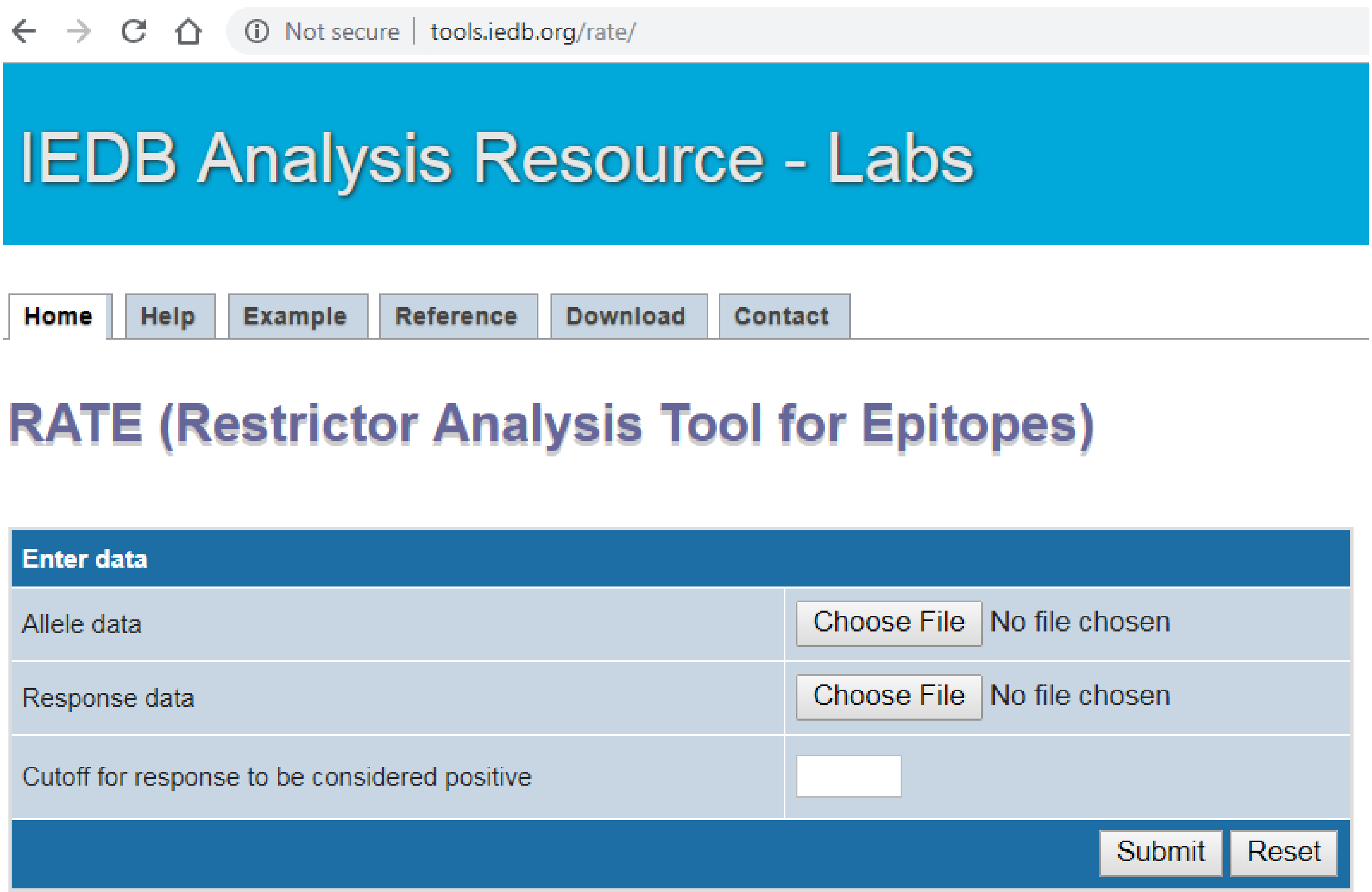
Paul, S., Arlehamn, C. S. L., … Peters, B., & Sette, A. (2017). Experimental validation of the RATE tool for inferring HLA restrictions of T cell epitopes. BMC immunology, 18(1), 20.

Paul, S., Dillon, M.B., … Peters, B. and Sette, A., 2015. A population response analysis approach to assign class II HLA-epitope restrictions. The Journal of Immunology, p.1403074

RATE

- Automated method to infer HLA restriction of given epitopes from immune response data of HLA typed subjects
- The method is based on comparing the frequency of donors who expressed (& not expressed) a particular allele and there was immune response (& no response) towards an epitope.

RATE – home page



Not secure | tools.iedb.org/rate/

IEDB Analysis Resource - Labs

RATE (Restrictor Analysis Tool for Epitopes)

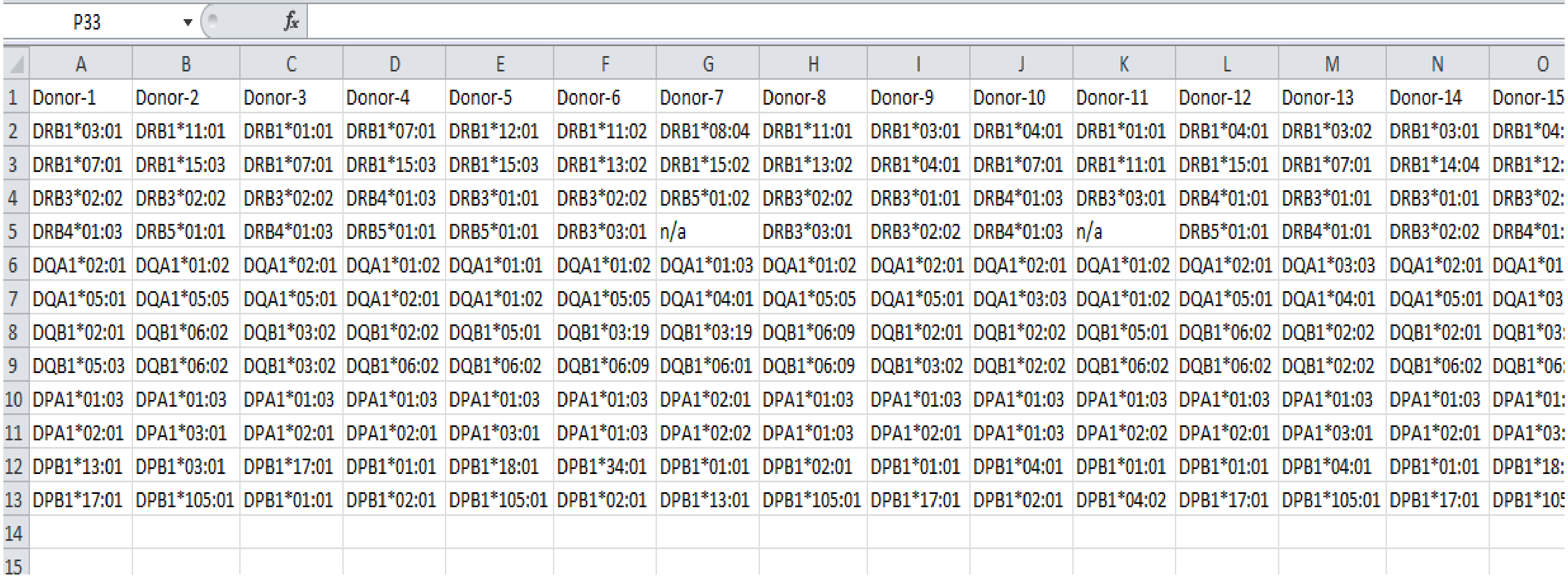
Enter data	
Allele data	Choose File No file chosen
Response data	Choose File No file chosen
Cutoff for response to be considered positive	

