Contract No. HHSN266200400006C

Immune Epitope Database and Analysis Program

2007 Annual IEDB Compendium

La Jolla Institute for Allergy and Immunology
9420 Athena Circle
La Jolla, CA 92037

858-752-6923
858-752-6987 (fax)
wfleri@liai.org

January 2008

Table of Contents

Introduction

The Immune Epitope Database and Analysis Resource (IEDB) is a public repository of immune epitope data and is sponsored by the National Institute for Allergy and Infectious Diseases (NIAID). The IEDB development started in December 2003 and it became available to the public in a beta test phase on 15 February 2006. The IEDB contains data related to antibody and T cell epitopes for humans, non-human primates, rodents, and other animal species. The IEDB also makes available a variety of analytical and epitope prediction tools and resources within its Analysis Resource.

This third Annual Compendium of the Immune Epitope Database and Analysis Resource consists of three sections. The first section contains a list of the antibody and T cell epitope information in the database as of 23 January 2008. The second section lists and describes the various features of the IEDB website available by the end of 2007. The third section lists the scientific publications in 2007 for which the IEDB played a contributory role.

Since the publication of last year's 2006 Annual Compendium, the quantity of data available in the IEDB has increased significantly with the addition of over 2000 fully curated references and over 30,000 records from direct data submissions from Large Scale Epitope Discovery contractors. By the end of the year, curation of data relating to NIAID Category A, B, and C priority pathogens, NIAID Emerging and Re-emerging infectious diseases, Malaria, Hepatitis B, Clostridium tetani, Leishmania, and Candida albicans was current through 30 September 2007. Curation of herpesviruses was almost complete and curation of allergen epitopes was in progress.

A number of new features were introduced to the website during 2007. They are listed below, and further information about them is available in Section 2 of this document.

- Tree structure for allele browser
- Browse by 3D structure
- Patent page
- Epitope dataset forum
- New XML format for export
- Antibody prediction tools BepiPred and DiscoTope
- ANN methods for T cell epitope prediction expanded to allow sequence lengths greater than nine
- MHC Class II epitope prediction methods expanded from only ARB to ARB, Smm_align, Sturniolo, and consensus methods
- Epitope cluster analysis

1 Antibody and T Cell Epitopes

Many new references and many new pathogens were added to the IEDB in 2007, as demonstrated in Table 1. The table lists the number of distinct B cell and T cell epitopes in the database by source species at the end of 2006 and 2007. Of the 848 source species listed, 275 were added in 2007. It should be noted that the source species is the species from which the epitopes originate, and may not be from an infecting organism or vaccine target. The curation of MHC binding peptides, cross-reactive epitopes, and autoimmune epitopes has resulted in the appearance of human and mouse epitopes on the list. Although the curation of autoimmune epitopes has not been a priority to date, they have entered the IEDB as a result of importing large epitope datasets and of curating them when they appear in priority infectious disease references.

The NCBI taxonomy ID is provided for each source species and strain. This list differs from the one available on the **Browse Records by Source Species** (http://immuneepitope.org/browse.do?dispatch=loadSpeciesBrowser) web page because the table below only counts distinct epitopes, not records. A **distinct epitope** is an epitope with given characteristics, regardless of the reference, that has a positive qualitative binding value for antigen or MHC association. The **Browse Records by Source Species** (http://immuneepitope.org/browse.do?dispatch=loadAlleleBrowser) list provides a count of records, which includes molecular structures in the database that have positive or negative qualitative binding values. A **record** contains data about one epitope (or structure) in one reference, where a reference is a scientific article or a direct data submission from a researcher. The IEDB collects both positive and negative data, a helpful feature for epitope prediction tool developers. The reader can obtain more information about the epitopes of a specific source species by going to http://www.immuneEpitope.org and performing a query.

In the table below, the leftmost column labeled "New 2007" indicates with an "X" if epitopes for the species/strain was added to the IEDB in 2007. A font color of red is used to highlight the information in the row. The columns labeled "B-06", "T-06", "B-07", and "T-07" indicate the cumulative number of distinct B cell and T cell epitopes in the database at the end of 2006 and 2007, respectively. The two rightmost columns display the differences in the B and T cell epitope counts from 2006 to 2007. The changes in B and T cell epitope counts are shown in red. In 2007, the number of distinct B cell epitopes increased by 5489, from 4122 to 9611, and the number of distinct T cell epitopes increased by 10538, from 16738 to 27276.

Table 1 Summary of B and T cell epitopes contained in the IEDB

New 2007	NCBI TAX ID	SPECIES/STRAIN	B-06	T-06	B-07	T-07	Δ B	Δ T
	5755	Acanthamoeba castellanii	14		14	1		1
	715	Actinobacillus pleuropneumoniae	2		2			
	714	Aggregatibacter actinomycetemcomitans		4		4		
X	11790	AKT8 murine leukemia virus				1		1
X	28314	Aleutian mink disease virus			1		1	
	1049	Allochromatium vinosum		1		1		
	5599	Alternaria alternata	3		4		1	
X	4212	Ambrosia artemisiifolia			1	2	1	2
	4215	Ambrosia elatior		1		4		3
X	4214	Ambrosia trifida				3		3
	171929	Anacardium occidentale	11		27		16	
	770	Anaplasma marginale	14	27	14	32		5
	320483	Anaplasma marginale str. Florida		4		8		4
	46607	Andes virus	50		50			
X	7460	Apis mellifera			4	131	4	131
X	4045	Apium graveolens				1		1
X	3702	Arabidopsis thaliana				2		2
	3818	Arachis hypogaea	41		48	26	7	26
X	4220	Artemisia vulgaris				18		18
X	5085	Aspergillus fumigatus			21	5	21	5
	7604	Asterias rubens		1		1		
X	79685	Avian erythroblastosis virus (strain ES4)			1		1	
	172851	Avian hepatitis E virus	7		7			
	195700	Avian rotavirus PO-13	4		4			
X	5866	Babesia bigemina				3		3
	5865	Babesia bovis		10		12		2
	5868	Babesia microti	2		2			
X	120505	Baboon cytomegalovirus				1		1
X	11764	Baboon endogenous virus strain M7				1		1
X	1386	Bacillus			5		5	
	1392	Bacillus anthracis	16	2	30	53	14	51
	1396	Bacillus cereus		2		2		
	1402	Bacillus licheniformis		5		5		
X	1409	Bacillus sp.			1		1	
	2	Bacteria	49		76		27	
	349344	Bat SARS coronavirus Rp3		236		236		
X	63673	Batillus cornutus			2		2	
	37962	Bayou virus	1		1			
X	31715	Bean-pod mottle virus (strain Kentucky G7)				1		1
X	3505	Betula pendula			11	149	11	149
X	10629	BK polyomavirus				3		3
X	271108	Bombyx mori NPV			2		2	
X	518	Bordetella bronchiseptica				1		1
	520	Bordetella pertussis	2		10	2	8	2
	12455	Borna disease virus		4	8	5	8	1

New 2007	NCBI TAX ID	SPECIES/STRAIN	B-06	T-06	B-07	T-07	Δ B	Δ T
	29518	Borrelia afzelii		1		1		
	139	Borrelia burgdorferi	76	55	113	80	37	25
	224326	Borrelia burgdorferi B31	5	5	5	5		
	40834	Borrelia duttonii	7		7			
	29519	Borrelia garinii	3		6		3	
	9913	Bos taurus	25	41	182	121	157	80
X	12064	Bovine enterovirus			13		13	
	11303	Bovine ephemeral fever virus	2		2			
X	10320	Bovine herpesvirus 1			4		4	
X	35244	Bovine herpesvirus 5			2		2	
X	79889	Bovine herpesvirus type 1.1			4	24	4	24
X	10323	Bovine herpesvirus type 1.1 (strain Cooper)			6		6	
X	10324	Bovine herpesvirus type 1.1 (strain P8-2)			1		1	
X	10562	Bovine papillomavirus - 4			3	3	3	3
X	10559	Bovine papillomavirus type 1			1		1	
X	11215	Bovine parainfluenza virus 3				1		1
X	11246	Bovine respiratory syncytial virus			1		1	
	31611	Bovine respiratory syncytial virus (strain 391-2)		1		1		
X	10927	Bovine rotavirus			2		2	
	11099	Bovine viral diarrhea virus 1	1		1			
	3707	Brassica juncea	9		9			
	235	Brucella abortus	3	3	3	3		
	29459	Brucella melitensis	4	16	5	17	1	1
X	236	Brucella ovis			2		2	
	29461	Brucella suis		2		2		
X	89462	Bubalus bubalis				1		1
X	350702	Burkholderia cenocepacia PC184				1		1
	13373	Burkholderia mallei		1		1		
	243160	Burkholderia mallei ATCC 23344		1		1		
	6239	Caenorhabditis elegans		3		4		1
	32019	Campylobacter fetus subsp. fetus	3		3			
	197	Campylobacter jejuni		5		5		
	5476	Candida albicans	2	16	81	42	79	26
X	42374	Candida dubliniensis			1		1	
X	292348	Canine calicivirus (strain 48)			2		2	
X	11232	Canine distemper virus			5		5	
X	11233	Canine distemper virus strain Onderstepoort			1	1	1	1
X	35258	Canine oral papillomavirus				25		25
X	31597	Canine parvovirus (TYPE 2 / STRAIN A72)			2		2	
X	9615	Canis familiaris				82		82
	11662	Caprine arthritis encephalitis virus (STRAIN G63)	6		6			
X	10325	Cercopithecine herpesvirus 1			1		1	
X	45455	Cercopithecine herpesvirus 15				18		18
X	47929	Cercopithecine herpesvirus 8				67		67

New 2007	NCBI TAX ID	SPECIES/STRAIN	B-06	T-06	B-07	T-07	Δ B	Δ T
	9863	Cervus nippon		1		1		
X	13415	Chamaecyparis obtusa				27		27
X	7154	Chironomus thummi			2	3	2	3
	7155	Chironomus thummi thummi	3		12	36	9	36
	813	Chlamydia trachomatis	11		38	25	27	25
X	315277	Chlamydia trachomatis A/HAR-13			5	1	5	1
X	272561	Chlamydia trachomatis D/UW-3/CX				25		25
	83555	Chlamydophila abortus	5		5			
	218497	Chlamydophila abortus S26/3	9		9			
	83558	Chlamydophila pneumoniae		4		11		7
	182082	Chlamydophila pneumoniae TW-183		9		14		5
	83554	Chlamydophila psittaci	64		64			
	9534	Chlorocebus aethiops		8		9		1
X	2711	Citrus sinensis			5		5	
X	11096	Classical swine fever virus			10		10	
	214432	Cloning vector pscFvCA-E8VHd		1		1		
	1491	Clostridium botulinum	164	6	308	74	144	68
X	36830	Clostridium botulinum E			1		1	
X	1496	Clostridium difficile			17		17	
	1502	Clostridium perfringens	2		2			
X	107819	Clostridium perfringens D			1		1	
	195102	Clostridium perfringens str. 13		1		1		
	1513	Clostridium tetani		26	76	264	76	238
	199306	Coccidioides posadasii		3		3		
X	9014	Colinus virginianus				2		2
	8932	Columba livia		11		13		2
	5347	Coprinellus congregatus		1		1		
X	13451	Corylus avellana			2	27	2	27
	1717	Corynebacterium diphtheriae	1	1	1	7		6
X	152794	Corynebacterium efficiens				1		1
	1718	Corynebacterium glutamicum		17		17		
	10714	Corynephage omega		2		2		
	10243	Cowpox virus		1		4		3
	265872	Cowpox virus (Brighton Red)		1		1		
	777	Coxiella burnetii		8		8		
	103903	Coxsackievirus B3 (strain Nancy)	1		1			
X	103904	Coxsackievirus B3 (strain Woodruff)				2		2
	103906	Coxsackievirus B4 (strain JVB / Benschoten / New York/51)		73		73		
	10029	Cricetulus griseus		3		3		
	10034	Cricetus cricetus		1		1		
	3369	Cryptomeria japonica	4	2	10	18	6	16
	5807	Cryptosporidium parvum		4		4		
X	3656	Cucumis melo			12		12	
X	84074	Cyclopes didactylus			1		1	
X	28909	Cynodon dactylon			1		1	
	7955	Danio rerio		1		1		

New 2007	NCBI TAX ID	SPECIES/STRAIN	B-06	T-06	B-07	T-07	Δ B	Δ T
X	4039	Daucus carota				1		1
X	243164	Dehalococcoides ethenogenes 195				1		1
	11053	Dengue virus type 1		10	1	18	1	8
X	408685	Dengue virus type 1 Brazil/97-11/1997				2		2
	33741	Dengue virus type 1 Singapore/S275/1990		3	1	3	1	
	11060	Dengue virus type 2	27	40	50	47	23	7
	11062	Dengue virus type 2 (isolate Malaysia M2)	4		4			
	31635	Dengue virus type 2 16681-PDK53		4	1	4	1	
	11064	Dengue virus type 2 Jamaica/1409/1983	4	1	251	11	247	10
	11066	Dengue virus type 2 Puerto Rico/PR159-S1/1969	5		7	2	2	2
	31634	Dengue virus type 2 Thailand/16681/84		50		50		
	11065	Dengue virus type 2 Thailand/NGS-C/1944	7	4	9	4	2	
X	11068	Dengue virus type 2 Thailand/PUO-218/1980			1	1	1	1
	11067	Dengue virus type 2 Tonga/EKB194/1974	6	1	6	1		
	11069	Dengue virus type 3		43		46		3
	11070	Dengue virus type 4		19	5	34	5	15
X	6954	Dermatophagoides farinae			4		4	
	6956	Dermatophagoides pteronyssinus	1	4	11	32	10	28
	7227	Drosophila melanogaster	2	2	2	3		1
X	12639	Duck hepatitis B virus			194	20	194	20
	11021	Eastern equine encephalitis virus	8		8			
X	6210	Echinococcus granulosus			2		2	
	31705	Echovirus 11 (strain Gregory)		1		1		
X	12643	Ectromelia virus				4		4
	5759	Entamoeba histolytica	5		28		23	
	10730	Enterobacteria phage 933W	1		1			
	10710	Enterobacteria phage lambda		7		7		
	10665	Enterobacteria phage T4	5	20	5	20		
X	10326	Equid herpesvirus 1			2		2	
X	12657	Equid herpesvirus 2			5		5	
X	10331	Equid herpesvirus 4			7		7	
	11047	Equine arteritis virus	3		3			
	11665	Equine infectious anemia virus		7	2	46	2	39
	11670	Equine infectious anemia virus (CLONE 1369)		1		1		
	9796	Equus caballus		1	1	56	1	55
	562	Escherichia coli	81	54	85	57	4	3
X	37762	Escherichia coli B			1		1	
	83334	Escherichia coli O157:H7		23		23		
	217992	Escherichia coli O6		6		6		
X	33682	Euglenozoa			15		15	
X	2759	Eukaryota			3	1	3	1
	57482	European bat lyssavirus 1		1		1		
X	420521	Expression vector pNIC-NHT-CF			1		1	
X	3617	Fagopyrum esculentum			28		28	

New 2007	NCBI TAX ID	SPECIES/STRAIN	B-06	T-06	B-07	T-07	Δ B	Δ T
X	46835	Fasciola gigantica			2		2	
X	6192	Fasciola hepatica			12	11	12	11
X	11981	Feline calicivirus (STRAIN F9)			2		2	
X	11980	Feline calicivirus (STRAIN JAPANESE F4)			6		6	
	11673	Feline immunodeficiency virus		2		2		
	11674	Feline immunodeficiency virus (isolate Petaluma)		11		11		
	11675	Feline immunodeficiency virus (isolate San Diego)		9		9		
	31676	Feline immunodeficiency virus (isolate TM2)		1		1		
	45409	Feline immunodeficiency virus (isolate wo)		1		1		
	36371	Feline immunodeficiency virus (strain UK2)		4		4		
	9685	Felis catus	3	3	5	44	2	41
X	4606	Festuca arundinacea			1		1	
X	12110	Foot-and-mouth disease virus			1	1	1	1
X	12112	Foot-and-mouth disease virus (strain A10-61)			1	1	1	1
	12114	Foot-and-mouth disease virus (strain A12)	1		1			
X	73482	Foot-and-mouth disease virus (strain O1)			8	1	8	1
	12118	Foot-and-mouth disease virus - type O	1		3	1	2	1
	31621	Four Corners hantavirus	2		2			
	263	Francisella tularensis		3		9		6
X	177416	Francisella tularensis subsp. tularensis SCHU S4				115		115
	11797	Friend murine leukemia virus (ISOLATE FB29)		1		2		1
X	4751	Fungi			10		10	
X	8053	Gadus callarias			10		10	
	9031	Gallus gallus		49	50	100	50	51
	54290	GB virus C	5		6		1	
	5741	Giardia intestinalis		3		3		
	184922	Giardia lamblia ATCC 50803		41		41		
	3847	Glycine max	17		20	1	3	1
X	9595	Gorilla gorilla gorilla				4		4
	45219	Guanarito virus		9		9		
	727	Haemophilus influenzae		31	21	32	21	1
X	374932	Haemophilus influenzae PittHH			1	1	1	1
X	6454	Haliotis rufescens			1		1	
	11602	Hantaan virus 76-118	9	6	10	17	1	11
	11601	Hantaan virus Lee	1		1			
X	13557	Hapalemur griseus				1		1
X	4232	Helianthus annuus			18		18	
	32025	Helicobacter hepaticus		1		1		
	210	Helicobacter pylori	9		19	3	10	3
	85962	Helicobacter pylori 26695	13		13			
	63330	Hendra virus	6		7		1	
X	10404	Hepadnaviridae			9	16	9	16

New 2007	NCBI TAX ID	SPECIES/STRAIN	B-06	T-06	B-07	T-07	Δ B	Δ T
	12092	Hepatitis A virus	5		11		6	
	12098	Hepatitis A virus (STRAIN HM-175)	94	2	97	2	3	
	12099	Hepatitis A virus (STRAIN LA)	1	10	1	10		
	12093	Hepatitis A virus (STRAIN LCDC-1)	4		4			
	10407	Hepatitis B virus	2	30	237	577	235	547
	10411	Hepatitis B virus (STRAIN ALPHA1)		11		28		17
	10414	Hepatitis B virus (STRAIN LSH / CHIMPANZEE ISOLATE)		7	1	11	1	4
	31512	Hepatitis B virus (subtype ADR / mutant)		3		3		
	10409	Hepatitis B virus (SUBTYPE ADR4)		31	3	66	3	35
	10410	Hepatitis B virus (SUBTYPE ADW / STRAIN 991)		2	6	9	6	7
	10412	Hepatitis B virus (SUBTYPE ADW / STRAIN INDONESIA/PIDW420)		8		10		2
	10413	Hepatitis B virus (SUBTYPE ADW / STRAIN JAPAN/PJDW233)		4		8		4
	31514	Hepatitis B virus (subtype ADW / strain Philippino/PFDW294)	1		2		1	
	45410	Hepatitis b virus (subtype ADW4 / strain brazil / isolate w4b)		15		21		6
X	31511	Hepatitis B virus subtype AD			1	3	1	3
	106820	Hepatitis B virus subtype ADR		2	8	29	8	27
	106821	Hepatitis B virus subtype ADW		1	15	44	15	43
	10415	Hepatitis B virus subtype ADW strain OKINAWA/PODW282		3		5		2
	10408	Hepatitis B virus subtype ADW2		37	76	101	76	64
	31515	Hepatitis B virus subtype ADW2 variant SF		2		2		
	10419	Hepatitis B virus subtype ADYW		5	4	48	4	43
	10418	Hepatitis B virus subtype AYW	1	78	47	274	46	196
	11103	Hepatitis C virus	467	1063	814	1317	347	254
	11104	Hepatitis C virus (isolate 1)	71	281	141	386	70	105
X	356391	Hepatitis C virus (isolate 6a33)				5		5
X	356413	Hepatitis C virus (isolate BEBE1)				10		10
	11105	Hepatitis C virus (isolate BK)	182	121	192	129	10	8
	333284	Hepatitis C virus (isolate Con1)	4	6	5	8	1	2
	356418	Hepatitis C virus (isolate ED43)		1		1		
	356419	Hepatitis C virus (isolate EUH1480)		3		7		4
X	329389	Hepatitis C virus (isolate Glasgow)			2	1	2	1
	11108	Hepatitis C virus (isolate H)	90	89	127	128	37	39
	63746	Hepatitis C virus (isolate H77)	8	101	20	180	12	79
	356410	Hepatitis C virus (isolate HC-G9)	1	3	2	4	1	1
	11111	Hepatitis C virus (isolate HC-J2)		1		1		
X	11112	Hepatitis C virus (isolate HC-J5)			2		2	
	11113	Hepatitis C virus (isolate HC-J6)	2	2	72	2	70	
X	11114	Hepatitis C virus (isolate HC-J7)			3		3	
	11115	Hepatitis C virus (isolate HC-J8)	1	3	57	3	56	
	31642	Hepatitis C virus (isolate HC-JT)		9	13	9	13	
	11110	Hepatitis C virus (isolate HCT18)	1		1	1		1

New 2007	NCBI TAX ID	SPECIES/STRAIN	B-06	T-06	B-07	T-07	Δ B	Δ T
	356416	Hepatitis C virus (isolate HCV-K3a/650)		6	1	17	1	11
	31644	Hepatitis C virus (isolate HCV-KF)	2	2	4	3	2	1
	356386	Hepatitis C virus (isolate India)		1		1		
	11116	Hepatitis C virus (isolate Japanese)	37	68	89	71	52	3
	356411	Hepatitis C virus (isolate JFH-1)		2		2		
	356417	Hepatitis C virus (isolate JK049)		3		5		2
	356415	Hepatitis C virus (isolate NZL1)	1	2	63	4	62	2
	31645	Hepatitis C virus (isolate Taiwan)	13	57	17	82	4	25
	356421	Hepatitis C virus (isolate Th580)		3		6		3
X	357355	Hepatitis C virus (isolate Tr Kj)			1	3	1	3
	356424	Hepatitis C virus (isolate VN004)		2		2		
X	40271	Hepatitis C virus genotype 2			2	1	2	1
X	421877	Hepatitis C virus isolate HC-J1			3		3	
X	421879	Hepatitis C virus isolate HCR6			3		3	
	31647	Hepatitis C virus subtype 1b	1	8	208	55	207	47
X	31649	Hepatitis C virus subtype 2a			1		1	
X	31650	Hepatitis C virus subtype 2b			4		4	
X	356426	Hepatitis C virus subtype 3a			2		2	
	12475	Hepatitis delta virus		7	6	9	6	2
X	10422	Hepatitis delta virus (ISOLATE AMERICAN)			35		35	
	12461	Hepatitis E virus	46		130	27	84	27
	31767	Hepatitis E virus (strain Burma)	160		160			
	31768	Hepatitis E virus (strain Mexico)	29		29			
	28300	Heron hepatitis B virus		1		1		
	10299	Herpes simplex virus (type 1 / strain 17)		1	27	17	27	16
	10301	Herpes simplex virus (type 1 / strain Angelotti)		3	1	4	1	1
X	10304	Herpes simplex virus (type 1 / strain F)			12	7	12	7
X	10303	Herpes simplex virus (type 1 / strain HFEM)			5		5	
X	10308	Herpes simplex virus (type 1 / strain Patton)			3		3	
X	10309	Herpes simplex virus (type 1 / strain SC16)			3		3	
X	10292	Herpesviridae			18	12	18	12
	10383	Herpesvirus saimiri (strain 11)		1		1		
	3981	Hevea brasiliensis	34		48	43	14	43
X	11097	Hog cholera virus (strain Alfort)				1		1
	11583	HoJo virus		1		1		
X	29679	Holcus lanatus			6		6	
	9606	Homo sapiens	74	2389	306	2804	232	415
	10515	Human adenovirus 2		2	2	5	2	3
	28285	Human adenovirus 5	3	2	3	3		1
X	28282	Human adenovirus type 12				1		1
	28284	Human adenovirus type 40		1		1		
	10519	Human adenovirus type 7		1		1		
	11137	Human coronavirus 229E		2		2		
X	39054	Human enterovirus 71			2		2	
X	10298	Human herpesvirus 1			187	101	187	101
	10306	Human herpesvirus 1 strain KOS		2	3	5	3	3

New 2007	NCBI TAX ID	SPECIES/STRAIN	B-06	T-06	B-07	T-07	Δ B	Δ T
	10310	Human herpesvirus 2		4	38	38	38	34
X	10313	Human herpesvirus 2 strain 333				25		25
	10315	Human herpesvirus 2 strain HG52		12		12		
X	10335	Human herpesvirus 3			32	70	32	70
X	10338	Human herpesvirus 3 (strain Dumas)			1	4	1	4
	10376	Human herpesvirus 4		11	42	536	42	525
	10377	Human herpesvirus 4 (strain B95-8)	1	92	47	400	46	308
X	31525	Human herpesvirus 4 (strain CAO)				1		1
X	36352	Human herpesvirus 4 type 1			1	12	1	12
X	12509	Human herpesvirus 4 type 2			1	1	1	1
X	10359	Human herpesvirus 5			26	404	26	404
	10360	Human herpesvirus 5 strain AD169		49	67	409	67	360
	10363	Human herpesvirus 5 strain Towne		5	1	15	1	10
	10368	Human herpesvirus 6	2		2			
X	10369	Human herpesvirus 6 (strain GS)			1		1	
	10370	Human herpesvirus 6 (strain Uganda-1102)	1	1	1	1		
	32604	Human herpesvirus 6B	3	2	3	2		
X	10372	Human herpesvirus 7			2		2	
	57278	Human herpesvirus 7 strain JI		1		1		
	37296	Human herpesvirus 8	27	133	28	150	1	17
	11676	Human immunodeficiency virus 1		32	2	83	2	51
	11709	Human immunodeficiency virus 2		3		4		1
	11685	Human immunodeficiency virus type 1 (ARV2/SF2 ISOLATE)		18		53		35
	11678	Human immunodeficiency virus type 1 (BH10 ISOLATE)		21		33		12
	11684	Human immunodeficiency virus type 1 (BH8 ISOLATE)		3		3		
	11693	Human immunodeficiency virus type 1 (BRAIN ISOLATE)		4		4		
	11686	Human immunodeficiency virus type 1 (BRU ISOLATE)		6	1	7	1	1
	11687	Human immunodeficiency virus type 1 (CDC-451 ISOLATE)		1		3		2
	11679	Human immunodeficiency virus type 1 (CLONE 12)		8		11		3
	11689	Human immunodeficiency virus type 1 (ELI ISOLATE)		10		13		3
	11706	Human immunodeficiency virus type 1 (HXB2 ISOLATE)		55		59		4
	11707	Human immunodeficiency virus type 1 (HXB3 ISOLATE)		3		3		
	362651	Human immunodeficiency virus type 1 (isolate YU2)		83		86		3
	11694	Human immunodeficiency virus type 1 (JH3 ISOLATE)		1		3		2
	11688	Human immunodeficiency virus type 1 (JRCSF ISOLATE)		33		35		2
	36375	Human immunodeficiency virus type 1 (KB-1 isolate)		2		2		

New 2007	NCBI TAX ID	SPECIES/STRAIN	B-06	T-06	B-07	T-07	Δ B	Δ T
	11697	Human immunodeficiency virus type 1 (MAL ISOLATE)		20		23		3
	11704	Human immunodeficiency virus type 1 (MFA ISOLATE)		3		4		1
	11696	Human immunodeficiency virus type 1 (MN ISOLATE)	1	27	1	34		7
	11695	Human immunodeficiency virus type 1 (NDK ISOLATE)		9		10		1
	11698	Human immunodeficiency virus type 1 (NEW YORK-5 ISOLATE)		3		4		1
X	11699	Human immunodeficiency virus type 1 (OYI ISOLATE)				4		4
	11701	Human immunodeficiency virus type 1 (RF/HAT ISOLATE)		31		35		4
	11691	Human immunodeficiency virus type 1 (SF162 ISOLATE)		1		1		
	11690	Human immunodeficiency virus type 1 (SF33 ISOLATE)		2		2		
	11703	Human immunodeficiency virus type 1 (STRAIN UGANDAN / ISOLATE U455)		7		15		8
	31678	Human immunodeficiency virus type 1 (WMJ1 isolate)		15		15		
	11705	Human immunodeficiency virus type 1 (WMJ2 ISOLATE)		2		2		
	11683	Human immunodeficiency virus type 1 (Z2/CDC-Z34 ISOLATE)		5		5		
	11708	Human immunodeficiency virus type 1 (ZAIRE 6 ISOLATE)		3		3		
	82834	Human immunodeficiency virus type 1 lw12.3 isolate		17		17		
X	11714	Human immunodeficiency virus type 2 (ISOLATE BEN)				7		7
X	11715	Human immunodeficiency virus type 2 (ISOLATE CAM2)				1		1
X	11713	Human immunodeficiency virus type 2 (ISOLATE D194)				1		1
	11716	Human immunodeficiency virus type 2 (ISOLATE D205,7)		1		1		
	11717	Human immunodeficiency virus type 2 (ISOLATE GHANA-1)		1		6		5
X	73484	Human immunodeficiency virus type 2 (isolate KR)				2		2
	11719	Human immunodeficiency virus type 2 (ISOLATE NIH-Z)		1		1		
X	11718	Human immunodeficiency virus type 2 (ISOLATE SBLISY)				2		2
	11721	Human immunodeficiency virus type 2 (ISOLATE ST)		1		1		
	10580	Human papillomavirus type 11		3		6		3
	333760	Human papillomavirus type 16	1	92	1	114		22
	333761	Human papillomavirus type 18		4		16		12
X	10583	Human papillomavirus type 1a				1		1

New 2007	NCBI TAX ID	SPECIES/STRAIN	B-06	T-06	B-07	T-07	Δ B	Δ T
	37112	Human papillomavirus type 29		1		1		
	10585	Human papillomavirus type 31		1		11		10
X	10586	Human papillomavirus type 33				6		6
	10592	Human papillomavirus type 44		1		1		
X	10593	Human papillomavirus type 45				2		2
X	10618	Human papillomavirus type 52				7		7
X	333765	Human papillomavirus type 53				1		1
	10596	Human papillomavirus type 56		1		5		4
X	10598	Human papillomavirus type 58				4		4
	37115	Human papillomavirus type 59		1		1		
	10600	Human papillomavirus type 6b		5		5		
	10602	Human papillomavirus type me180		1		1		
X	36412	Human parainfluenza 1 virus (strains A1426 / 86-315 / 62M-753)				2		2
	10798	Human parvovirus B19	79	19	95	43	16	24
	12081	Human poliovirus 1 Mahoney	1		1			
	12082	Human poliovirus 1 strain Sabin	1		1			
X	11250	Human respiratory syncytial virus			1	1	1	1
X	11251	Human respiratory syncytial virus (subgroup B / strain 18537)			1		1	
X	11260	Human respiratory syncytial virus A strain Long			10	1	10	1
	11259	Human respiratory syncytial virus A2		16	16	31	16	15
	10962	Human rotavirus serotype 1 / strain WA	1		2	5	1	5
	10960	Human rotavirus strain St. Thomas 3		2		2		
	11927	Human T-cell lymphotrophic virus type 1 (Caribbean isolate)		12		13		1
	11926	Human T-cell lymphotrophic virus type 1 (strain ATK)	1		10	8	9	8
X	39015	Human T-cell lymphotropic virus type 1 (african isolate)				1		1
	11908	Human T-lymphotropic virus 1		2	18	52	18	50
X	11909	Human T-lymphotropic virus 2			17	1	17	1
	9580	Hylobates lar		4		4		
	77644	IncQ plasmid pIE1120		1		1		
X	11290	Infectious hematopoietic necrosis virus			9		9	
	11320	Influenza A virus	25	68	35	103	10	35
	387139	Influenza A virus (A/Aichi/2/1968(H3N2))	27	27	36	42	9	15
	385576	Influenza A virus (A/Alaska/6/1977(H3N2))		1		1		
	383602	Influenza A virus (A/Anas acuta/Primorje/695/1976(H2N3))		4		15		11
	135322	Influenza A virus (A/Ann Arbor/6/60(H2N2))		166		194		28
	62446	Influenza A virus (A/Argentina/3779/94(H3N2))		1		1		
	385630	Influenza A virus (A/Bangkok/1/1979(H3N2))	1	32	1	37		5
	11327	Influenza A virus (A/Beijing/11/56 (H1N1))		22		22		
	62449	Influenza A virus (A/Beijing/281/94(H3N2))	1		1			
	62450	Influenza A virus (A/Beijing/32/92(H3N2))		24		24		

New 2007	NCBI TAX ID	SPECIES/STRAIN	B-06	T-06	B-07	T-07	Δ B	Δ T
	385587	Influenza A virus (A/budgerigar/Hokkaido/1/1977(H4N6))		5		5		
X	11384	Influenza A virus (A/chicken/FPV/Weybridge(H7N7))				1		1
X	97348	Influenza A virus (A/Chicken/Hong Kong/G23/97(H9N2))				1		1
X	298603	Influenza A virus (A/chicken/Hubei/489/2004(H5N1))				1		1
X	285712	Influenza A virus (A/chicken/Jilin/9/2004(H5N1))				1		1
X	300750	Influenza A virus (A/chicken/Korea/S1/2003(H9N2))				1		1
	385617	Influenza A virus (A/chicken/Pennsylvania/1370/1983(H5N2))	1		1			
	380985	Influenza A virus (A/Chile/1/1983(H1N1))		1		1		
	62541	Influenza A virus (A/Christ_Church/2/88(H3N2))		1		1		
	107493	Influenza A virus (A/Cordoba/3278/96(H3N2))		1		1		
X	88295	Influenza A virus (A/duck/Alberta/35/76(H1N1))				3		3
X	383550	Influenza A virus (A/duck/England/1/1956(H11N6))				1		1
X	365080	Influenza A virus (A/duck/Guangxi/1793/2004(H5N1))				1		1
	80266	Influenza A virus (A/Duck/Hokkaido/8/80 (H3N8))		5		6		1
X	210671	Influenza A virus (A/duck/Hong Kong/366/78(H9N2))				1		1
	353253	Influenza A virus (A/duck/Novosibirsk/56/2005(H5N1))		2		2		
	385580	Influenza A virus (A/duck/Ukraine/1/1963(H3N8))	3		3			
	221011	Influenza A virus (A/England/878/69(H3N2))		1	1	3	1	2
	137578	Influenza A virus (A/England/939/69 x A/PR/8/34)		1		1		
	229411	Influenza A virus (A/Fort Monmouth/1/47-MA(H1N1))		8		8		
	260806	Influenza A virus (A/FPV/Dutch/27(H7N7))	1		1			
	382786	Influenza A virus (A/FPV/Rostock/1934(H7N1))	2	5	2	5		
	107558	Influenza A virus (A/France/75/97(H3N2))		1		1		
	31661	Influenza A virus (A/Harbin/1/88(H1N2))		3		5		2
	108859	Influenza A virus (A/Hong Kong/1/68(H3N2))		10		13		3
	130760	Influenza A virus (A/Hong Kong/1073/99(H9N2))	1		1			
	130763	Influenza A virus (A/Hong Kong/156/97(H5N1))		1		2		1
	317652	Influenza A virus (A/Hong Kong/2/68(H3N2))		5		6		1