RATE – Input

- Two data files need to be provided - Allele data & response data
- The input data need to be in specified format
- Tab separated format in plain text file

RATE – Input: Allele data

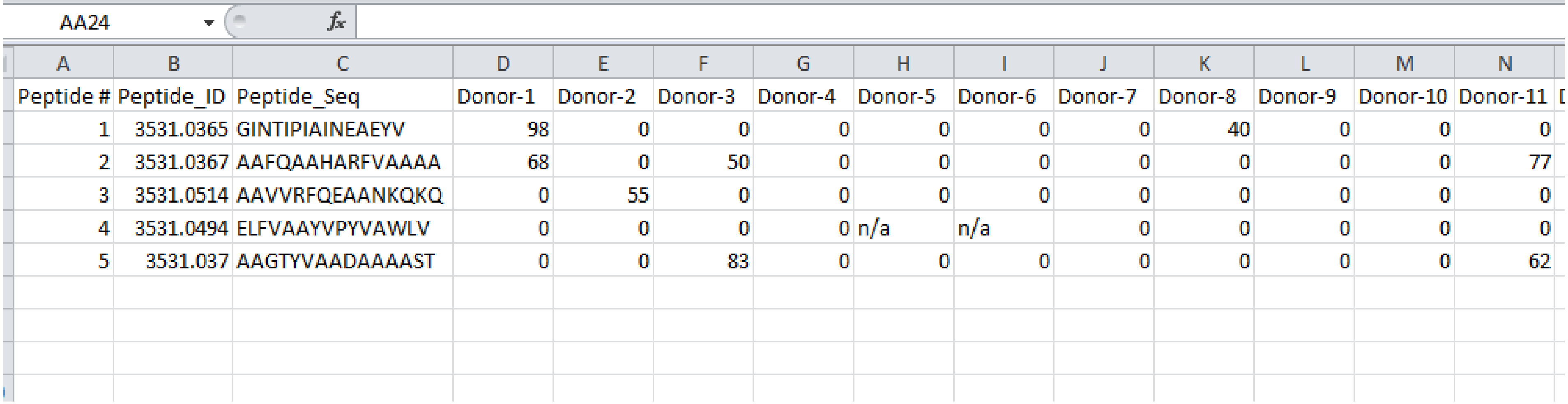
- Sample data shown here – 12 class II alleles of 6 loci for each donor are listed in separate columns



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Donor-1	Donor-2	Donor-3	Donor-4	Donor-5	Donor-6	Donor-7	Donor-8	Donor-9	Donor-10	Donor-11	Donor-12	Donor-13	Donor-14	Donor-15
2	DRB1*03:01	DRB1*11:01	DRB1*01:01	DRB1*07:01	DRB1*12:01	DRB1*11:02	DRB1*08:04	DRB1*11:01	DRB1*03:01	DRB1*04:01	DRB1*01:01	DRB1*04:01	DRB1*03:02	DRB1*03:01	DRB1*04:
3	DRB1*07:01	DRB1*15:03	DRB1*07:01	DRB1*15:03	DRB1*15:03	DRB1*13:02	DRB1*15:02	DRB1*13:02	DRB1*04:01	DRB1*07:01	DRB1*11:01	DRB1*15:01	DRB1*07:01	DRB1*14:04	DRB1*12:
4	DRB3*02:02	DRB3*02:02	DRB3*02:02	DRB4*01:03	DRB3*01:01	DRB3*02:02	DRB5*01:02	DRB3*02:02	DRB3*01:01	DRB4*01:03	DRB3*03:01	DRB4*01:01	DRB3*01:01	DRB3*01:01	DRB3*02:
5	DRB4*01:03	DRB5*01:01	DRB4*01:03	DRB5*01:01	DRB5*01:01	DRB3*03:01	n/a	DRB3*03:01	DRB3*02:02	DRB4*01:03	n/a	DRB5*01:01	DRB4*01:01	DRB3*02:02	DRB4*01:
6	DQA1*02:01	DQA1*01:02	DQA1*02:01	DQA1*01:02	DQA1*01:01	DQA1*01:02	DQA1*01:03	DQA1*01:02	DQA1*02:01	DQA1*02:01	DQA1*01:02	DQA1*02:01	DQA1*03:03	DQA1*02:01	DQA1*01
7	DQA1*05:01	DQA1*05:05	DQA1*05:01	DQA1*02:01	DQA1*01:02	DQA1*05:05	DQA1*04:01	DQA1*05:05	DQA1*05:01	DQA1*03:03	DQA1*01:02	DQA1*05:01	DQA1*04:01	DQA1*05:01	DQA1*03
8	DQB1*02:01	DQB1*06:02	DQB1*03:02	DQB1*02:02	DQB1*05:01	DQB1*03:19	DQB1*03:19	DQB1*06:09	DQB1*02:01	DQB1*02:02	DQB1*05:01	DQB1*06:02	DQB1*02:02	DQB1*02:01	DQB1*03:
9	DQB1*05:03	DQB1*06:02	DQB1*03:02	DQB1*06:02	DQB1*06:02	DQB1*06:09	DQB1*06:01	DQB1*06:09	DQB1*03:02	DQB1*02:02	DQB1*06:02	DQB1*06:02	DQB1*02:02	DQB1*06:02	DQB1*06:
10	DPA1*01:03	DPA1*01:03	DPA1*01:03	DPA1*01:03	DPA1*01:03	DPA1*01:03	DPA1*02:01	DPA1*01:03	DPA1*01:03	DPA1*01:03	DPA1*01:03	DPA1*01:03	DPA1*01:03	DPA1*01:03	DPA1*01:
11	DPA1*02:01	DPA1*03:01	DPA1*02:01	DPA1*02:01	DPA1*03:01	DPA1*01:03	DPA1*02:02	DPA1*01:03	DPA1*02:01	DPA1*01:03	DPA1*02:02	DPA1*02:01	DPA1*03:01	DPA1*02:01	DPA1*03:
12	DPB1*13:01	DPB1*03:01	DPB1*17:01	DPB1*01:01	DPB1*18:01	DPB1*34:01	DPB1*01:01	DPB1*02:01	DPB1*01:01	DPB1*04:01	DPB1*01:01	DPB1*01:01	DPB1*04:01	DPB1*01:01	DPB1*18:
13	DPB1*17:01	DPB1*105:01	DPB1*01:01	DPB1*02:01	DPB1*105:01	DPB1*02:01	DPB1*13:01	DPB1*105:01	DPB1*17:01	DPB1*02:01	DPB1*04:02	DPB1*17:01	DPB1*105:01	DPB1*17:01	DPB1*10
14															
15															

RATE – Input: Response data

- The response data (here SFC values) for each epitope in each of the donors are provided



Peptide #	Peptide_ID	Peptide_Seq	Donor-1	Donor-2	Donor-3	Donor-4	Donor-5	Donor-6	Donor-7	Donor-8	Donor-9	Donor-10	Donor-11
1	3531.0365	GINTIPIAINEAEYV	98	0	0	0	0	0	0	40	0	0	0
2	3531.0367	AAFQAAHARFVAAAA	68	0	50	0	0	0	0	0	0	0	77
3	3531.0514	AAVVRFQEAANKQKQ	0	55	0	0	0	0	0	0	0	0	0
4	3531.0494	ELFVAAYVPYVAWLV	0	0	0	0	n/a	n/a	0	0	0	0	0
5	3531.037	AAGTYVAADAAAAST	0	0	83	0	0	0	0	0	0	0	62

RATE – how it works

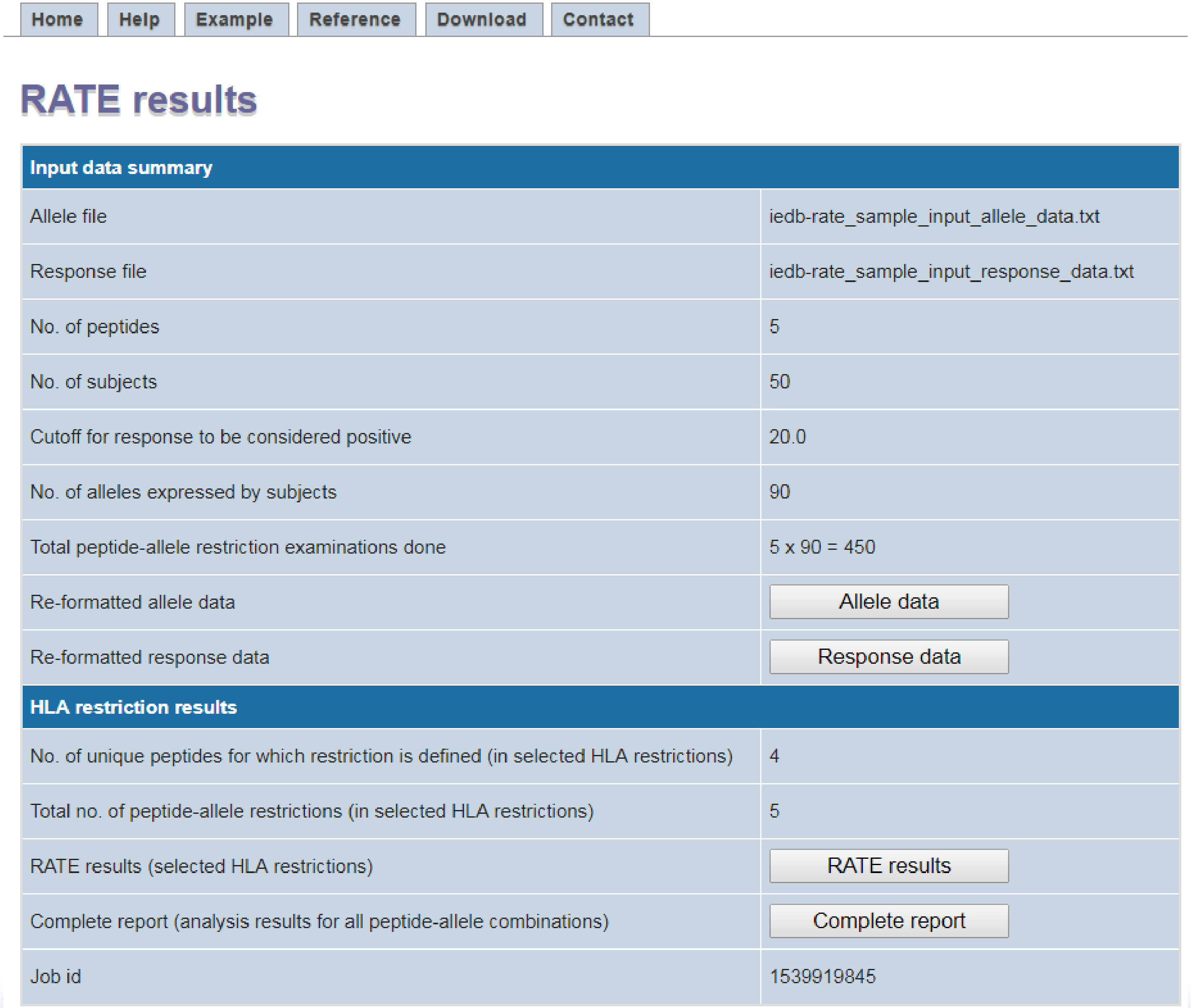
- Restrictions are determined based on
  - Odds ratio and significance estimated using Fisher's exact test.