New 2007	NCBI TAX ID	SPECIES/STRAIN	B-06	T-06	B-07	T-07	Δ B	Δ T
	88104	Influenza A virus (A/Hong Kong/483/1997(H5N1))	1		1			
X	153969	Influenza A virus (A/Hong Kong/497/97(H3N2))				1		1
	220500	Influenza A virus (A/Japan/305/57(H2N2))	2	49	3	64	1	15
X	11422	Influenza A virus (A/Kiev/59/79(H1N1))				2		2
	220985	Influenza A virus (A/Korea/426/68(H2N2))		1		1		
	393557	Influenza A virus (A/Leningrad/1954/1(H1N1))		2		6		4
	62559	Influenza A virus (A/Los_Angeles/(H3N2))	5		5			
	95888	Influenza A virus (A/mallard duck/PA/10218/84(H5N2))	5		5			
	220503	Influenza A virus (A/Memphis/1/71(H3N2))	3	6	18	12	15	6
	383583	Influenza A virus (A/Memphis/101/1972(H3N2))		1		1		
	252233	Influenza A virus (A/Memphis/102/72(H3N2))		5		5		
	228928	Influenza A virus (A/Memphis/31/98(H3N2))	5		6		1	
X	38973	Influenza A virus (A/Memphis/4/1973(H3N2))				2		2
	11440	Influenza A virus (A/Memphis/6/86(H3N2))	3	17	3	21		4
	311374	Influenza A virus (A/Mongolia/111/91 (H1N1))		2		2		
	62488	Influenza A virus (A/Nanchang/58/1993(H3N2))		1		1		
	132841	Influenza A virus (A/Netherlands/785e/90 (H3N2))		1		1		
	62564	Influenza A virus (A/New_York/15/94(H3N2))		1		1		
	62496	Influenza A virus (A/New_York/17/94(H3N2))		1		1		
	370128	Influenza A virus (A/Northern Territory/60/1968(H3N2))		4		4		
	260805	Influenza A virus (A/NT/60/68/(H3N2))	1	95	1	101		6
	62503	Influenza A virus (A/Ohio/3/95(H3N2))		1		1		
X	223935	Influenza A virus (A/Okuda/57(H2N2))			1		1	
X	11448	Influenza A virus (A/parrot/Ulster/73(H7N1))				2		2
	119209	Influenza A virus (A/Philippines/2/82(H3N2))	1		1			
	385624	Influenza A virus (A/Port Chalmers/1/1973(H3N2))	7		7			
	211044	Influenza A virus (A/Puerto Rico/8/34(H1N1))	14	398	22	766	8	368
	183764	Influenza A virus (A/Puerto Rico/8/34/Mount Sinai(H1N1))	6	87	6	124		37
X	106423	Influenza A virus (A/Quail/Hong Kong/AF157/92(H9N2))				1		1
	221016	Influenza A virus (A/Queensland/7/70(H3N2))		2		2		
	221012	Influenza A virus (A/Rio/6/69(H3N2))		1		2		1
X	384493	Influenza A virus			1		1	

New 2007	NCBI TAX ID	SPECIES/STRAIN	B-06	T-06	B-07	T-07	Δ B	Δ T
		(A/seal/Mass/1/1980(H7N7))						
	62512	Influenza A virus (A/Shangdong/5/94(H3N2))		1		1		
	63105	Influenza A virus (A/Shanghai/16/89(H3N2))		2		3		1
X	62412	Influenza A virus (A/Shanghai/6/90(H3N2))				3		3
	220949	Influenza A virus (A/Singapore/1/57(H2N2))	1		1			
	150154	Influenza A virus (A/swine/29/37 (H1N1))		2		2		
X	169169	Influenza A virus (A/swine/Cotes d'Armor/1482/99(H1N1))				1		1
X	253676	Influenza A virus (A/swine/Hong Kong/1197/02(H3N2))				1		1
	11498	Influenza A virus (A/swine/Hong Kong/126/82(H3N2))		1		1		
X	219643	Influenza A virus (A/swine/Hong Kong/127/82(H3N2))				1		1
	219637	Influenza A virus (A/swine/Hong Kong/6/76(H3N2))		2		2		
	219641	Influenza A virus (A/swine/Hong Kong/81/78(H3N2))		1		9		8
X	145307	Influenza A virus (A/swine/Hong Kong/9/98(H9N2))			2		2	
X	11504	Influenza A virus (A/swine/Indiana/1726/88(H1N1))				1		1
	300743	Influenza A virus (A/swine/Korea/S10/2004(H1N1))		1		1		
	300744	Influenza A virus (A/swine/Korea/S175/2004(H1N1))		1		1		
	186460	Influenza A virus (A/swine/New Jersey/11/76(H1N1))	1		1			
	88305	Influenza A virus (A/swine/Wisconsin/1/61(H1N1))		2		2		
	82372	Influenza A virus (A/Sydney/05/97-like(H3N2))		1		1		
	11465	Influenza A virus (A/Taiwan/1/86(H1N1))	1	2	1	2		
	384509	Influenza A virus (A/tern/Australia/G70C/1975(H11N9))	3		10		7	
	183796	Influenza A virus (A/Texas/1/77(H3N2))	2	44	2	49		5
	380301	Influenza A virus (A/turkey/Ontario/7732/1966(H5N9))	5	6	5	6		
X	381517	Influenza A virus (A/Udorn/307/1972(H3N2))				5		5
	62596	Influenza A virus (A/USSR/26/(H3N2))		1		1		
	381516	Influenza A virus (A/USSR/90/1977(H1N1))	12	14	17	18	5	4
	392809	Influenza A virus (A/Victoria/3/1975(H3N2))	44	8	48	13	4	5
	284217	Influenza A virus (A/Viet Nam/1194/2004(H5N1))		3		3		
	11484	Influenza A virus (A/whale/Maine/1/84(H13N9))	1		3		2	
	381518	Influenza A virus (A/Wilson-Smith/1933(H1N1))		108		110		2

New 2007	NCBI TAX ID	SPECIES/STRAIN	B-06	T-06	B-07	T-07	Δ B	Δ T
	382835	Influenza A virus (A/WSN/1933(H1N1))	8		8	1		1
	63106	Influenza A virus (A/Wuhan/359/95(H3N2))	3		3			
	132504	Influenza A virus (A/X-31(H3N2))	13	130	25	146	12	16
	11357	Influenza A virus (STRAIN A/DUCK/HOKKAIDO/5/77)		1		1		
	11408	Influenza A virus (STRAIN A/EQUINE/NEW MARKET/76)		1		1		
	11482	Influenza A virus (STRAIN A/VICTORIA/5/68)		2		4		2
	41857	Influenza A virus H3N2		2		7		5
	11520	Influenza B virus	1	5	1	7		2
	11521	Influenza B virus (B/Ann Arbor/1/1986)		3		3		
X	184816	Influenza B virus (B/Kadoma/122/99)			3		3	
	256080	Influenza B virus (B/Kobe/1/2003)	1		1			
	107412	Influenza B virus (B/Lee/40)	5	5	5	5		
	11541	Influenza B virus (B/Oregon/5/80)	15		17		2	
	150127	Influenza B virus (B/Osaka/983/97-V3)	1		1			
	38994	Influenza B virus (strain B/finland/150/90)	2		2			
	11531	Influenza B virus (STRAIN B/HONG KONG/8/73)	1	1	1	1		
	11553	Influenza C virus (C/Ann Arbor/1/50)	9		9			
	9725	Inia geoffrensis		1		1		
	42097	Isla Vista virus	1		1			
	11072	Japanese encephalitis virus	3		3			
	11075	Japanese encephalitis virus strain JAOARS982	11	10	12	10	1	
	11076	Japanese encephalitis virus strain Nakayama		1	1	2	1	1
	51240	Juglans regia	1		1			
	11619	Junin virus		1		1		
	13101	Juniperus ashei	5		5			
	42894	Khabarovsk virus	1		1			
	573	Klebsiella pneumoniae	15		15			
	11078	Kunjin virus (STRAIN MRM61C)		4		5		1
	11577	La Crosse virus		1		1		
	33727	Lake Victoria marburgvirus - Musoke	4		4			
	33728	Lake Victoria marburgvirus - Popp	1		1			
	11620	Lassa virus		5		5		
	11621	Lassa virus GA391		6		12		6
	11622	Lassa virus Josiah	4	63	4	115		52
X	5658	Leishmania				4		4
X	5667	Leishmania aethiopica			10		10	
X	5659	Leishmania amazonensis				1		1
X	5660	Leishmania braziliensis			6		6	
X	5661	Leishmania donovani			5	33	5	33
X	44271	Leishmania donovani chagasi			2		2	
X	5671	Leishmania infantum			123	1	123	1
X	5664	Leishmania major			5	107	5	107

New 2007	NCBI TAX ID	SPECIES/STRAIN	B-06	T-06	B-07	T-07	Δ B	Δ T
X	347515	Leishmania major strain Friedlin				1		1
X	5679	Leishmania panamensis			3	8	3	8
X	5682	Leishmania pifanoi				20		20
X	11646	Lentivirus				1		1
X	36936	Lepidoglyphus destructor			5		5	
	1642	Listeria innocua	2	1	2	1		
	1639	Listeria monocytogenes	23	181	24	194	1	13
	265669	Listeria monocytogenes str. 4b F2365		3		5		2
	1643	Listeria welshimeri		33		33		
	217686	Little cherry virus 1		1		1		
	4522	Lolium perenne	7	3	7	32		29
X	36386	Louping ill virus (strain 31)			1		1	
	11623	Lymphocytic choriomeningitis virus	5	57	5	82		25
	11624	Lymphocytic choriomeningitis virus (strain Armstrong)	2	198	2	309		111
	11627	Lymphocytic choriomeningitis virus (strain WE)		57		68		11
	9541	Macaca fascicularis		2		3		1
X	9544	Macaca mulatta				1		1
X	10373	Macaca mulatta cytomegalovirus				35		35
X	3750	Malus x domestica			2	1	2	1
X	40674	Mammalia			4	4	4	4
	7130	Manduca sexta		5		6		1
	75985	Mannheimia haemolytica	14		14			
X	11234	Measles virus			1	34	1	34
	36408	Measles virus strain AIK-C		1		1		
	11235	Measles virus strain Edmonston		10	90	109	90	99
	11237	Measles virus strain IP-3-CA		6		9		3
	11239	Measles virus strain Yamagata-1		2	1	5	1	3
	12107	Mengo virus	1		1			
	10036	Mesocricetus auratus	23		29		6	
	11801	Moloney murine leukemia virus		1		1		
	31691	Moloney murine sarcoma virus (strain TS110)		2		2		
X	10244	Monkeypox virus			5		5	
	300180	Mopeia Lassa reassortant 29		1		1		
	11629	Mopeia virus		3		3		
X	11161	Mumps virus			5		5	
X	11171	Mumps virus (STRAIN MIYAHARA VACCINE)			2		2	
	10366	Murid herpesvirus 1	2		2	54		54
	33708	Murid herpesvirus 4		7		27		20
X	69156	Murine cytomegalovirus (strain K181)				25		25
	10367	Murine cytomegalovirus (strain Smith)		5		31		26
	12760	Murine hepatitis virus strain 4		3		3		
	11142	Murine hepatitis virus strain A59	1	1	1	1		
	11144	Murine hepatitis virus strain JHM		5	2	6	2	1
X	338561	Murine herpesvirus strain 72				1		1

New 2007	NCBI TAX ID	SPECIES/STRAIN	B-06	T-06	B-07	T-07	Δ B	Δ T
X	11812	Murine sarcoma virus 3611			4		4	
	301478	Murray valley encephalitis virus (strain MVE-1-51)	24	21	24	22		1
	10090	Mus musculus	71	210	109	275	38	65
X	4641	Musa acuminata				1		1
	1763	Mycobacterium	4	5	38	27	34	22
	1764	Mycobacterium avium		1	2	8	2	7
X	262316	Mycobacterium avium subsp. paratuberculosis K-10				1		1
	1765	Mycobacterium bovis		32	54	184	54	152
	233413	Mycobacterium bovis AF2122/97	7	10	7	90		80
X	410289	Mycobacterium bovis BCG str. Pasteur 1173P2			1	68	1	68
	1767	Mycobacterium intracellulare	1		1	1		1
	1768	Mycobacterium kansasii	7		7			
	1769	Mycobacterium leprae		160	107	356	107	196
X	272631	Mycobacterium leprae TN				35		35
X	1783	Mycobacterium scrofulaceum			4		4	
X	197612	Mycobacterium sp. 185-409				1		1
	1773	Mycobacterium tuberculosis	83	240	188	955	105	715
X	348776	Mycobacterium tuberculosis C				12		12
	83331	Mycobacterium tuberculosis CDC1551		13		18		5
X	336982	Mycobacterium tuberculosis F11				1		1
	83332	Mycobacterium tuberculosis H37Rv	1	147	20	284	19	137
X	28903	Mycoplasma bovis			21		21	
	2097	Mycoplasma genitalium		1		1		
X	2104	Mycoplasma pneumoniae			43	2	43	2
X	272634	Mycoplasma pneumoniae M129			2		2	
X	59463	Myotis lucifugus			1		1	
X	8656	Naja atra			1		1	
X	35670	Naja naja			1		1	
X	8658	Naja pallida			1	1	1	1
	485	Neisseria gonorrhoeae		9		9		
	122586	Neisseria meningitidis MC58		1		1		
	4097	Nicotiana tabacum	1		1			
	121791	Nipah virus	3		4		1	
	122928	Norovirus genogroup 1	1		1			
	122929	Norovirus genogroup 2	1		1			
	95340	Norwalk-like virus	1		1			
X	8996	Numida meleagris			1		1	
X	4146	Olea europaea			1	14	1	14
	42764	Oliveros virus		1		1		
X	9733	Orcinus orca				1		1
	784	Orientia tsutsugamushi	89		89			
	9986	Oryctolagus cuniculus		5		6		1
	9940	Ovis aries	54	3	88	3	34	
	1406	Paenibacillus polymyxa		1		1		

New 2007	NCBI TAX ID	SPECIES/STRAIN	B-06	T-06	B-07	T-07	Δ B	Δ T
	9598	Pan troglodytes	8	10	8	10		
X	121759	Paracoccidioides brasiliensis				8		8
X	33127	Parietaria judaica			4	2	4	2
	10780	Parvoviridae	1		2	1	1	1
	747	Pasteurella multocida		8	2	8	2	
X	6687	Penaeus monodon			2		2	
	5076	Penicillium chrysogenum	10		10			
	6978	Periplaneta americana	3		3			
X	31604	Peste-des-petits-ruminants virus			9	1	9	1
X	3885	Phaseolus vulgaris				1		1
	15957	Phleum pratense	21		26	26	5	26
	9755	Physeter catodon		30	1	31	1	1
	11630	Pichinde virus		10		10		
	3888	Pisum sativum		2		2		
	141833	Plasmid pIPO2T		2		2		
	5821	Plasmodium berghei		1		1		
	5833	Plasmodium falciparum		1		2		1
	57270	Plasmodium falciparum Palo Alto/Uganda		5		5		
	73239	Plasmodium yoelii yoelii		2		2		
X	352914	Plasmodium yoelii yoelii str. 17XNL			4	2	4	2
X	11245	Pneumovirus			2		2	
X	4545	Poa pratensis			21	17	21	17
X	4479	Poaceae			9	2	9	2
	5145	Podospora anserina		1		1		
X	138953	Poliovirus			3	3	3	3
	12088	Poliovirus type 3 (strains P3/LEON/37 AND P3/LEON 12A[1]B)	1		1			
X	188763	Pongine herpesvirus 4				2		2
	9600	Pongo pygmaeus		2		2		
	28344	Porcine respiratory and reproductive syndrome virus	1		13		12	
X	53179	Porcine rubulavirus			3		3	
	11151	Porcine transmissible gastroenteritis coronavirus strain Purdue	1		1			
	242619	Porphyromonas gingivalis W83		2		2		
	11603	Prospect Hill virus	1		1			
	584	Proteus mirabilis	1		1			
X	42229	Prunus avium				1		1
X	3760	Prunus persica			11		11	
	287	Pseudomonas aeruginosa	23	34	37	35	14	1
	208964	Pseudomonas aeruginosa PAO1	1		1			
	294	Pseudomonas fluorescens		3		3		
	303	Pseudomonas putida		5		5		
	306	Pseudomonas sp.		1		1		
	74138	Pseudomonas sp. DJ-12		1		1		
	71238	Pseudomonas sp. G-179		1		1		
	159091	Pseudomonas sp. KIE171		1		1		

New 2007	NCBI TAX ID	SPECIES/STRAIN	B-06	T-06	B-07	T-07	Δ B	Δ T
	237609	Pseudomonas sp. KL28		1		1		
	150396	Pseudomonas sp. MT-1		1		1		
	11604	Puumala virus	10	1	10	4		3
X	39000	Puumala virus (strain k27)				6		6
	39002	Puumala virus (strain sotkamo/v-2969/81)	243	1	244	1	1	
	39003	Puumala virus (strain udmurtia/894cg/91)	1		1			
	314536	Rabbit hemorrhagic disease virus-FRG	1		1			
	11292	Rabies virus	9	11	9	11		
	11293	Rabies virus (strain AVO1)	3	9	3	10		1
	11295	Rabies virus (strain ERA)	10	12	16	14	6	2
	11296	Rabies virus (strain HEP-FLURY)	3		3			
	39005	Rabies virus (strain ontario skunk)		1		1		
	103929	Rabies virus (strain Pasteur / PV)	3	2	15	4	12	2
	11300	Rabies virus (strain SAD B19)		1		1		
	45418	Rabies virus (strain vnukovo-32)		1		1		
	10116	Rattus norvegicus	8	77	8	77		
X	12814	Respiratory syncytial virus				1		1
	186539	Reston ebolavirus	2		2			
X	103930	Rhesus cytomegalovirus strain 68-1				5		5
X	60189	Rhipicephalus decoloratus			4		4	
X	43767	Rhodococcus equi			15		15	
	3988	Ricinus communis	8	4	8	4		
	781	Rickettsia conorii		11		11		
	782	Rickettsia prowazekii		10		10		
	783	Rickettsia rickettsii		1		1		
	35793	Rickettsia sibirica		3		3		
	785	Rickettsia typhi		1		1		
	11588	Rift Valley fever virus	4		6		2	
	11589	Rift valley fever virus (STRAIN ZH-548 M12)		1		1		
	11243	Rinderpest virus (strain L)	4		4			
	46920	Rio Mamore virus	1		1			
	37207	Rio Segundo virus	1		1			
	1	root	39	504	79	630	40	126
	11032	Ross river virus (STRAIN T48)	3		3			
X	11041	Rubella virus			33	42	33	42
	11043	Rubella virus (strain M33)	1		1			
	11045	Rubella virus (strain THERIEN)		1	1	1	1	
	45709	Sabia virus		11		11		
	4932	Saccharomyces cerevisiae		3	1	4	1	1
	9521	Saimiri sciureus		1		1		
X	596	Salmonella enterica subsp. enterica serovar Muenchen			1		1	
	601	Salmonella typhi	4	22	4	22		
	602	Salmonella typhimurium		36	1	40	1	4
	227859	SARS coronavirus	156	1916	193	1972	37	56
	299335	SARS coronavirus B039		13		13		

New 2007	NCBI TAX ID	SPECIES/STRAIN	B-06	T-06	B-07	T-07	Δ B	Δ T
X	228407	SARS coronavirus BJ01			2		2	
	285949	SARS coronavirus civet020		2		2		
	260550	SARS coronavirus CUHK-L2	1		1			
X	229992	SARS coronavirus Frankfurt 1			5		5	
	227984	SARS coronavirus Tor2		1		1		
	228330	SARS coronavirus Urbani	21	42	21	44		2
	344702	SARS coronavirus ZJ0301		59		59		
X	6182	Schistosoma japonicum				9		9
	6183	Schistosoma mansoni		1	3	28	3	27
	4896	Schizosaccharomyces pombe		1		1		
	11033	Semliki forest virus	4		25		21	
X	11191	Sendai virus				4		4
	11194	Sendai virus (strain Enders)		8		10		2
	11195	Sendai virus (strain Fushimi)		8		8		
X	302272	Sendai virus (strain Ohita)				1		1
	11198	Sendai virus (Z)		2		2		
	11610	Seoul virus SR11	1		1			
	615	Serratia marcescens		1		1		
	4182	Sesamum indicum	11		11			
	623	Shigella flexneri	5	4	70	4	65	
	624	Shigella sonnei		1		1		
	11723	Simian immunodeficiency virus		395		447		52
X	11737	Simian immunodeficiency virus (F236/SMH4 ISOLATE) (SOOTY MANGABEY)				7		7
	11735	Simian immunodeficiency virus (K6W ISOLATE)		46		80		34
	11736	Simian immunodeficiency virus (K78 ISOLATE)		6		13		7
X	11733	Simian immunodeficiency virus (MM142-83 ISOLATE)				53		53
X	11734	Simian immunodeficiency virus (MM251 ISOLATE)				1		1
	11711	Simian immunodeficiency virus - mac		1		1		
X	31682	Simian immunodeficiency virus - mac1A11				9		9
	31683	Simian immunodeficiency virus - stm		5		5		
X	160753	Simian immunodeficiency virus 17E-Fr				2		2
	10633	Simian virus 40		12		17		5
X	57667	Simian-Human immunodeficiency virus				1		1
	37705	Sin Nombre virus	2	11	2	11		
X	11780	Snyder-Theilen feline sarcoma virus			1		1	
X	2133	Spiroplasma citri				1		1
	6584	Spisula solidissima		7		7		
	7108	Spodoptera frugiperda		1		1		
X	11081	St. Louis encephalitis virus (strain MS1-7)			10		10	
	1280	Staphylococcus aureus	76	27	79	27	3	
	93062	Staphylococcus aureus subsp. aureus COL	1		1			
	359787	Staphylococcus aureus subsp. aureus JH1	1		1			

New 2007	NCBI TAX ID	SPECIES/STRAIN	B-06	T-06	B-07	T-07	Δ B	Δ T
	282458	Staphylococcus aureus subsp. aureus MRSA252	40		40			
	158878	Staphylococcus aureus subsp. aureus Mu50	1		1			
	196620	Staphylococcus aureus subsp. aureus MW2		7		7		
	158879	Staphylococcus aureus subsp. aureus N315	15		15			
X	1317	Streptococcus downei			4	4	4	4
X	119602	Streptococcus dysgalactiae subsp. equisimilis			1		1	
	1309	Streptococcus mutans	1		4	1	3	1
	1313	Streptococcus pneumoniae		2		2		
	1314	Streptococcus pyogenes	59	75	124	94	65	19
	160490	Streptococcus pyogenes M1 GAS	15		15			
	286636	Streptococcus pyogenes MGAS10394		1		1		
	370554	Streptococcus pyogenes MGAS10750	2		2			
	370553	Streptococcus pyogenes MGAS2096		1		1		
	301449	Streptococcus pyogenes serotype M5	55	99	62	99	7	
	301450	Streptococcus pyogenes serotype M6	4	4	7	4	3	
	1915	Streptomyces lincolnensis		1		1		
	128949	Sudan ebolavirus - Maleo (1979)		1		1		
X	10345	Suid herpesvirus 1			9	2	9	2
	31523	Suid herpesvirus 1 (strain Indiana-Funkhauser / Becker)		1		1		
	9823	Sus scrofa		6	2	6	2	
	32630	synthetic construct		1	1	7	1	6
	11631	Tacaribe virus		1		1		
X	6202	Taenia				1		1
	13281	Taphozous georgianus		2		2		
	204711	Theilovirus		8		8		
X	271	Thermus aquaticus			1		1	
	11084	Tick-borne encephalitis virus		1	3	1	3	
X	11087	Tick-borne encephalitis virus (STRAIN SOFJIN)			18		18	
	83192	Topografov virus	1		1			
X	7787	Torpedo californica			4		4	
	7788	Torpedo marmorata		3		3		
	5811	Toxoplasma gondii	20	28	20	28		
X	11149	Transmissible gastroenteritis virus			1		1	
X	160	Treponema pallidum			4		4	
	9337	Trichosurus vulpecula		3		3		
	4565	Triticum aestivum		75	12	164	12	89
X	279889	Triticum aestivum var. arduini				4		4
X	5690	Trypanosoma			1	29	1	29
X	5691	Trypanosoma brucei			7		7	
	5693	Trypanosoma cruzi	1	61	114	316	113	255
X	353153	Trypanosoma cruzi strain CL Brener			5	106	5	106

New 2007	NCBI TAX ID	SPECIES/STRAIN	B-06	T-06	B-07	T-07	Δ B	Δ T
	37133	Tula virus	1		1			
	10245	Vaccinia virus		119		126		7
	126794	Vaccinia virus Ankara		8		12		4
	10249	Vaccinia virus Copenhagen		132		138		6
	10253	Vaccinia virus Tian Tan		1		1		
	10254	Vaccinia virus WR	7	2991	8	3252	1	261
	12870	Variola major virus		2		2		
	10255	Variola virus		2		5		3
	11037	Venezuelan equine encephalitis virus (strain TC-83)	21		21			
	11038	Venezuelan equine encephalitis virus (strain Trinidad donkey)	7		14		7	
	7742	Vertebrata	7	6	52	7	45	1
X	11276	Vesicular stomatitis virus				2		2
X	11278	Vesicular stomatitis virus (serotype Indiana / strain Glasgow)				2		2
	11283	Vesicular stomatitis virus (serotype New Jersey / strain Ogden)		1		1		
	11284	Vesicular stomatitis virus (strain Orsay)		1		1		
	11285	Vesicular stomatitis virus (strain San Juan)		2		5		3
X	7454	Vespula vulgaris				36		36
	666	Vibrio cholerae	65	34	79	34	14	
X	345076	Vibrio cholerae V52			5		5	
	670	Vibrio parahaemolyticus		17		17		
	223926	Vibrio parahaemolyticus RIMD 2210633	2		2			
	672	Vibrio vulnificus		26		26		
	11288	Viral hemorrhagic septicemia virus (STRAIN 07-71)	1		1			
	3972	Viscum album	20		20			
	11082	West Nile virus	5	36	17	117	12	81
	307044	West Nile virus strain 385-99	1		1			
	46919	Whitewater Arroyo virus		2		2		
	66077	Wolbachia sp. wMel		1		1		
X	35269	Woodchuck hepatitis virus			3	25	3	25
X	10430	Woodchuck hepatitis virus 1			3		3	
X	341946	Woodchuck hepatitis virus 2			4		4	
X	10433	Woodchuck hepatitis virus 8				1		1
	8355	Xenopus laevis		16		16		
	8364	Xenopus tropicalis		1		1		
	11089	Yellow fever virus		3	1	3	1	
	11090	Yellow fever virus 17D		4		4		
X	11091	Yellow fever virus Pasteur 17D-204			3		3	
	630	Yersinia enterocolitica	3	12	3	12		
	632	Yersinia pestis	26	38	36	64	10	26
X	214092	Yersinia pestis CO92				4		4
X	633	Yersinia pseudotuberculosis				20		20
X	186538	Zaire ebolavirus				21		21
	128952	Zaire ebolavirus - Mayinga (Zaire, 1976)	9	4	11	11	2	7

New 2007	NCBI TAX ID	SPECIES/STRAIN	B-06	T-06	B-07	T-07	Δ B	Δ T
	4577	Zea mays		1		1		
X	34245	Zinnia elegans				1		1

2 Website Features

The IEDB website functionality can be divided into five categories – Query, Resources, Account Information, Support, and About IEDB, which largely correspond to the pull-down menus at the top of the home page. The subsections that follow describe the website features within these categories. These features can also be classified by those available to any user and those available only to registered users. Any user can become a registered user in a simple process described in Section 2.4.1. The features are listed in Table 2.

The reader will find it helpful to keep in mind how data are stored in the IEDB. Each item contained in the IEDB consists of a reference (article or submission) containing information about one or more epitopes and associated binding or response information. The same epitope can exist in multiple references. Therefore, the results of a query will contain the same epitope numerous times if the epitope is contained in multiple references.

Table 2 IEDB website features available to anonymous and registered users

Anonymous User Features	Registered User Features
• Tour the IEDB • Submit feedback • View online help • View home page • Perform an Advanced Query • Perform a Simple Query • Perform a Search • Download Advanced Query results • Download Simple Query or Search results • Utilize analysis tools • Utilize finders when performing queries • View forum discussion topics • Search forum discussion topics • Download export files • Download documents • View external links • Use all features of the Analysis Resource	All features for anonymous users PLUS: • Submit a help request to the help desk • Save advanced queries to be used at a later time • Submit a discussion topic to a forum • Reply to a previously submitted discussion topic in a forum

2.1 Home Page

The IEDB Home Page is the default screen displayed when users enter the IEDB system. Besides providing a general description of the IEDB project, the home page displays system level status and notification of scheduled updates or maintenance. The page also contains a variety of Quick Links and other information including Summary Metrics and project related News and Updates. As users browse the IEDB system, they can return to the home page anytime by clicking *Home* on the far left of the main menu bar.

Summary Metrics are displayed in the lower left side of the screen. These numbers are intended to be a gauge of the volume of data available in the system. The Summary Metrics provide the

number of References, Records, Distinct Structures, and Distinct Epitopes that have been curated. These terms are fully defined in Section 2.2.4.

2.2 Query

There are six ways users can find information in the IEDB - Search, Simple Query, Advanced Query, Browse by Species, Browse by Allele, and Browse by 3D Structure. Users can retrieve information from curated literature or electronic submissions. Electronic submissions will usually contain relevant unpublished and unpatented epitope data. These six methods are elaborated in the subsections below.

The main menu bar appears on all screens (Figure 2.1). On the right hand side of the menu bar is a search field. This *Search* will return records that contain the term(s) entered, as in a Google search.

Figure 2.1 Picture of main menu bar

The *Advanced Query* is based on a standard Query By Example (QBE) approach, which is a method of forming queries where a user can enter conditions for each data field they want included in the query. The *Advanced Query* allows users to define example criteria for each field in the system. As there are over 300 fields, the *Advanced Query* function is both powerful and comprehensive. *Advanced Query* results are displayed in columnar format similar to a spreadsheet. The system allows the user to select which columns are displayed in the results and to download data.

The *Simple Query* is also based on the QBE approach. This form however, displays only the fields necessary to perform popular queries. It is geared to help users get the quick answers to the questions they typically ask the IEDB and is easier to read and understand. The results of a *Search* or *Simple Query* are organized into various lists the user can pick from. This allows the user to drill down quickly to the level that is of interest.

2.2.1 Browse for Records

The IEDB allows users to browse for records in three different ways – by MHC allele, by source species, and by 3D structure. Browse by 3D Structure was introduced early in 2007 and Browse by MHC allele received a new user interface.

2.2.1.1 Browse Records by MHC Allele

All users can find records associated with a specific MHC allele by browsing records by allele. To browse records by allele, the user selects *Browse Records by Allele* under the *Query* heading on the main menu or the Quick Links on the IEDB home page. The interface for the Browse by Allele changed in 2007 from a very long table of MHC alleles and their corresponding species to a tree structure that makes it much easier for users to find and investigate information on specific MHC alleles. As Figure 2.2 shows, the tree structure expands (and collapses) so users can drill down on species, MHC type, and allele to find the number of records in the IEDB for their MHC allele of interest. This number serves as a link that will display the records associated with the selected allele.

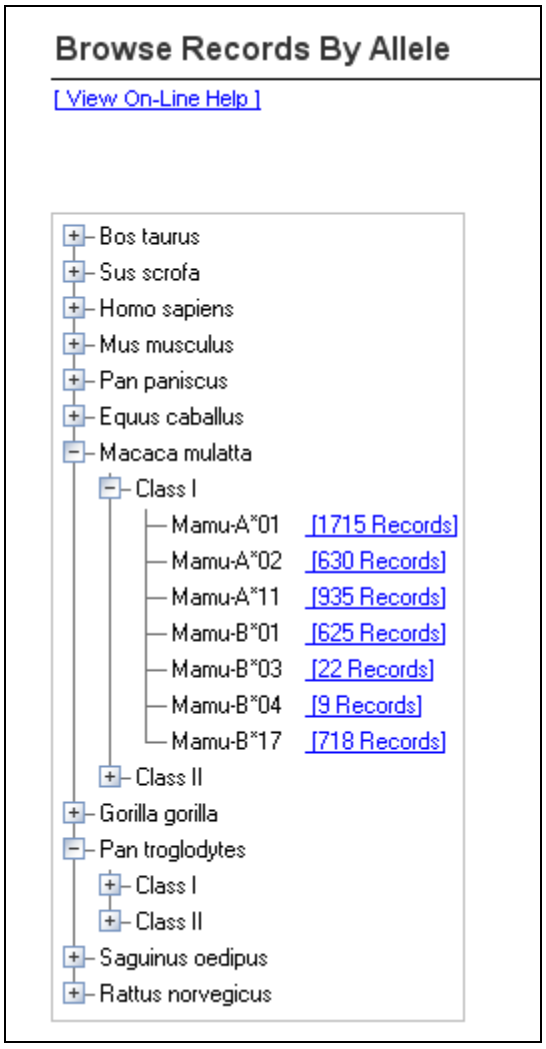

Figure 2.2 Browse records by MHC allele interface with new tree structure

2.2.1.2 Browse Records by Source Species

Users can find records associated with a specific epitope source species by browsing records by species. To do this, the user accesses the *Browse Records by Species* page via the Query pull-down menu or the Quick Links on the IEDB home page. A table will be displayed with all the epitope source species contained in the IEDB. The number of records associated with each species will be shown as a link. Clicking on a link will display the records associated with the selected species.

2.2.1.3 Browse Records by 3D Structure

Users can find IEDB records associated with PDB structures by browsing records by 3D structure. To do this, the user accesses the *Browse Records by 3D Structure* page via the Query pull-down menu or the Quick Links on the IEDB home page. A table will be displayed that indicates the number of B Cell, MHC binding, and T Cell assays with structural data. The number of records/assays for each category will be shown as a link. Clicking on a link will display the associated records.

2.2.2 Perform a Search

A Search will allow all users to locate records in the database using a keyword, identifier, or sequence. Wild card characters '%' and '_' can be used in the search field. The '%' character will match zero or more characters and the '_' character will match exactly one character. Additionally, the operators 'and', 'or', and 'not' can be utilized. The search field is case insensitive, so "ABC" is the same as "abc". Regular expressions cannot be used in the search field.

To perform a search, users enter criteria such as a keyword, identifier, or sequence into the search field on the main menu (Figure 2.1). They then click the Search button and view the Result Summary, which is described in Section 2.2.4.

In addition to using the search function from the menu bar, users can submit searches via a properly constructed URL. The example below will perform a search using the URL for records that contain 'dengue'. This would return the equivalent of typing 'dengue' in the search option on the menu bar and clicking the Search button.

http://www.immuneEpitope.org/httpQuery.do?dispatch=runquery&searchValue=dengue

2.2.3 Perform a Simple Query

The simple query was designed to allow users to perform common queries without having to search through the numerous fields available on the Advanced Query. To access the Simple Query search form, users select *Perform a Simple Search* under the *Query* heading on the main menu. The system will display the Simple Query Search Form (Figure 2.3). Data in the IEDB consists of References, Structures, and Assays. A reference can have many structures and a structure can have many assays. Each reference will have at least one structure and each structure will have at least one assay. When users perform a search, the system will determine which records match their search criteria down to the assay since that is the lowest level in the IEDB data hierarchy. The search form is written with descriptive logic to help the user understand how providing search criteria will affect the results.