$$OR = \frac{(A^{+}R^{+}) \times (A^{-}R^{-})}{(A^{-}R^{+}) \times (A^{+}R^{-})}$$

  - “Relative frequency”, a parameter estimated by the tool based on the frequency of alleles and donors responded

$$RF = \frac{A^{+}R^{+} / (A^{+}R^{+} + A^{+}R^{-})}{(A^{+}R^{+} + A^{-}R^{+}) / \text{Total donors}}$$

RATE – results

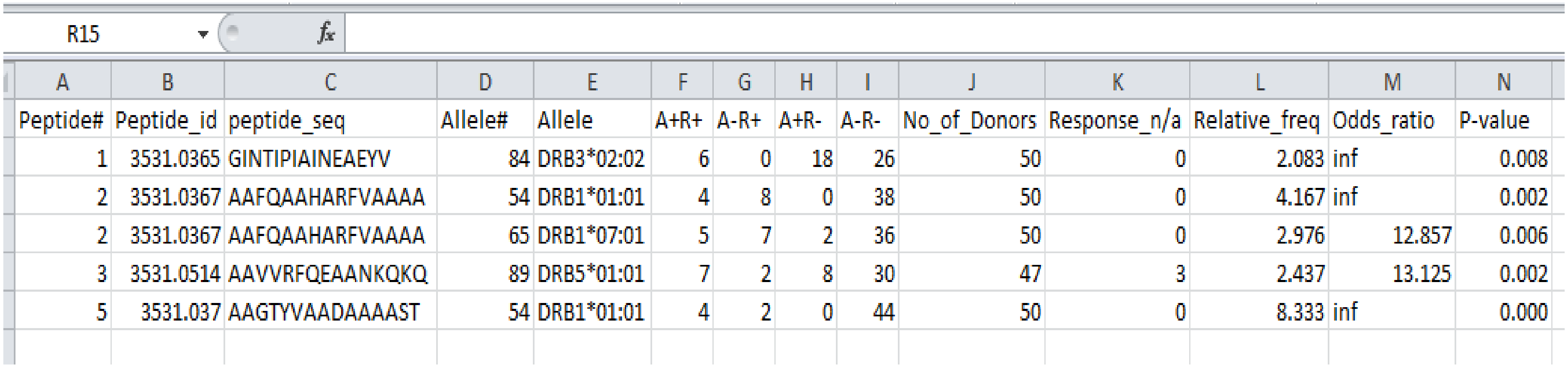


RATE results

Input data summary	
Allele file	iedb-rate_sample_input_allele_data.txt
Response file	iedb-rate_sample_input_response_data.txt
No. of peptides	5
No. of subjects	50
Cutoff for response to be considered positive	20.0
No. of alleles expressed by subjects	90
Total peptide-allele restriction examinations done	5 x 90 = 450
Re-formatted allele data	Allele data
Re-formatted response data	Response data
HLA restriction results	
No. of unique peptides for which restriction is defined (in selected HLA restrictions)	4
Total no. of peptide-allele restrictions (in selected HLA restrictions)	5
RATE results (selected HLA restrictions)	RATE results
Complete report (analysis results for all peptide-allele combinations)	Complete report
Job id	1539919845

RATE – results

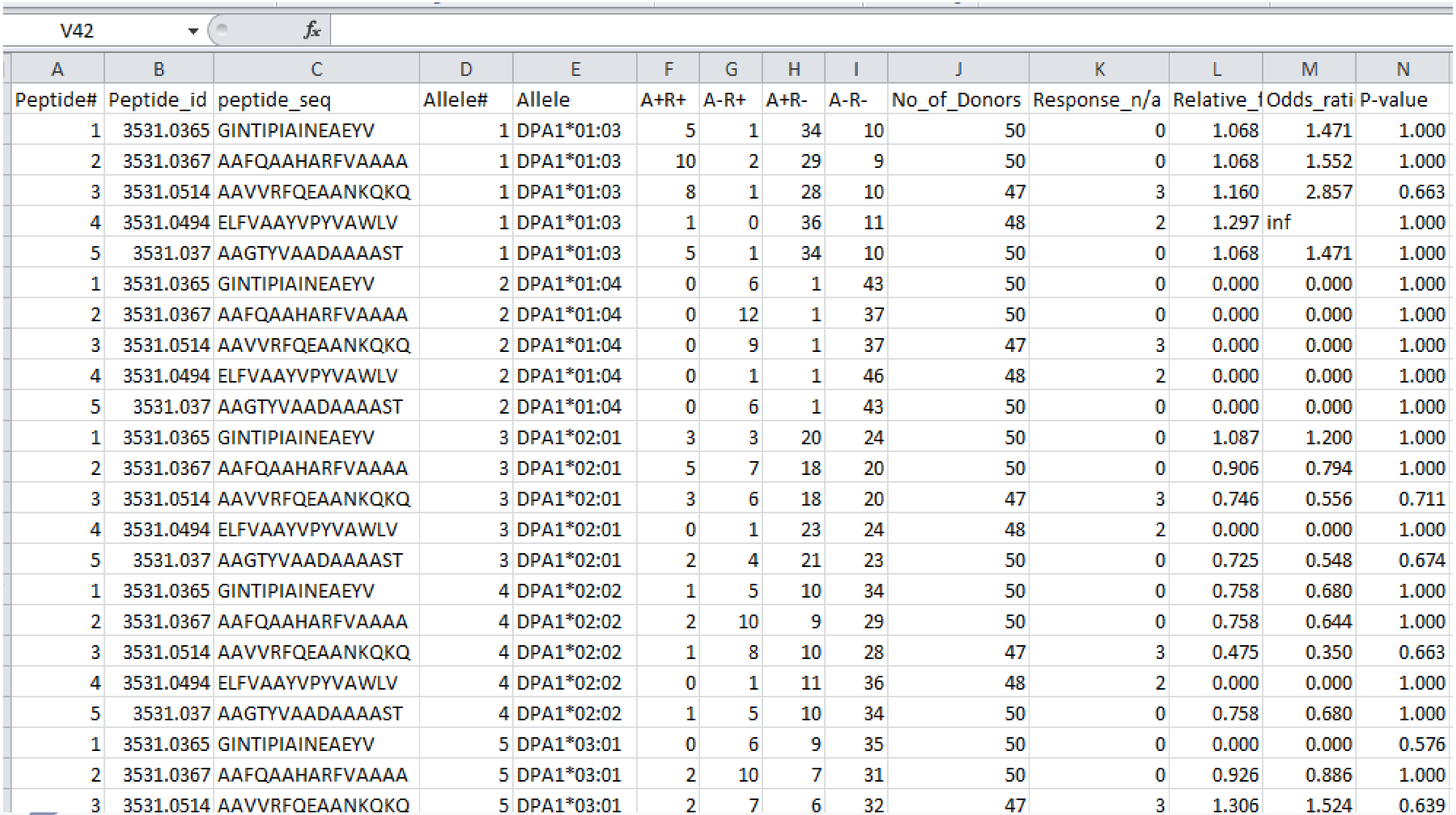
- Epitope-allele combinations with RF ≥ 1.3 and p-value < 0.01 are considered significant.
- Significant results selected by RATE for the example files