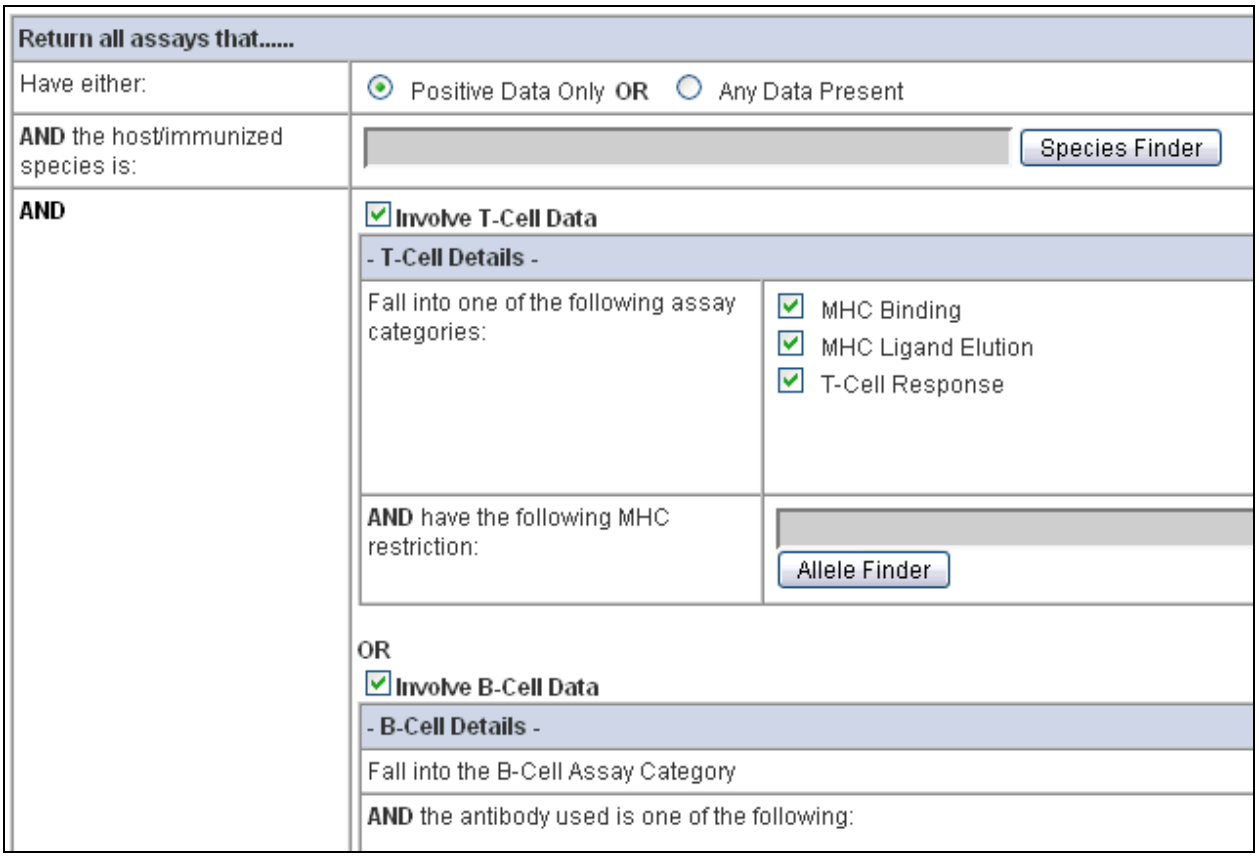

Figure 2.3 Top portion of the Simple Query Search Form

Some fields will allow multiple selections as search criteria. In these cases the selections are treated as a set. Records will be considered a match if they include at least one of the selected values in the set. Some fields will use finders (Section 2.2.6) to help users when the number of possible choices is extensive. For example the Species Finder and Allele finder are visible in the Simple Query Search Form seen in Figure 2.3.

If the T Cell Data section is selected, at least one of the assay categories (MHC Binding, MHC Ligand Elution, and T Cell Response) must be selected. If the B Cell Data section is selected, at least one of the antibody types (Monoclonal and Polyclonal) and at least one of the reference types (Literature, Submission) must be selected.

A Regular expression (http://www.digitalmars.com/ctg/regular.html) can be entered in the Linear Sequence field. Depending on the way a regular expression is written, users can search for single or multiple sequences.

The 'Citation or abstract contains' field at the bottom of the form utilizes the following wildcard characters: '%' and '_'. A '%' character matches zero or more characters and a '_' matches exactly one character. Also the operators 'And', 'Or', and 'Not' can be used with these fields. The following illustrates the precedence of the above operators from highest precedence to lowest: Not, And, Or. Parentheses can be used to alter precedence. Common words such as 'the', 'a', 'but', etc. are ignored if used as search criteria. Operators may be nested, but if no operators are used between words, results will include records that contain the entered string.

To perform a simple query, users select *Perform a Simple Query* under the *Query* heading on the main menu or in the Quick Links box on the IEDB home page. Users enter their search criteria, click the *Submit* button, and view the Results Summary (Section 2.2.4).

2.2.4 Search and Simple Query – Results Summary Page

The Result Summary page displays the results of a Search or Simple Query. The number of references, records, distinct epitopes, distinct structures, and assays (displayed as links) that met the specified search criteria are listed. The user can click on the number in the Count column (Figure 2.4) to view all the results for that category. The Result Summary allows the user to quickly assess the results of their search and drill down the level of detail that is of interest.

Record Type		Count
References		47
Records		331
Distinct Epitopes		
	Peptide	238
	DNA, RNA	0
	Lipid, Carbohydrate, Glycolipid, Other	0
Distinct Structures		
	Peptide	325
	DNA, RNA	0
	Lipid, Carbohydrate, Glycolipid, Other	0
Reported Measurements		
	MHC Binding Assays	229
	MHC Ligand Elution Assays	0
	T Cell Assays	456
	B Cell Assays	347

Figure 2.4 Sample of Result Summary Table

The number of **references** is the total number of references containing the matching records. Each **record** contains data about one epitope (or structure) in one reference. Any number of assays can be associated with a record. A protein/DNA sequence/carbohydrate etc. that may or may not induce an immune response or MHC binding is considered a structure. A **structure** is an epitope if it produces a positive qualitative measurement for at least one of its associated assays. A structure is not an epitope if it produces a negative qualitative measurement for each

of its associated assays. All epitopes are structures, but not all structures are epitopes. Multiple epitopes/structures may be described in a single reference. If the same epitope is described in two references, and a user defines criteria that match the epitope, the two records would be returned which are identical except for the reference information and possibly the assay information.

A **distinct epitope** is an epitope with given characteristics regardless of the reference. If the same epitope is described in two different references, and a user defines criteria that match the epitope, only one distinct epitope will be displayed in the results, although all associated assays from both references would be reflected in the reported number of assays. The number of distinct epitopes will always be less than or equal to the number of distinct structures.

A **distinct structure** is a structure with given characteristics regardless of the reference. If the same structure is described in two different references, and a user defines criteria that matches the structure, only one distinct structure will be displayed in the results. The number of records will always be greater than or equal to the number of distinct structures.

The distinct epitopes and distinct structures are categorized according to chemical type (e.g., protein/peptide, carbohydrate, and lipid).

Assays are divided by assay categories and are associated with records. Therefore, if the same epitope and associated assay are described in two references, and a user enters search criteria for the epitope, two identical assays (with the exception of the reference information) will be listed.

When users click on the number of references on the Result Summary screen, the system will display the **Reference List**. This list will display all the references that matched the user's search criteria. The columns in the list on the Reference List screen are not comprehensive. Enough data is listed to identify each reference uniquely, including authors, authors' affiliations, article title, year, PubMed ID, and journal name. To view all the information related to a given reference, click the corresponding Details link. The list can be sorted by clicking on the column headers. This will sort the list based on the values in that column. Clicking on the same column again reverses the sort order. PubMed identifiers are displayed as a hyperlink in the list. Clicking on the PubMed identifier in the system will open the PubMed citation in a new window.

When users click on the number of records on the Result Summary screen the system will display the **Record List**. This list will display all the structures that matched the user's search criteria. The columns in the list on the Record List screen are not comprehensive, but provide enough data on the reference and epitope (IEDB ID, name, chemical type, structure, and source) to identify each structure uniquely. To view all the information related to a given structure the user can click the corresponding Details link. Columns can be sorted as described above.

A user can click on the number of distinct structures or distinct epitopes on the Result Summary screen the system to display the **Distinct Structure List** or **Distinct Epitope List**, which list all the distinct structures/epitopes, respectively, that matched the user's search criteria. The columns displayed include the number of corresponding records, the number of positive and negative responses (epitopes have zero negative responses by definition), chemical type, chemical

structure information, and source species. As in the other lists, only a subset of the data is displayed in order to identify each structure or epitope uniquely, and the user can view all the information by clicking on the hyperlink in the "# Records" column.

The **Assay List** is displayed when users click on the number of assays on the Result Summary screen. This list will display all the assays that matched the user's search criteria. Only data relevant to the assay type are included. Because a subset of all available fields is included, the user can view all the information related to a given assay by clicking the corresponding Details link.

From the Assay List display, users can perform additional analysis using tools in the Analysis Resource. Users can analyze one or more rows on the same page of the results by checking the corresponding check boxes, selecting the type of analysis in the Analyze Selected Records box (Figure 2.5) and clicking the *Analyze* button. Users can select from T Cell epitope prediction, B Cell epitope prediction, population coverage, conservancy analysis, and homology mapping. Each tool offers independent online help or instructions.

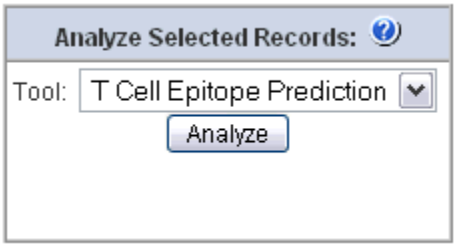

Figure 2.5 Analyze Selected Records box

The results of the reference, records, distinct epitope, distinct structure, and assay lists can be downloaded in a tab-delimited Excel format by clicking on the Excel link at the bottom of the screen. All rows will be downloaded regardless of the page shown or how many rows are displayed per page.

2.2.5 Advanced Query

The Advanced Query function allows the user to search the database using hundreds of different fields. It is the most comprehensive type of search in the IEDB system. The Advanced Query has numerous additional features that are not available through the Search and Simple Query methods. Most of these features are available to any anonymous user, but several are only available to registered users. Further information on these features is provided in the following subsections.

2.2.5.1 View Query Criteria

The Advanced Query function allows the user to leverage several hundred fields spread across various tabs and sections to perform a search. The location and volume of the fields can make it difficult to see what criteria in total has been entered and or selected. On the Advanced Query Search Form, the Query Results, and the Manage Saved Queries screens the user can opt to view the criteria for the selected query. After the user selects the *View Query Criteria* link the system will display the list of criteria for the selected query (Figure 2.6). From the Query Criteria screen

the user can execute the query by clicking the *Query Results* button or load the criteria into the Advanced Query Search Form for editing purposes by click the *Revise Query* button.

Criteria	Value
Maximum number of records to return:	10000
Epitope Chemical Type:	in list: Peptide/Protein
Reference Type:	= Literature
Article Year:	= 2005

Figure 2.6 Sample Advanced Query criteria

2.2.5.2 Save Queries

Registered users can save a query after they have performed an Advanced Query. Users cannot save a query after performing a Search or a Simple Query. A saved query will contain all of the entered filter criteria for the query and the format selected to report the results. When users save a query, they are required to give it a search title so it can be found in the future. Users might want to save a query so they can perform the query again at a later date when additional epitopes from other references have been added to the database. Once a query has been saved, users can load a saved query and then perform the query, or delete a saved query. Saved queries can also be modified, run, and saved again. Queries can be saved indefinitely if they are renewed by the user every year.

2.2.5.3 View Advanced Query Results

The system will display the results of an Advanced Query after the user's search criteria have been submitted from the Advanced Query Search Form or the View Query Criteria screen. The heart of the results screen is the result table (Figure 2.7). The user can change what columns are displayed in the result table by changing the Display Options or by defining a custom report.

The results table will display one row per assay. The reader may recall that each reference may contain multiple structures and each structure may have multiple assays. As such users may see epitope and reference information repeat like in the example below. To view all information about a single row in the results table, users can click the corresponding Details link. Other links will appear in the *Links* column such as the Epitope Viewer conditionally (Section 2.3.1.1.4).

10,000 items found, displaying 1 to 10.

Pages [First/Prev] 1, 2, 3, 4, 5, 6, 7, 8 [Next/Last]

☐	Links	Reference	Structure	Source
☐	Details	Nadia Caccamo Eur J Immunol 2004	SEFAYGSFVRTVSL	Mycobacterium tubercul
☐	Details	Nadia Caccamo Eur J Immunol 2004	SEFAYGSFVRTVSL	Mycobacterium tubercul
☐	Details	Nadia Caccamo Eur J Immunol 2004	SEFAYGSFVRTVSL	Mycobacterium tubercul
☐	Details	Nadia Caccamo Eur J Immunol 2004	SEFAYGSFVRTVSL	Mycobacterium tubercul
☐	Details	Nadia Caccamo Eur J Immunol 2004	SEFAYGSFVRTVSL	Mycobacterium tubercul
☐	Details	Nadia Caccamo Eur J Immunol 2004	SEFAYGSFVRTVSLPGADE	Mycobacterium tubercul
☐	Details	Nadia Caccamo Eur J Immunol 2004	SEFAYGSFVRTVSLPGADE	Mycobacterium tubercul
☐	Details	Nadia Caccamo Eur J Immunol 2004	SEFAYGSFVRTVSLPGADE	Mycobacterium tubercul
☐	Details	Nadia Caccamo Eur J Immunol 2004	SEFAYGSFVRTVSLPGADE	Mycobacterium tubercul
☐	Details	Nadia Caccamo Eur J Immunol 2004	SEFAYGSFVRTVSLPGADE	Mycobacterium tubercul

Figure 2.7 Sample Advanced Query Results Table

The result table will paginate the results to facilitate viewing. The page number and navigation links above and below the table can be used to move between pages of the results. Users can adjust the number of rows visible in the *Display Options* box at the top of the screen. To download the contents of the results table (all pages), users can click the *Export All Results* link below the results table. This will allow users to save the results in Microsoft Excel format to their local computer.

Users will often first perform a broad search then narrow their results by refining their search after looking at the results. If they want to change their search criteria, they can click the Revise Search link in the options box at the top of the screen rather than using the Back button in their browser.

Registered users can save their search criteria to execute a query again in the future. Users can also view the search criteria that were used to produce the results by clicking the View Query Criteria link in the options box at the top of the screen.

Users can also perform additional analysis using tools in the Analysis Resource. Users can analyze one or more rows on the same page of the results by checking the corresponding check boxes, selecting the type of analysis in the Analyze Selected Records box (Figure 2.5) and clicking the *Analyze* button. Each tool will offer independent online help or instructions.

2.2.5.4 Download Advanced Query Results

All users can download the results of an Advanced Query in a tab-delimited format that can be read by Excel. To download the results of a query, users must first perform the query for which they wish to download the results. The columns that are downloaded are the columns displayed in the Results Table when the user selects the Excel link.

2.2.5.5 Accessing the EpitopeViewer

All users are able to utilize the EpitopeViewer after an Advanced Query has been performed. The EpitopeViewer is an application for three dimensional viewing of receptor-antigen interactions that can be accessed from the View Results screen. The EpitopeViewer can be used with all assays that have receptor-antigen interaction data available.

2.2.6 Finders Overview

Several finders (Allele, Assay, Disease, Source, and Species) are available to help facilitate selections and control vocabulary usage (improves result outputs). At times the potential list of selections can be quite extensive, and the finders help users make selections from large lists. All finders can be utilized when performing an Advanced Query. The allele, source, and species finders can also be used when performing a Simple Query. Multiple selections can be made when utilizing finders during a query. Records will be returned that contain one of the entered values, or at least one of the entered values in the case of the MHC Types Present field.

The disease and species finders use wild card characters by default on both ends of entered search criteria when the Search names button is selected. An exact match is found when criteria are entered and the Search ID's button is selected. The allele finder uses wild card characters by default on both ends of criteria entered into the allele field. The source finder uses wild card characters by default on both ends of criteria entered into the name or strain fields. All finder search fields can search on common strings such as 'the', 'and', 'a', 'or', 'but', etc.

2.2.6.1 MHC Allele Finder

The MHC Allele Finder facilitates the selection of one or more MHC alleles. Initially the Allele Finder lists all alleles ordered by allele name. The Allele Finder allows the user to find alleles by name and or class {I, II, non-classical} in the Find box (Figure 2.8). After the user supplies their search criteria and clicks the Search button, the system will filter the list of MHC alleles using the name and or class provided. The allele finder uses wild card characters by default on both ends of criteria entered into the allele field. The system then returns any alleles that contain the value in the name field and match the class selected.

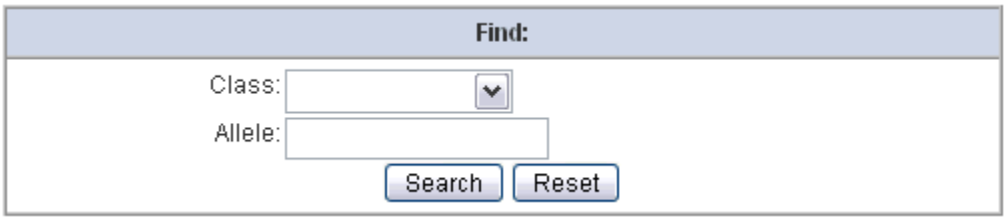

Figure 2.8 Search form on the MHC Allele Finder

2.2.6.2 Assay Type Finder

The Assay Type Finder is used to facilitate the selection of one or more assay types and lists all assay types in the selected assay category. The Assay Type Finder allows the user to find assay types by assay type name, assay group and or units in the Find box (Figure 2.9). After the user supplies their search criteria and clicks the Search button, the system filters the list of assay types using the selections provided in the Find box.