A	B	C	D	E	F	G	H	I	J	K	L	M	N
Peptide#	Peptide_id	peptide_seq	Allele#	Allele	A+R+	A-R+	A+R-	A-R-	No_of_Donors	Response_n/a	Relative_freq	Odds_ratio	P-value
1	3531.0365	GINTIPIAINEAEYV	84	DRB3*02:02	6	0	18	26	50	0	2.083	inf	0.008
2	3531.0367	AAFQAAHARFVAAAA	54	DRB1*01:01	4	8	0	38	50	0	4.167	inf	0.002
2	3531.0367	AAFQAAHARFVAAAA	65	DRB1*07:01	5	7	2	36	50	0	2.976	12.857	0.006
3	3531.0514	AAVVRFQEAANKQKQ	89	DRB5*01:01	7	2	8	30	47	3	2.437	13.125	0.002
5	3531.037	AAGTYVAADAAAAST	54	DRB1*01:01	4	2	0	44	50	0	8.333	inf	0.000

RATE – complete report

- Complete report - all epitopes



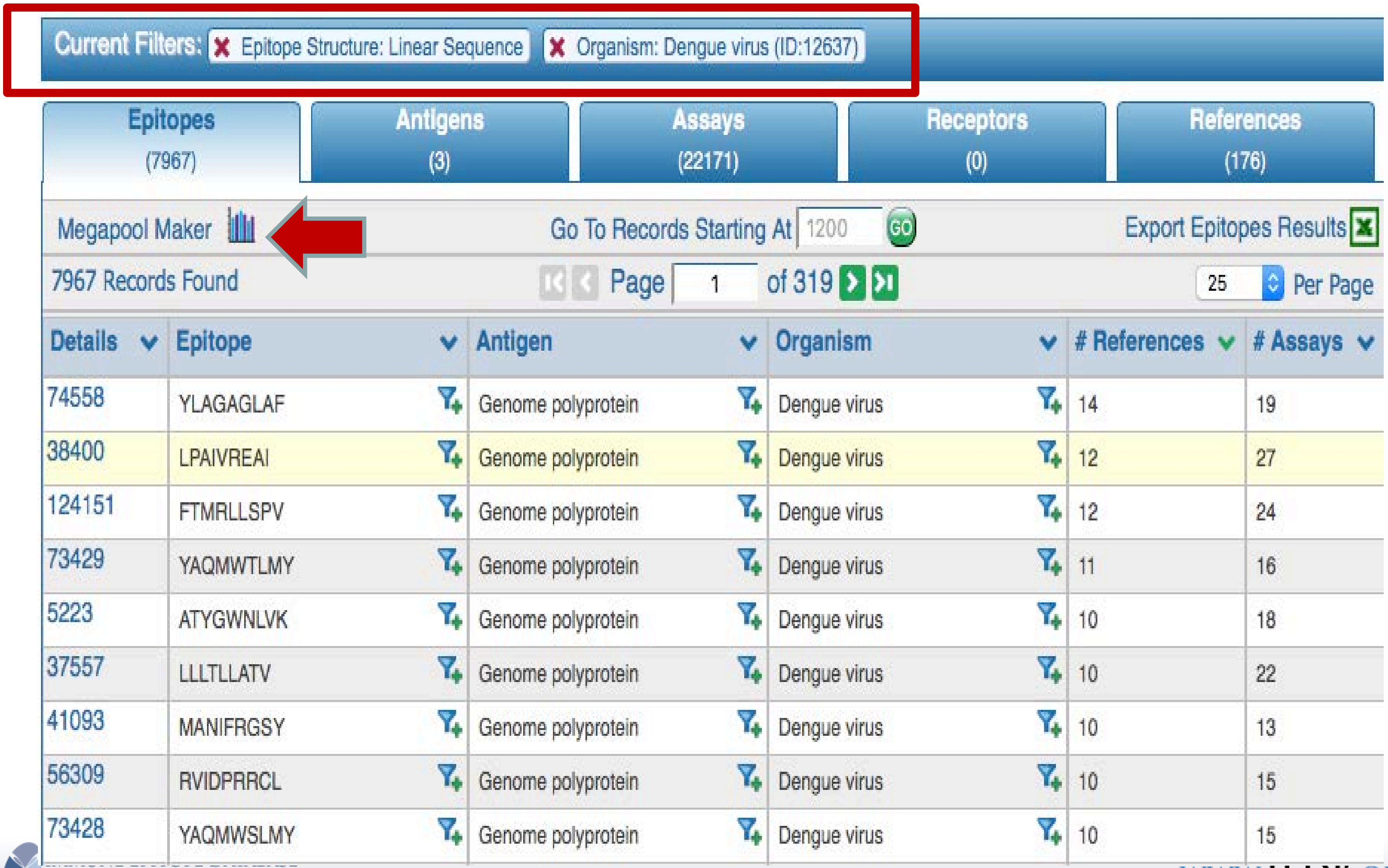
A	B	C	D	E	F	G	H	I	J	K	L	M	N
Peptide#	Peptide_id	peptide_seq	Allele#	Allele	A+R+	A-R+	A+R-	A-R-	No_of_Donors	Response_n/a	Relative_	Odds_rati	P-value
1	3531.0365	GINTIPIAINEAEYV	1	DPA1*01:03	5	1	34	10	50	0	1.068	1.471	1.000
2	3531.0367	AAFQAAHARFVAAAA	1	DPA1*01:03	10	2	29	9	50	0	1.068	1.552	1.000
3	3531.0514	AAVVRFQEAANKQKQ	1	DPA1*01:03	8	1	28	10	47	3	1.160	2.857	0.663
4	3531.0494	ELFVAAYVPYVAWLV	1	DPA1*01:03	1	0	36	11	48	2	1.297	inf	1.000
5	3531.037	AAGTYVAADAAAAST	1	DPA1*01:03	5	1	34	10	50	0	1.068	1.471	1.000
1	3531.0365	GINTIPIAINEAEYV	2	DPA1*01:04	0	6	1	43	50	0	0.000	0.000	1.000
2	3531.0367	AAFQAAHARFVAAAA	2	DPA1*01:04	0	12	1	37	50	0	0.000	0.000	1.000
3	3531.0514	AAVVRFQEAANKQKQ	2	DPA1*01:04	0	9	1	37	47	3	0.000	0.000	1.000
4	3531.0494	ELFVAAYVPYVAWLV	2	DPA1*01:04	0	1	1	46	48	2	0.000	0.000	1.000
5	3531.037	AAGTYVAADAAAAST	2	DPA1*01:04	0	6	1	43	50	0	0.000	0.000	1.000
1	3531.0365	GINTIPIAINEAEYV	3	DPA1*02:01	3	3	20	24	50	0	1.087	1.200	1.000
2	3531.0367	AAFQAAHARFVAAAA	3	DPA1*02:01	5	7	18	20	50	0	0.906	0.794	1.000
3	3531.0514	AAVVRFQEAANKQKQ	3	DPA1*02:01	3	6	18	20	47	3	0.746	0.556	0.711
4	3531.0494	ELFVAAYVPYVAWLV	3	DPA1*02:01	0	1	23	24	48	2	0.000	0.000	1.000
5	3531.037	AAGTYVAADAAAAST	3	DPA1*02:01	2	4	21	23	50	0	0.725	0.548	0.674
1	3531.0365	GINTIPIAINEAEYV	4	DPA1*02:02	1	5	10	34	50	0	0.758	0.680	1.000
2	3531.0367	AAFQAAHARFVAAAA	4	DPA1*02:02	2	10	9	29	50	0	0.758	0.644	1.000
3	3531.0514	AAVVRFQEAANKQKQ	4	DPA1*02:02	1	8	10	28	47	3	0.475	0.350	0.663
4	3531.0494	ELFVAAYVPYVAWLV	4	DPA1*02:02	0	1	11	36	48	2	0.000	0.000	1.000
5	3531.037	AAGTYVAADAAAAST	4	DPA1*02:02	1	5	10	34	50	0	0.758	0.680	1.000
1	3531.0365	GINTIPIAINEAEYV	5	DPA1*03:01	0	6	9	35	50	0	0.000	0.000	0.576
2	3531.0367	AAFQAAHARFVAAAA	5	DPA1*03:01	2	10	7	31	50	0	0.926	0.886	1.000
3	3531.0514	AAVVRFQEAANKQKQ	5	DPA1*03:01	2	7	6	32	47	3	1.306	1.524	0.639