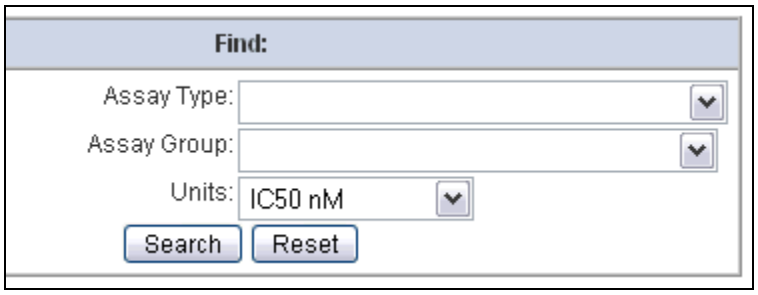

Figure 2.9 Search form on the Assay Type Finder

2.2.6.3 Disease Finder

The disease finder is used to facilitate the selection of a disease state and input the selection into a Disease State field. It includes all diseases from The International Statistical Classification of Diseases and Related Health Problems, tenth revision (ICD-10) and displays diseases with their corresponding ICD-10 codes in a hierarchical tree. The first level of the tree displays similar groups of diseases, and each additional level of the tree further breaks down the groups of diseases. Variations of each disease are not included. Searching capabilities are provided so users can quickly select a disease state.

The Disease Finder will allow the user to find diseases using the disease name or ICD-10 code (Figure 2.10). When the user performs a search, the system will display the first match and then allow the user to move forward and backward through the matching records using Next and Previous buttons (Figure 2.11). When the user provides a name, any disease name that contains the character string provided will be considered a match.

Figure 2.10 Search form on the Disease Finder

The selections in the disease finder are displayed in a tree (Figure 2.11). To expand a node of the tree, the user clicks the plus sign next to the name. To collapse a node, a user clicks the corresponding minus sign. The ICD-10 code is displayed next to each selection in square brackets. For example the ICD-10 code for Bartonellosis in the Figure 2.11 is A44.

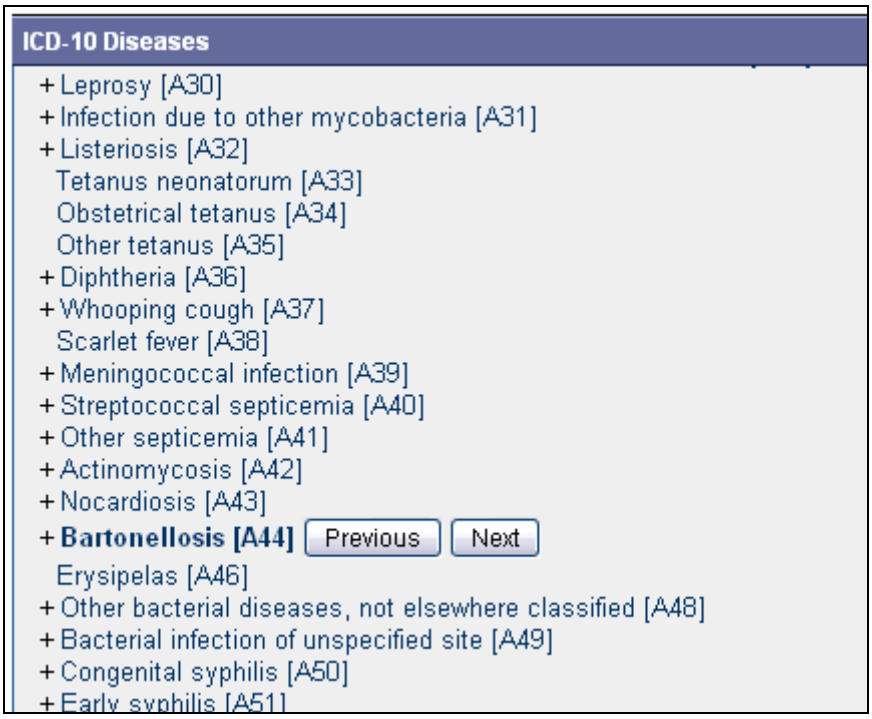

Figure 2.11 Disease List

2.2.6.4 Source Finder

The Source Finder is used to facilitate the selection of source antigens, immunogens, and epitopes. Records in the Source Finder come from GenBank, UniProt and IEDB curators. Among the finders, the Source Finder has the most comprehensive and flexible search form, and even includes the Species Finder. Due to the large volume of possible selections, the Source Finder initially won't display any selections. Users need to perform a search to narrow the list down. After the user enters search criteria (Figure 2.12) and clicks the Search button, the system will list the matching sources from which the user may select, as seen in Figure 2.13. The user can then select their desired sources from the list by click on *Select* in the far left column.

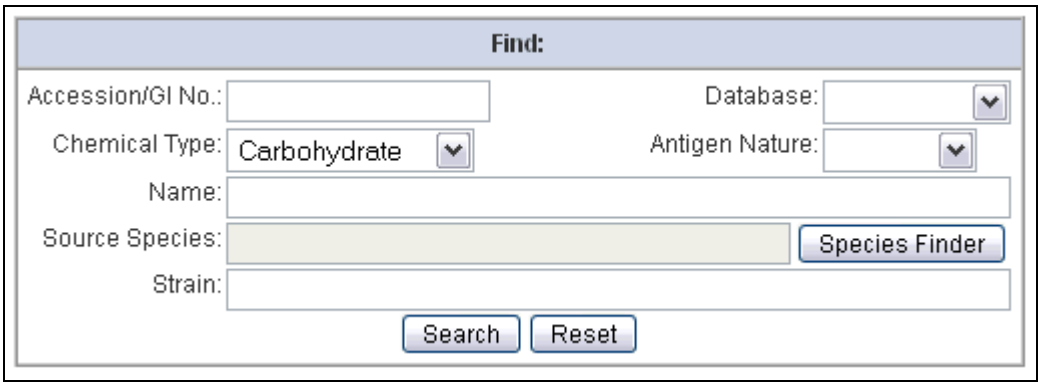

Figure 2.12 Search form on the Source Finder

3,681 items found, displaying 1 to 10.

Pages [First/Prev] **1**, 2, 3, 4, 5, 6, 7, 8 [Next/Last]

Options	Accession	Database	Name(s)	Source Species	Antigen Nature	Chemic Type
Select	38479432	GenBank	surface layer protein SapA8 ; sapA8	Campylobacter fetus subsp. fetus		Peptide/P
Select	83031685	GenBank	PB1-F2 protein; PB1-F2	Influenza A virus (A/Puerto Rico/8/34(H1N1))		Peptide/P
Select	10048255	GenBank	malaria exported protein-1; EXP-1	Plasmodium falciparum		Peptide/P
Select	10048261	GenBank	sporozoite surface protein 2; SSP2	Plasmodium falciparum		Peptide/P
Select	1017427	GenBank	elastic titin; titin	Homo sapiens		Peptide/P
Select	1042034	GenBank	1-phosphatidylinositol-4-phosphate 5-kinase isoform C; PtdIns4P 5-kinase isoform C; 1-phosphatidylinositol-4-phosphate	Homo sapiens		Peptide/P

Figure 2.13 Source List

2.2.6.5 Species Finder

The species finder is used to facilitate the selection of a species or virus from the NCBI Taxonomy Database. Common selections are displayed first to speed the selection of the usual suspects (image below). To view the entire NCBI taxonomy data set click *NCBI Taxonomy Tree* in the accordion slider. The Common Selection bar will move down to reveal the taxonomy tree. We have also provided an Allergen tree. The Allergen tree consists of species from the NCBI taxonomy, but is organized to help allergists locate common allergen selections more easily. To use the Allergen tree click the Allergen Tree heading in the accordion slider.

The species finder is used to facilitate the selection of a species or virus from the NCBI Taxonomy Database. Common selections are displayed initially to speed the selection of the usual suspects, as seen in Figure 2.14. To view the entire NCBI taxonomy data set click *NCBI Taxonomy Tree* in the accordion slider. The Common Selection bar will move down to reveal the taxonomy tree. An Allergen tree has also been provided. The Allergen tree consists of species from the NCBI taxonomy, but is organized to help allergists locate common allergen selections more easily. To use the Allergen tree click the Allergen Tree heading in the accordion slider.

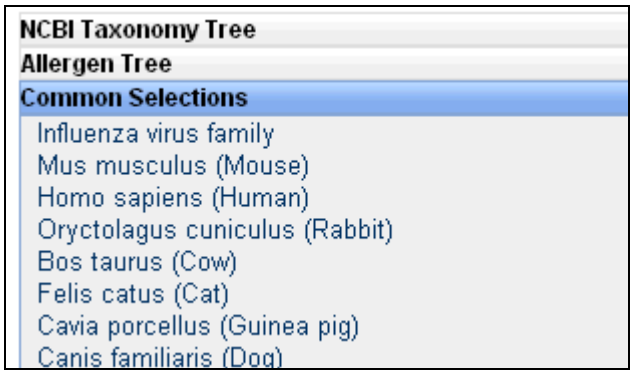

Figure 2.14 Common selections in Species Finder

The Species Finder will allow the user to find species using their name or taxonomy identifier (assigned by NCBI). When the user performs a search, the system will display the first match then allow the user to move forward and backward through the matching records using Next and Previous buttons as in Figure 2.15. When the user provides a name, any species name or synonym that contains the name provided will be considered a match. Search results will always appear in the NCBI Taxonomy tree, not the Allergen Tree or Common Selections.

The selections in the species finder are displayed in a tree (Figure 2.15). The taxonomy identifier is displayed next to each node of the tree in square brackets. For example, the taxonomy identifier for the selection in the example below is 301536. To see the synonyms for a selection, users can place their computer mouse over the scientific name.

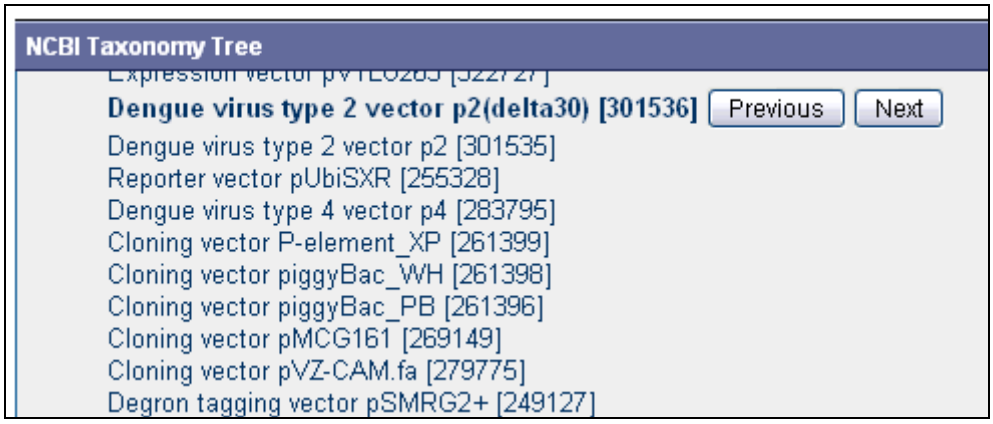

Figure 2.15 NCBI Taxonomy Tree

2.2.7 Analysis Tools

All users will be able to utilize analysis tools after an Advanced Query, Simple Query, or Search has been executed. From the View Results screen (for advanced queries) or an Assay List page (for simple queries and searches), the user will have the ability to perform various types of analysis on results. The analysis tools are described in greater detail in Section 2.3.1.

2.3 Resources

There are six major types of resource features available to users. Users can utilize the analysis and epitope prediction capabilities of the Analysis Resource. They can also make use of the

IEDB forum to discuss and exchange information and ideas. In addition, users can export the contents of the database for their own use and manipulation, and download an ever increasing number of documents from the website. Finally, links to other websites of interest and a list of potentially relevant patent items are provided.

2.3.1 Analysis Resource

The purpose of the Analysis Resource of the IEDB is to provide computational tools that enhance the value of the IEDB database to the user. Providing access to tools in one centralized location helps make users aware of available solutions to their problems. All of the information contained within the Analysis Resource, including analysis tools and algorithms developed by the IEDB staff, are freely available to the scientific community.

The tools provided in the analysis resource fall into two categories – analysis tools and prediction tools. Analysis tools help extract and interpret data contained in the database. Currently four tools exist - population coverage (Section 2.3.1.1.1), epitope conservancy analysis (Section 2.3.1.1.2), epitope cluster analysis (Section 2.3.1.1.3), epitope visualization (Section 2.3.1.1.4), and mapping B-cell epitopes to the PDB structures (Section 2.3.1.1.5**Error! Reference source not found.**).

Predictive tools extrapolate beyond data held in the database. They can be used to predict epitopes in protein sequences or predict properties of known epitopes, such as their MHC binding affinity.

For predictive tools, it is important to differentiate between the **tool** making predictions, and the **method** used to generate that tool, given a set of training data. For example, the artificial neural network method or approach, when trained on a particular data set, will yield a predictive tool. As the ANN method is trained on different data sets, different corresponding predictive tools will result. In this way, the ANN method can be used to develop an MHC class I prediction tool and a separate MHC class II prediction tool. These tools can be refined as more data are available for training. One benefit of the IEDB is that it allows implementing methods to automatically generate new predictive tools as the database grows.

Predictive tools can be subdivided into categories by what they aim to predict. The current tools fall into the subcategories listed below:

- T cell epitopes – MHC class I and II binding prediction
- T cell epitopes – Processing prediction (proteasomal cleavage/TAP transport)
- B-cell epitope prediction

The next subsections will describe each analytical and predictive tool in more detail.

2.3.1.1 Analysis Tools

2.3.1.1.1 Population coverage

T cells recognize a complex between a specific MHC type and a particular pathogen-derived epitope and thus a given epitope will elicit a response only in individuals that express an MHC

molecule capable of binding that particular epitope. MHC molecules are extremely polymorphic (over a thousand different variants are known in humans). Therefore, selecting multiple peptides with different MHC binding specificities will afford increased coverage of the patient population targeted as vaccine recipients. The issue of population coverage in relation to MHC polymorphism is further complicated by the fact that different MHC types are expressed at dramatically different frequencies in different ethnicities. Thus, without careful consideration, a vaccine with ethnically biased population coverage could result. To address this issue, the actual/predicted binding capacity of potential epitopes to as many different MHC molecules possible (and when available, also restriction data of T cell responses recognizing the epitope) can be used to project the population coverage in different ethnicities of different vaccine candidates or epitope sets. Accordingly, epitope-based vaccines or diagnostics can be designed to maximize population coverage, while minimizing complexity (that is, the number of different epitopes included in the diagnostic or vaccine), and also minimizing the variability of coverage obtained or projected in different ethnic groups.

An important consideration in the process of epitope selection is that the patient population coverage afforded by a given set is not simply corresponding to the sum of the coverage of its individual components. Thus, to calculate the coverage afforded by a given mixture of epitopes, a more comprehensive approach and a suitable algorithm has been developed for this specific purpose (Bui et al. BMC Bioinformatics 2006). This method calculates the fraction of individuals predicted to respond to a given epitope set on the basis of HLA genotypic frequencies, assuming non-linkage disequilibrium between HLA loci, and on the basis of MHC binding and/or T cell restriction data. The algorithm is briefly explained here. First, genotypic frequencies of various MHC are tabulated. Each time a peptide binds to a given MHC, a "hit" is recorded for that MHC. The process is repeated for all peptides. Then the hits for MHC are tallied. Next, the frequency of each possible diploid MHC combination (phenotype) is calculated. For n MHC types, this corresponds to an n x n tabulation of the frequency at which each specific pair of MHCs will be found in the population from which the MHC frequencies are derived. A similar table is generated to contain the number of hits per each of the MHC combinations by adding the number of hits associated with each of the two alleles of MHC in the combination (a simple exception is the case of homozygous combinations, where the number of hits is simply the number of hits of the given MHC). From these two tables, a frequency distribution is assembled, tabulating the genotypic frequency of all MHC combinations associated with a certain number of hits. The result of the analysis is displayed as a frequency distribution histogram and a cumulative frequency plot.

We have derived HLA allele genotypic frequencies from the dbMHC database (http://www.ncbi.nlm.nih.gov/mhc/) and stored them in a database on the IEDB tool server. At present, dbMHC provides allele frequencies for 78 populations and 11 different geographical areas. It is envisioned that the compiled data will be updated regularly as further HLA frequency data become available. Furthermore, customized frequency data can be utilized in the calculation, should studies of specific and particular patient populations be of interest to a given user. Multiple population coverages can be simultaneously calculated and an average population coverage is generated. Since MHC class I and II restricted epitopes elicit immune responses from two different T cell populations (CTL and Th cells, respectively), the program provides three different coverage calculation modes – (1) class I separate, (2) class II separate, and (3)

class I and class II combined.

2.3.1.1.2 Epitope conservancy

In a diagnostic or epitope-based vaccine setting, focusing on conserved epitopes allows for targeting responses around pathogen variability, whether it exists prior to infection, or develops in the natural course of disease. The use of conserved epitopes would be expected to focus the immune response on sequences crucial for retaining biological function of the pathogen proteins, and thus with intrinsically lower variability, even under immune pressure. The epitope conservancy analysis tools implemented here aims to address the issue of variability (or conservation) of epitopes, and to assist in the selection of epitopes with the desired pattern of conservation. The algorithm has been implemented to calculate the degree of conservancy of an epitope within a given protein sequence set at different degree of identities. The degree of conservation is defined as the number of protein sequences that contain the epitope at a given identity level, divided by the total number of protein sequences found in the dataset analyzed (Bui et al. BMC Bioinformatics 2007).

2.3.1.1.3 Epitope Cluster Analysis

This tool groups epitopes into clusters based on sequence identity. A cluster is defined as a group of sequences that has a sequence similarity greater than the minimum sequence identity threshold specified. Epitope sequences can be either directly entered in the text area or uploaded from a file. Two acceptable sequence formats are PLAIN and FASTA. The user can select the sequence identity threshold at which they want to calculate epitope clusters. Clusters are displayed in a table format where clusters are indicated by table rows which have the same color. All calculated cluster results can be saved to a file by clicking on the "Download data to file" button.

2.3.1.1.4 Visualization Tool

The convenient and easy to use EpitopeViewer, a Java application running JOGL, has been developed for three-dimensional visualization of immune epitopes and analyses of their interactions with antigen-specific receptors of the immune system (antibodies, T cell receptors, MHC molecules) for structures available in the Protein Data Bank (PDB). The EpitopeViewer is based on the Molecular Biology Toolkit (MBT; http://mbt.sdsc.edu/) developed at the San Diego Supercomputer Center (SDSC). It uses data both from the PDB and the IEDB, and visualizes one epitope at a time from a particular PDB structure (Beaver, et al., Immunome Res 2007).

The EpitopeViewer provides the following functionality:

- Link to the PDB web-page displaying a particular structure.
- Visualization of the 3D structure of epitope/antigen in complex with immune receptor(s) as curated within the IEDB and available in the PDB.
- Visualization of the 3D structure of epitope and antigen mapped to a PDB structure using the Homology Mapping tool.
- Visualization of sequences of epitope/antigen and immune receptor(s).

- 3D-visualization of intermolecular (epitope-paratope, epitope-antibody CDR, epitope-MHC, pMHC-TCR, pMHC-TCR CDR), inter-atom and inter-residue interactions curated within the IEDB and/or calculated on the fly from the PDB file with essential details (contact type, atoms, distance) provided;

- 2D-plot of inter-residue interactions between epitope and immune receptor.

- Generation of publication-quality pictures of structures, sequences, and plots of contacting residues.

2.3.1.1.5 Homology Mapping Tool

This tool maps linear epitopes to 3D structures of proteins (Beaver, et al., Immunome Res 2007). This is done by comparing the epitope source protein sequence with that of proteins with known 3D structures in the PDB. The tool generates an alignment between the query sequence of the epitope source sequence and a homologous sequence from the PDB, and allows to visualize the result in an EpitopeViewer. For input, the tool uses the SwissProt ID of the antigen protein, the epitope sequence, and the position of the epitope in the antigen sequence as curated within the IEDB or input by the user. The tool applies the NCBI BLAST algorithm for performing sequence homology search, and provides options for the sophisticated user to choose cutoff values on parameters used in the search programs (such as e-value and penalty on gap initiation and gap extension). The tool output page displays the alignment between the query sequence of the antigen containing the epitope and the sequence from the PDB representing significant hits (matches). The region within the epitope is highlighted in the alignment, and the sequence identity for the epitope and homologous region is provided. The EpitopeViewer application for visualization of homologous epitope/antigen and its further structural analysis is launched from the output page.

2.3.1.2 Predictive tools

2.3.1.2.1 T Cell Epitopes - MHC binding prediction

The Analysis Resource provides tools for predicting peptide binding to MHC class I and II molecules. For class I binding predictions, users can select predictions performed with tools derived from three different methods – artificial neural network, average relative binding, and stabilized matrix. For class II binding predictions, three new methods were introduced in 2007 to augment the previously available ARB method. These additional methods are the SMM_align method, a method devised by Sturniolo et al. and used in TEPITOPE, and consensus method derived from the three aforementioned methods. Tutorials and example data are available for both the class I and II tools.

2.3.1.2.1.1 Peptide Binding to MHC Class I Molecules

Users can select from three different methods for predicting class I epitopes – ANN, ARB, and SMM, which are described further below. During 2007, the ANN method was generalized to handle sequences of nine or more residues. Previously, it was constrained to handling only 9-mers.

Artificial Neural Network
Artificial neural networks (ANN) are computer algorithms modeled after the brain. They consist of many simple processing units which are wired together in a communication network. Each unit is a simplified model of a neuron which sends off a new signal if it receives a sufficiently strong input signal from the other units to which it is connected. The strength of these connections can be varied in order for the network to perform a desired pattern of node signal activity, which is learned from a set of input training data. The training data in this case are peptide sequences with quantitative affinities for a specific MHC molecule.

Many different implementations of artificial neural networks exist. The one utilized here is described for HLA-A2 binding predictions by Nielsen et al. (Protein Science, 2003) and has been applied to a number of different alleles (http://www.cbs.dtu.dk/services/NetMHC/).

Average Relative Binding (ARB)
Average relative binding (ARB) matrix binding prediction method is based on the assumption that each residue along the peptide molecule independently contributes to binding affinity. When a residue *R* occurs at position *i* in the peptide, it is assumed to contribute a constant amount of R_i to the free energy of binding of the peptide. The effect of each of the 20 possible amino acids at each possible position along the peptide sequence, therefore, can be estimated by a matrix of coefficients. The overall binding propensity of each peptide sequence, an algorithm "score", is calculated by multiplying the R_i coefficients. Predicted IC_{50} values, which provide quantitative K_D (IC_{50}) predictions, are then calculated by mathematical transformations of the algorithm scores (Bui et al., Immunogenetics 2005).

Stabilized Matrix Method (SMM)
The Stabilized Matrix Method (SMM) described by Peters and Sette (BMC Bioinformatics, 2005) can be applied to calculate matrices from quantitative affinity data of peptides binding to MHC molecules. The advantage of this method is that it suppresses the noise present in the training data, caused by the inevitable experimental error as well as the limited number of data points.

2.3.1.2.1.2 MHC Class I Binding Prediction Resource

In addition to prediction tools, the Analysis Resource makes data sets and method evaluations available to users. The IEDB MHC Class I Binding Prediction Resource (http://mhcbindingpredictions.immuneepitope.org) contains training data, test data, and other resources for tool developers interested in predictions of peptide binding to MHC class I molecules. The user can follow links to a manuscript describing the resource in detail (Peters et al., PLoS Comput. Biol. 2006), a dataset of experimental affinities of peptide to MHC molecules, and a description of the framework used for the evaluation of prediction methods. A link to this site appears at the bottom of the Analysis Resource main page.

As described in the manuscript, predictions were obtained from public web-servers for all relevant peptide-MHC affinities in the dataset. The correlation between predicted and measured affinities was evaluated using scatter plots, linear regression, and ROC analyses. The evaluation of these external tools can be accessed on the site by name of the method or the MHC allele. As

carefully noted in the manuscript, this is not a fair evaluation of the value of each method, primarily because the data available to each method are highly divergent.

A similar evaluation of the prediction performance of three prediction methods available in the IEDB Analysis Resource (ANN, ARB, and SMM) was carried out using cross-validation on the dataset. In contrast to the comparison of external predictions, this is a fair evaluation of prediction performance of the three methods, as training and testing data were the same for each method. Again, the evaluations of these three internal methods can be accessed by name of the method or the MHC allele.

2.3.1.2.1.3 Peptide Binding to MHC Class II Molecules

Users can select from three different methods for predicting class II epitopes – ARB, and SMM-align, Sturniolo, and Consensus. The Consensus method has been selected as the default method. An evaluation conducted by the IEDB team that will be published in 2008 has indicated that this method generally performs better than the others. The other three methods are described further below. Before 2007, only the ARB method was available for class II epitope prediction.

SMM-align

The MHC class II binding groove is open at both ends making the correct alignment of a peptide in the binding groove a crucial part of identifying the core of an MHC class II binding motif. The stabilization matrix alignment method, SMM-align, allows for direct prediction of peptide:MHC binding affinities. The method uses amino terminal peptide flanking residues (PFR) to get a consistent gain in predictive performance by favoring binding registers with a minimum PFR length of two amino acids. The method predicts quantitative peptide:MHC binding affinity values. The method has been trained and evaluated on a data set that covers the nine HLA-DR supertypes suggested and three mouse H2-IA allele. The method is described by Nielsen et al. (BMC Bioinformatics, 2007).

Sturniolo

This matrix-based approach is used in the TEPITOPE class II epitope prediction program. It is described in Sturniolo et al. (Nat. Biotechnol., 1999).

Consensus

The consensus method was developed by the IEDB team by exploiting features of the other three aforementioned methods. A manuscript describing the method and its comparison to the other methods is in preparation and should be published later in 2008.

2.3.1.2.2 T Cell Epitopes – Processing Prediction

2.3.1.2.2.1 Proteasomal cleavage/TAP transport/MHC class I combined predictor

For the prediction of antigen processing through the MHC class I antigen presentation pathway, we incorporated predictions of proteasomal cleavage and TAP transport similar to the MHCPathway website described in (Tenzer et al, CMLS, 2005). The predictions are based on in vitro experiments characterizing the sequence specificity of proteasomal cleavage and TAP transport. The goal of the prediction is to identify MHC-I ligands (peptides that are naturally processed from their source proteins and presented by MHC class I molecules).

The proteasomal cleavage predictions evaluate how efficiently a peptide or its N-terminally prolonged precursors can be liberated from its source protein. The TAP transport predictions evaluate how efficiently a peptide or its N-terminal prolonged precursors are transported into the ER by TAP (Peters, J Immunol, 2003). When this information is taken together and combined with MHC class I binding predictions, the tool yields a prediction of the efficiency with which a peptide is presented on the cell surface.

2.3.1.2.2.2 Neural network based prediction of proteasomal cleavage sites (NetChop) and T cell epitopes (NetCTL)

NetChop produces neural network predictions for cleavage sites of the human proteasome (Kesmir et al., 2002). NetChop takes into account the characteristics of the structurally modified proteasomes found in cells stimulated by gamma-interferon under physiological conditions. The NetChop algorithm was trained on in vitro data and MHC Class I ligand data. The use of this training set, combined with the artificial neural network methodology, makes the prediction of cleavage sites more accurate. NetChop has been trained only on human data, but since the proteasome structure is quite conserved, the algorithm developers believe that the tool is capable of making reliable predictions for at least the other mammalian proteasomes.

NetCTL predicts CTL epitopes in protein sequences integrating prediction of peptide MHC binding, proteasomal C terminal cleavage and TAP transport efficiency. The method is described in detail in Larsen et al. (Eur J Immunol., 2005).

2.3.1.2.3 B Cell Epitope Prediction

2.3.1.2.3.1 Prediction of epitopes from protein sequence

Six different tools are provided that predict antibody epitope candidates from amino acid sequences. Five are based on amino acid property scales and a new sixth method was introduced in 2007 that is includes the Hidden Markov Model. Parameters such as hydrophilicity, flexibility, accessibility, and antigenic propensity of polypeptides chains have been correlated with the location of continuous epitopes in a few well-characterized proteins. Based on these observations, amino acid property scales have been developed to predict antigenic determinants. Each scale consists of 20 values assigned to each of the amino acid residues on the basis of their relative propensity to possess the property described by the scale. The following amino acid property scales have been selected and implemented based on their popularity and coverage of different categories.

- Hydrophobicity/hydrophilicity
    - Parker hydrophilicity prediction
- Flexibility
    - Karplus and Schulz flexibility prediction
- Surface exposure
    - Emini surface accessibility prediction
- Antigenicity
    - Kolaskar and Tongaonkar antigenicity prediction
- Secondary structure
    - Chou and Fasman beta turn prediction

The new tool, BepiPred, combines the predictions of a hidden Markov model and the propensity scale of Parker et al. It is described in Larsen et al. (Immunome Research, 2006).

2.3.1.2.3.2 DiscoTope - Prediction of epitopes from protein structure

DiscoTope was introduced as a new tool in 2007 and is the only tool designed specifically to predict discontinuous epitopes. It uses protein three-dimensional structural data in addition sequence data. The method is based on amino acid statistics, spatial information, and surface accessibility in a compiled data set of discontinuous epitopes determined by X-ray crystallography of antibody/antigen protein complexes. The method is described in Haste Andersen et al. (Protein Sci., 2006).

2.3.2 Forums

The IEDB system includes a discussion forum to promote discussion, improve our user experience, and encourage community outreach. All users can search for discussion topics in forums, but only registered users can submit and reply to a discussion topic in a forum. Five forum categories are currently available – General, Queries/Reporting, Analysis Tools, NIAID B Cell Epitope Prediction Tools Workshop, and Epitope Datasets. New forums will be created by an Application Administrator in response to feedback, help requests, and frequently asked questions. Apart from the IEDB system, the forum has its own help content provided by Jive Forums. The forums can be accessed at http://www.immuneepitope.org/jive/index.jspa.

2.3.3 Database Export

The contents of the Immune Epitope Database are exported daily to files in XML format that can be downloaded by any user. There is one XML file for each reference in the IEDB. All of them are compressed and packaged together in a zip file (Figure 2.16). Users can learn more about the IEDB XML format in the reference materials available in the Documents area (Section 2.3.4).

Export List	Size
IEDB_2008_1_17_0_31.zip	20104k

Figure 2.16 Database export file

The file names are displayed as hyperlinks. The naming convention is IEDB_Year_Month_Day_Hours_Minutes.zip. For example, the file shown in Figure 2.16 was created January 17, 2008, at 12:31am. Users can download the file by right-clicking on the corresponding hyperlink and selecting "save link as" or "save target as", depending on the browser.

2.3.4 Documents

A variety of IEDB reference materials is available for download by all users. The Docuements page lists the files available for download by category. The files available for download include an Introduction to IEDB and Analysis Resource, the Curation Manual, IEDB Ontology, IEDB Annual Workshop Executive Summaries, the Annual Compendium, and quarterly newsletters. Additional reference materials will be added for download over time.

2.3.5 Links to External Sources

The IEDB system provides a list of links to external resources solely for the convenience of Immune Epitope Database visitors. The Immune Epitope Database has no interest in, responsibility for, or control over the linked-site. The Immune Epitope Database makes no promises or warranties of any kind, express or implied, including those of fitness for any particular purpose, as to the content of the linked-site. To view the links available, select Links under the Resources heading on the main menu. The hyperlinks on the links page are grouped by category:

- Antibody Related Links
- Bioinformatics Resource Centers
- Public Databases, Prediction Algorithms, and Other Tools
- MHC and TCR Related Links
- Protein Related Links
- Laboratory Resources
- Biodefense Resources

2.3.6 Patents

The Patents List page was introduced in 2007. As part of the IEDB curation effort, the Derwent World Patent Index has been searched for potentially relevant patent items. These are presented in a table on the web page that lists the publication number, patent title, inventors, assignee name, patent abstract, date filed, and date published. The enhanced abstracts of all of the 795 listed patent items have been reviewed, but the actual patents have not been read or curated. This information is presented for those users who wish to explore these patent items further.

The list mostly includes patents related to Category A-C priority pathogens, emerging and re-emerging infectious diseases, Malaria, Hepatitis B, Clostridium tetani, Leishmania, and Candida albicans, as well as other diseases. Users can search the table by using the "find" feature of their browser.

2.4 Account Information

2.4.1 Become a Registered User

IEDB has two types of users: anonymous users and registered users. Anonymous users are persons who use the IEDB application and do not register with IEDB. Users can become registered with the IEDB by submitting an e-mail address and a password, and indicating the institution they belong to, the country where they live, and if they would like a newsletter subscription. Registered users are able to use features that are unavailable to anonymous users, such as submitting help requests or posting on discussion forums.

2.4.2 Registered User Login

Users must be registered with IEDB in order to log in to the system. Users will be automatically prompted to log in to the IEDB upon attempting to use a function that requires user registration.

2.4.3 Modify Account Information

Account modifications consist of e-mail address, institution, country, password, and/or news letter subscription changes. Users must already be registered in order to have an account to modify.

2.4.4 Terms of Use

The terms of use statement can be viewed by selecting *Terms of Use* under the *Account* heading on the main menu. The Terms of Use page is a collection of statements that outline the conditions related to using the IEDB system. The Terms of use includes our privacy notice, copyright information, and various disclaimers.

2.5 Support Overview

The IEDB offers numerous ways to learn how to use the IEDB and to receive help if a problem is encountered. On almost every page of the IEDB, instructions are listed in a box at the top of the screen that describes how to perform the most important functions on the page. The instructions can be removed from all pages by clicking on the Hide Instructions link in the Quick Links box on the home page. Additionally, a more detailed description of each page can be accessed by clicking the  button to the right of a page title. When performing an Advanced Query, the user can click on a field name for a detailed description of the field. The online help can also be accessed by moving the cursor over the Support tab and selecting Online Help, which describes all features of the IEDB in detail.

Numerous tours that describe how to perform different types of queries and provide an overview of the IEDB are available in the "About IEDB" pull-down menu. Additionally, the forums described in the Resources section (2.3.2) are available so users can view, post, and reply to discussion topics relating to the IEDB or immunology in general.

The IEDB has done its best to make the application intuitive and to describe how to use all features of the IEDB. However, questions may arise that the online help, tours, and forums do not answer. In these cases, users can submit feedback or submit a help request.

2.5.1 Submit Feedback

All users are able to submit feedback, which will transmit an email message to the IEDB team. Feedback is intended for questions, input, and suggestions, such as new features they would like to see added in the future. Feedback helps the IEDB team update the system to provide users with the best possible experience. If users need help using the system or handling an unexpected result, a Help Request is probably more appropriate.

2.5.2 Submit Help Request

Only registered users logged in to the system can submit help requests. Help requests will be sent to the help desk. Help requests should consist of problems that users have with the application, such as a certain function of the system not working, and not questions on how to use features of the application. The online help is designed to answer questions regarding features of the application. A full name, day phone number, and description of the problem are required to submit a help request. After submitting a help request, a confirmation e-mail will be

sent to the user's registered e-mail address, which will include the help request number. When the help request issue has been resolved, an e-mail will be sent to notify the user that the help request has been satisfied.

Help requests are generally responded to within one business day. Purely technical requests are often responded to and resolved in the same one day period. Requests that are specific to analytical tools or the method used to curate data are answered initially to inform the requestor that their question/comment is being forwarded to team specialists, and an approximate date of full response is provided. Based on the complexity of the request or if that request prompts the team to make changes to the system or curate additional data/source organisms, requests are resolved immediately in some cases, while others are resolved in future IEDB system builds, or later curation.

2.5.3 View Online Help

All users are able to view the online help. Online help is designed to be an online user guide. It describes in detail all of the features that the system offers, explains how users go about doing everything the system has to offer, and tries to answer any questions that users might have when using the system.

The Contents tab is selected by default. The Table of Contents is displayed on the left and a help topic is displayed in the section on the right. Users can click on a folder in the Table of Contents and it will expand. Help topics contained within the selected folder will appear. Users can click on a topic and the corresponding topic will appear in the section on the right. The contents pane organizes help topics as they might appear in a user manual.

Users can select the Index tab to view similar topics grouped together and displayed alphabetically. Users can also select the Search tab and enter criteria into the search field. Relevant topics with be displayed. The index view organizes help topics as they might appear in the index of a book.