EPIFILTER

- Generate a validated set of reference epitopes for B cell, MHC class I and Class II
- Available with IEDB query search results
- Different filters can be applied to epitopes and assays
- Recommended settings are set as default and can be changed in “advanced options”
- Results can be exported in csv file.

Ref: Pro et. al. 2015, Journal of Immunology Research

EPIFILTER – tool link



Current Filters: ✗ Epitope Structure: Linear Sequence ✗ Organism: Dengue virus (ID:12637)

Go To Records Starting At: 1200

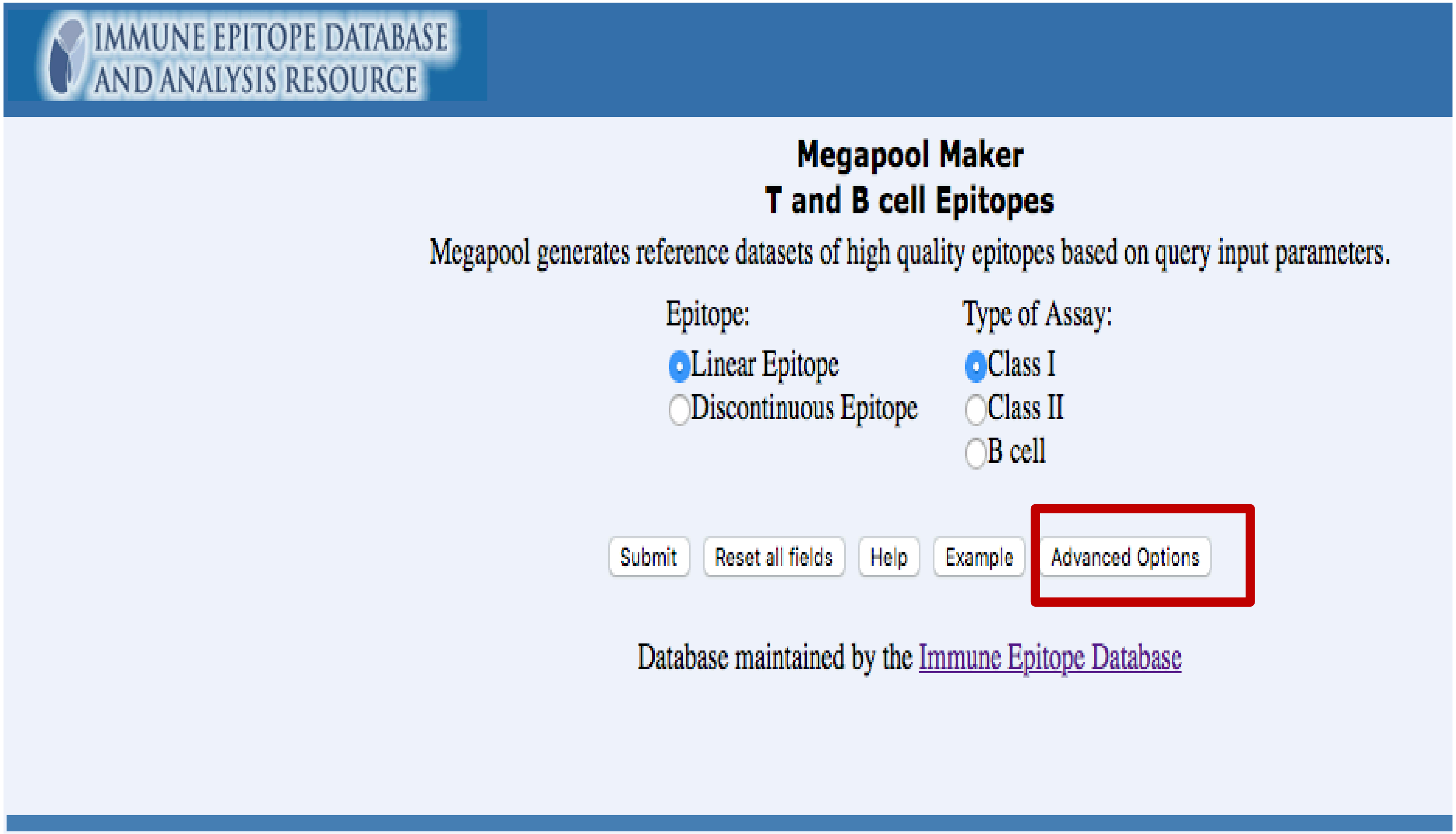
7967 Records Found

Page 1 of 319

25 Per Page

Details	Epitope	Antigen	Organism	# References	# Assays
74558	YLAGAGLAF	Genome polyprotein	Dengue virus	14	19
38400	LPAIVREAI	Genome polyprotein	Dengue virus	12	27
124151	FTMRLLSPV	Genome polyprotein	Dengue virus	12	24
73429	YAQMWTLMY	Genome polyprotein	Dengue virus	11	16
5223	ATYGWNLVK	Genome polyprotein	Dengue virus	10	18
37557	LLLTLLATV	Genome polyprotein	Dengue virus	10	22
41093	MANIFRGSY	Genome polyprotein	Dengue virus	10	13
56309	RVIDPRRCL	Genome polyprotein	Dengue virus	10	15
73428	YAQMWSLMY	Genome polyprotein	Dengue virus	10	15

EPIFILTER – parameter selection page



Megapool Maker
T and B cell Epitopes

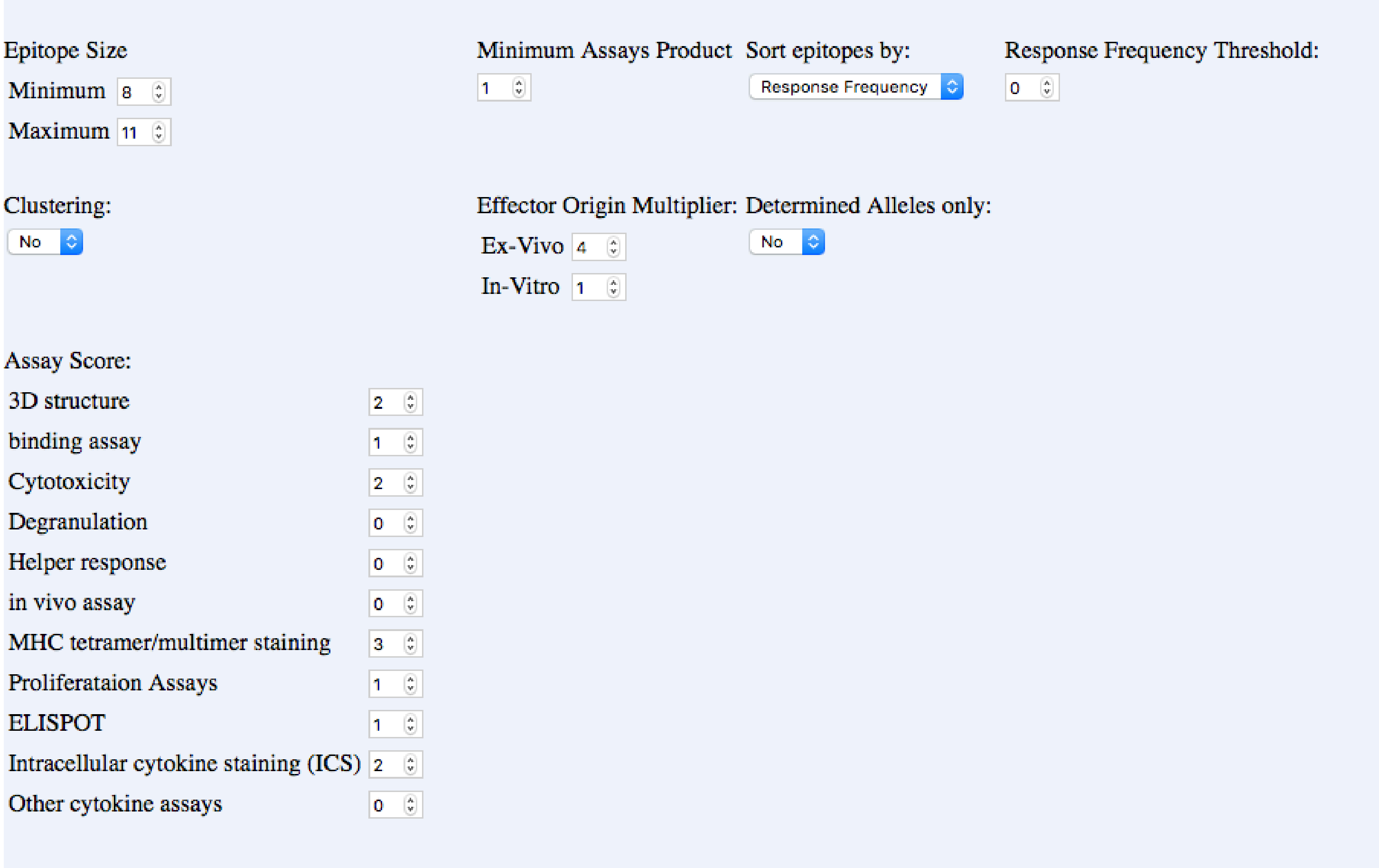
Megapool generates reference datasets of high quality epitopes based on query input parameters.

Epitope:
● Linear Epitope
○ Discontinuous Epitope

Type of Assay:
● Class I
○ Class II
○ B cell

Database maintained by the Immune Epitope Database

EPIFILTER – advanced options



Epitope Size
Minimum 8
Maximum 11

Minimum Assays Product
1

Sort epitopes by:
Response Frequency

Response Frequency Threshold:
0

Clustering:
No

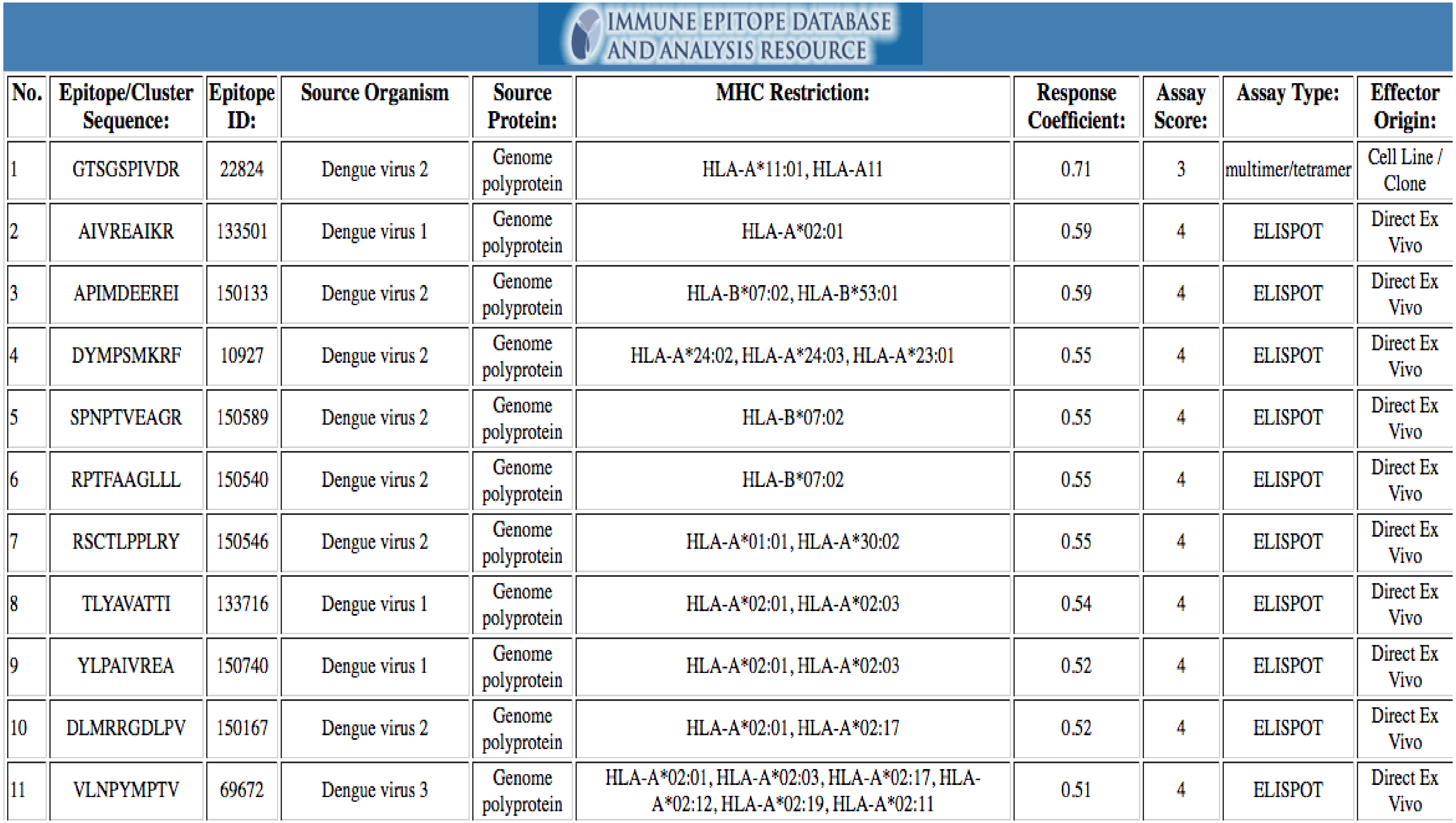
Effector Origin Multiplier:
Ex-Vivo 4
In-Vitro 1

Determined Alleles only:
No

Assay Score:

3D structure	2
binding assay	1
Cytotoxicity	2
Degranulation	0
Helper response	0
in vivo assay	0
MHC tetramer/multimer staining	3
Proliferataion Assays	1
ELISPOT	1
Intracellular cytokine staining (ICS)	2
Other cytokine assays	0

EPIFILTER – result page



No.	Epitope/Cluster Sequence:	Epitope ID:	Source Organism	Source Protein:	MHC Restriction:	Response Coefficient:	Assay Score:	Assay Type:	Effector Origin:
1	GTSGSPIVDR	22824	Dengue virus 2	Genome polyprotein	HLA-A*11:01, HLA-A11	0.71	3	multimer/tetramer	Cell Line / Clone
2	AIVREAIKR	133501	Dengue virus 1	Genome polyprotein	HLA-A*02:01	0.59	4	ELISPOT	Direct Ex Vivo
3	APIMDEEREI	150133	Dengue virus 2	Genome polyprotein	HLA-B*07:02, HLA-B*53:01	0.59	4	ELISPOT	Direct Ex Vivo
4	DYMPSMKRF	10927	Dengue virus 2	Genome polyprotein	HLA-A*24:02, HLA-A*24:03, HLA-A*23:01	0.55	4	ELISPOT	Direct Ex Vivo
5	SPNPTVEAGR	150589	Dengue virus 2	Genome polyprotein	HLA-B*07:02	0.55	4	ELISPOT	Direct Ex Vivo
6	RPTFAAGLLL	150540	Dengue virus 2	Genome polyprotein	HLA-B*07:02	0.55	4	ELISPOT	Direct Ex Vivo
7	RSCTLPPLRY	150546	Dengue virus 2	Genome polyprotein	HLA-A*01:01, HLA-A*30:02	0.55	4	ELISPOT	Direct Ex Vivo
8	TLYAVATTI	133716	Dengue virus 1	Genome polyprotein	HLA-A*02:01, HLA-A*02:03	0.54	4	ELISPOT	Direct Ex Vivo
9	YLPAIVREA	150740	Dengue virus 1	Genome polyprotein	HLA-A*02:01, HLA-A*02:03	0.52	4	ELISPOT	Direct Ex Vivo
10	DLMRRGDLPV	150167	Dengue virus 2	Genome polyprotein	HLA-A*02:01, HLA-A*02:17	0.52	4	ELISPOT	Direct Ex Vivo
11	VLNPYMPTV	69672	Dengue virus 3	Genome polyprotein	HLA-A*02:01, HLA-A*02:03, HLA-A*02:17, HLA-A*02:12, HLA-A*02:19, HLA-A*02:11	0.51	4	ELISPOT	Direct Ex Vivo

Questions?