2.6 About IEDB

The "About IEDB" pull-down menu contains links to Flash Media tours on using the IEDB, an acknowledgement page, a list of relevant references, and instructions on how to cite the IEDB in publications and other work.

2.6.1 Tour the IEDB

Several video tours are available for viewing and download. They describe the IEDB in general, how to perform different types of queries, and how to interpret your query results. These tours are available in Flash Media format and are listed by subject. The title, length, and description for each video clip are listed. The script for each tour is provided in Microsoft Excel format. The tours and their descriptions are listed in Table 3.

Table 3 Flash media tours describing the basic IEDB functionality are available

Tour Category	Tour Name	Description
Overview	IEDB Overview	Provides an overview of the Immune Epitope Database and Analysis Resource or IEDB. The following modules are discussed: Query, Analysis tools, Discussion forum, and Online help.
Query Development Overviews	Search	Describes how to perform a Search.
	Simple Query	The Simple Query input form and how to input search criteria are discussed in this tour.
	Advanced Query	Outlines search criteria development and how registered users save advanced queries and load saved queries.
Query Results Overviews	Search and Simple Query	The result summary page and how to analyze and download results are discussed for a Search or Simple Query in this tour.
	Advanced Query	Describes how to format, download, and analyze Advanced Query results.
Sample Queries	Advanced Epitope Structure Query	Uses regular expression syntax to retrieve epitopes with a specific linear sequence using an Advanced Query. In this query, a regular expression is entered into the Epitope Linear Sequence field.
	Advanced Epitope Structure and Source Query 1	Demonstrates how to retrieve epitopes with a specific linear sequence and source name using an Advanced Query. The search fields include Epitope Linear Sequence and Epitope Source Name.
	Advanced Epitope Structure and Source Query 2	Demonstrates how to retrieve peptide epitopes for a species and use the custom report to format query results. The search fields include Epitope Chemical Type and Epitope Source Species.
	Advanced B Cell Query	Demonstrates how to find all epitopes with B Cell Response assays using monoclonal antibodies that react to a particular protein from a specific species strain using an Advanced Query. The search fields include B Cell Antibody Type and Epitope Source Name.
	Simple B Cell Query	Demonstrates how to find all epitopes with B Cell Response assays using monoclonal antibodies that react to a particular protein from a specific species strain using a Simple Query. The search fields include B Cell Antibody Type and Epitope Protein or Gene.
	Simple T Cell Query	Demonstrates how to retrieve T cell epitopes with a given MHC restriction using a Simple Query. The Assay Categories and MHC Restriction fields are used.

2.6.2 Acknowledgements

A host of talented individuals have worked hard to make the Immune Epitope Database a reality. A roster of the current team members can be viewed on this page.

2.6.3 Publications

A list of publications relevant to the IEDB can be found on the Publications page.

2.6.4 Citing the IEDB

Data and tools within the IEDB are presented as a public resource. Users are requested to consider citing the IEDB when they present information obtained from the IEDB or use tools contained in the Analysis Resource. It is expected that the authors of an entry as well as the IEDB are properly cited whenever their work is referred to:

1. The IEDB website should be cited using the URL: www.immuneepitope.org

2. The journal reference for the IEDB should be cited as:

   Peters B, Sidney J, Bourne P, Bui HH, Buus S, Doh G, Fleri W, Kronenberg M, Kubo R, Lund O, Nemazee D, Ponomarenko JV, Sathiamurthy M, Schoenberger S, Stewart S, Surko P, Way S, Wilson S, Sette A. The immune epitope database and analysis resource: from vision to blueprint. PLoS Biol. 2005 Mar;3(3):e91. PMID: 15760272.

3 Scientific Publications

This section lists the scientific publications in 2007 for which the IEDB played a contributory role. The first subsection lists publications authored by the IEDB contractor team. The second subsection lists articles that cited the IEDB. This list was generated by searching for "immune epitope database" or "IEDB" in PubMed and Google Scholar for 2007, and by searching for the designated IEDB citation paper (Peters et al., PLoS Biol. 2005) in Google Scholar. The IEDB was cited twice as often in 2007 as it was in 2006.

3.1 Publications of the IEDB team

1. Beaver JE, Bourne PE, Ponomarenko JV. EpitopeViewer: a Java application for the visualization and analysis of immune epitopes in the Immune Epitope Database and Analysis Resource (IEDB). Immunome Res. 2007 Feb 21;3:3. PMID: 17313688

2. Blythe MJ, Zhang Q, Vaughan K, de Castro R Jr, Salimi N, Bui HH, Lewinsohn DM, Ernst JD, Peters B, Sette A. An analysis of the epitope knowledge related to Mycobacteria. Immunome Res. 2007 Dec 14;3(1):10. PMID: 18081934

3. Bui HH, Peters B, Assarsson E, Mbawuike I, Sette A. Ab and T cell epitopes of influenza A virus, knowledge and opportunities. Proc Natl Acad Sci U S A. 2007 Jan 2;104(1):246-51. PMID: 17200302

4. Bui HH, Sidney J, Li W, Fusseder N, Sette A. Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines. BMC Bioinformatics. 2007 Sep 26;8(1):361. PMID: 17897458

5. Ernst JD, Lewinsohn DM, Behar S, Blythe M, Schlesinger LS, Kornfeld H, Sette A. Meeting Report: NIH Workshop on the Tuberculosis Immune Epitope Database. Tuberculosis (Edinb). 2007 Dec 6. PMID: 18068490

6. Greenbaum JA, Andersen PH, Blythe M, Bui HH, Cachau RE, Crowe J, Davies M, Kolaskar AS, Lund O, Morrison S, Mumey B, Ofran Y, Pellequer JL, Pinilla C, Ponomarenko JV, Raghava GP, van Regenmortel MH, Roggen EL, Sette A, Schlessinger A, Sollner J, Zand M, Peters B. Towards a consensus on datasets and evaluation metrics for developing B-cell epitope prediction tools. J Mol Recognit. 2007 Mar-Apr;20(2):75-82. PMID: 17205610

7. Lundegaard C, Lund O, Kesmir C, Brunak S, Nielsen M. Modeling the adaptive immune system: predictions and simulations. Bioinformatics. 2007 Dec 15;23(24):3265-75. PMID: 18045832

8. Nielsen M, Lundegaard C, Blicher T, Lamberth K, Harndahl M, Justesen S, Røder G, Peters B, Sette A, Lund O, Buus S. NetMHCpan, a method for quantitative predictions of peptide binding to any HLA-A and -B locus protein of known sequence. PLoS ONE. 2007 Aug 29;2(8):e796. PMID: 17726526

9. Peters B, Sette A. Integrating epitope data into the emerging web of biomedical knowledge resources. Nat Rev Immunol. 2007 Jun;7(6):485-90. PMID: 17479127

10. Ponomarenko JV, Bourne PE., Antibody-protein interactions: benchmark datasets and prediction tools evaluation. BMC Struct Biol. 2007 Oct 2;7(1):64. PMID: 17910770

11. Sette A, Peters B., Immune epitope mapping in the post-genomic era: lessons for vaccine development. Curr Opin Immunol. 2007 Feb;19(1):106-10. PMID: 17113275

12. Wang P, Morgan AA, Zhang Q, Sette A, Peters B. Automating document classification for the Immune Epitope Database. BMC Bioinformatics. 2007 Jul 26;8:269. PMID: 17655769

3.2 Publications Citing the IEDB

1. Amela I, Cedano J, Querol E. Pathogen proteins eliciting antibodies do not share epitopes with host proteins: a bioinformatics approach. PLoS ONE. 2007 Jun 6;2(6):e512. PMID: 17551592

2. Antonets DV, Bakulina AY, Portnyagina OY, Sidorova OV, Novikova OD, Maksyutov AZ. Prediction of antigenically active regions in the OmpF-like porin of Yersinia pseudotuberculosis. Dokl Biochem Biophys. 2007 May-Jun;414:124-6. PMID: 17695318

3. Assarsson E, Sidney J, Oseroff C, Pasquetto V, Bui HH, Frahm N, Brander C, Peters B, Grey H, Sette A. A quantitative analysis of the variables affecting the repertoire of T cell specificities recognized after vaccinia virus infection. J Immunol. 2007 Jun 15;178(12):7890-901. PMID: 17548627

4. Beaver JE, Bourne PE, Ponomarenko JV. EpitopeViewer: a Java application for the visualization and analysis of immune epitopes in the Immune Epitope Database and Analysis Resource (IEDB). Immunome Res. 2007 Feb 21;3:3. PMID: 17313688

5. Bui HH, Botten J, Fusseder N, Pasquetto V, Mothe B, Buchmeier MJ, Sette A. Protein sequence database for pathogenic arenaviruses. Immunome Res. 2007 Feb 8;3:1. PMID: 17288609

6. Bui HH, Peters B, Assarsson E, Mbawuike I, Sette A. Ab and T cell epitopes of influenza A virus, knowledge and opportunities. Proc Natl Acad Sci U S A. 2007 Jan 2;104(1):246-51. PMID: 17200302

7. Calvo-Calle JM, Strug I, Nastke MD, Baker SP, Stern LJ. Human CD4+ T cell epitopes from vaccinia virus induced by vaccination or infection. PLoS Pathog. 2007 Oct 12;3(10):1511-29. PMID: 17937498

8. Deluca DS, Blasczyk R. The immunoinformatics of cancer immunotherapy. Tissue Antigens. 2007 Oct;70(4):265-71. PMID: 17767547

9. Dinakarpandian D, Lee Y, Dinakar C. Applications of medical informatics in allergy/immunology. Ann Allergy Asthma Immunol. 2007 Jul;99(1):2-9; quiz 9-12, 41. PMID: 17650823

10. Gajria B, Bahl A, Brestelli J, Dommer J, Fischer S, Gao X, Heiges M, Iodice J, Kissinger JC, Mackey AJ, Pinney DF, Roos DS, Stoeckert CJ Jr, Wang H, Brunk BP, ToxoDB: an integrated Toxoplasma gondii database resource. Nucleic Acids Res. 2008 Jan;36(Database issue):D553-6. Epub 2007 Nov 14. PMID: 18003657

11. Glasner JD, Plunkett G 3rd, Anderson BD, Baumler DJ, Biehl BS, Burland V, Cabot EL, Darling AE, Mau B, Neeno-Eckwall EC, Pot D, Qiu Y, Rissman AI, Worzella S, Zaremba S, Fedorko J, Hampton T, Liss P, Rusch M, Shaker M, Shaull L, Shetty P, Thotakura S, Whitmore J, Blattner FR, Greene JM, Perna NT. Enteropathogen Resource Integration Center (ERIC): bioinformatics support for research on biodefense-relevant enterobacteria. Nucleic Acids Res. 2008 Jan;36(Database issue):D519-23. Epub 2007 Nov 13. PMID: 17999997

12. Greenbaum JA, Andersen PH, Blythe M, Bui HH, Cachau RE, Crowe J, Davies M, Kolaskar AS, Lund O, Morrison S, Mumey B, Ofran Y, Pellequer JL, Pinilla C, Ponomarenko JV, Raghava GP, van Regenmortel MH, Roggen EL, Sette A, Schlessinger A, Sollner J, Zand M, Peters B. Towards a consensus on datasets and evaluation metrics for developing B-cell epitope prediction tools. J Mol Recognit. 2007 Mar-Apr;20(2):75-82. PMID: 17205610

13. Greene JM, Collins F, Lefkowitz EJ, Roos D, Scheuermann RH, Sobral B, Stevens R, White O, Di Francesco V. National Institute of Allergy and Infectious Diseases bioinformatics resource centers: new assets for pathogen informatics. Infect Immun. 2007 Jul;75(7):3212-9. PMID: 17420237

14. Heiny AT, Miotto O, Srinivasan KN, Khan AM, Zhang GL, Brusic V, Tan TW, August JT., Evolutionarily conserved protein sequences of influenza a viruses, avian and human, as vaccine targets. PLoS ONE. 2007 Nov 21;2(11):e1190. PMID: 18030326

15. Jaaskelainen S, Riikonen P, Salakoski T, Vihinen M. Evaluation of Protein Hydropathy Scales. Bioinformatics and Biomedicine, 2007. BIBM 2007. IEEE International Conference on. 2007:245-251.

16. Karpenko O, Huang L, Dai Y. A probabilistic meta-predictor for the MHC class II binding peptides. Immunogenetics. 2007 Dec 19. PMID: 18092156

17. Kennedy R, Poland GA. T-Cell epitope discovery for variola and vaccinia viruses. Rev Med Virol. 2007 Mar-Apr;17(2):93-113. PMID: 17195963

18. Kessler JH, Melief CJ. Identification of T-cell epitopes for cancer immunotherapy. Leukemia. 2007 Sep;21(9):1859-74. PMID: 17611570

19. Lundegaard C, Lund O, Kesmir C, Brunak S, Nielsen M. Modeling the adaptive immune system: predictions and simulations. Bioinformatics. 2007 Dec 15;23(24):3265-75. PMID: 18045832

20. Parida R, Shaila MS, Mukherjee S, Chandra NR, Nayak R. Computational analysis of proteome of H5N1 avian influenza virus to define T cell epitopes with vaccine potential. Vaccine. 2007 Oct 23;25(43):7530-9. PMID: 17900763

21. Rapberger R, Lukas A, Mayer B. Identification of discontinuous antigenic determinants on proteins based on shape complementarities. J Mol Recognit. 2007 Mar-Apr;20(2):113-21. PMID: 17421048

22. Samarghitean C, Väliaho J, Vihinen M. IDR knowledge base for primary immunodeficiencies. Immunome Res. 2007 Mar 29;3:6. PMID: 17394641

23. Sette A, Peters B., Immune epitope mapping in the post-genomic era: lessons for vaccine development. Curr Opin Immunol. 2007 Feb;19(1):106-10. PMID: 17113275

24. Squires B, Macken C, Garcia-Sastre A, Godbole S, Noronha J, Hunt V, Chang R, Larsen CN, Klem E, Biersack K, Scheuermann RH., BioHealthBase: informatics support in the elucidation of influenza virus host pathogen interactions and virulence. Nucleic Acids Res. 2008 Jan;36(Database issue):D497-503. Epub 2007 Oct 26. PMID: 17965094

25. Tong JC, Tan TW, Ranganathan S. Methods and protocols for prediction of immunogenic epitopes. Brief Bioinform. 2007 Mar;8(2):96-108. PMID: 17077136

26. Tung CW, Ho SY. POPI: predicting immunogenicity of MHC class I binding peptides by mining informative physicochemical properties. Bioinformatics. 2007 Apr 15;23(8):942-9. PMID: 17384427

27. Wang P, Morgan AA, Zhang Q, Sette A, Peters B. Automating document classification for the Immune Epitope Database. BMC Bioinformatics. 2007 Jul 26;8:269. PMID: 17655769

4 References

Beaver JE, Bourne PE, Ponomarenko JV. EpitopeViewer: a Java application for the visualization and analysis of immune epitopes in the Immune Epitope Database and Analysis Resource (IEDB). Immunome Res. 2007 Feb 21;3:3. PMID: 17313688

Bui HH, Sidney J, Peters B, Sathiamurthy M, Sinichi A, Purton KA, Mothe BR, Chisari FV, Watkins DI, Sette A. Automated generation and evaluation of specific MHC binding predictive tools: ARB matrix applications. Immunogenetics. 2005 Jun;57(5):304-14. Epub 2005 May 3.

Bui HH, Sidney J, Dinh K, Southwood S, Newman MJ, Sette A. Predicting population coverage of T-cell epitope-based diagnostics and vaccines. BMC Bioinformatics. 2006 Mar 17;7(1):153. PMID: 16545123

Bui HH, Sidney J, Li W, Fusseder N, Sette A. Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines. BMC Bioinformatics. 2007 Sep 26;8(1):361. PMID: 17897458

Chou P, Fasman G. Prediction of the secondary structure of proteins from their amino acid sequence. Adv Enzymol Relat Areas Mol Biol. 1978:45–148. PMID: 364941

Emini E, Hughes J, Perlow D, Boger J. Induction of hepatitis A virus-neutralizing antibody by a virus specific synthetic peptide. J Virol. 1985;55:836–839. PMID: 2991600

Haste Andersen P, Nielsen M, Lund O. Prediction of residues in discontinuous B-cell epitopes using protein 3D structures. Protein Sci. 2006 Nov;15(11):2558-67. Epub 2006 Sep 25. PMID: 17001032

Karplus PA, Schulz GE. Prediction of chain flexibility in proteins. Naturwissenschaften. 1985. 72:212-213.

Keşmir C, Nussbaum AK, Schild H, Detours V, Brunak S. Prediction of proteasome cleavage motifs by neural networks. Protein Eng. 2002 Apr;15(4):287-96. PMID: 11983929

Kolaskar AS, Tongaonkar PC. A semi-empirical method for prediction of antigenic determinants on protein antigens. FEBS Lett. 1990 Dec 10;276(1-2):172-4. PMID: 1702393

Larsen JE, Lund O, Nielsen M. Improved method for predicting linear B-cell epitopes. Immunome Res. 2006 Apr 24;2:2. PMID: 16635264

Larsen MV, Lundegaard C, Lamberth K, Buus S, Brunak S, Lund O, Nielsen M. An integrative approach to CTL epitope prediction: a combined algorithm integrating MHC class I binding, TAP transport efficiency, and proteasomal cleavage predictions. Eur J Immunol. 2005 Aug;35(8):2295-303. PMID: 15997466

Nielsen M, Lundegaard C, Lund O. Prediction of MHC class II binding affinity using SMM-align, a novel stabilization matrix alignment method. BMC Bioinformatics. 2007 Jul 4;8:238. PMID: 17608956

Nielsen M, Lundegaard C, Worning P, Lauemoller SL, Lamberth K, Buus S, Brunak S, Lund O., Reliable prediction of T-cell epitopes using neural networks with novel sequence representations, Protein Sci. 2003 May;12(5):1007-17

Parker J, Guo D, Hodges R. New hydrophilicity scale derived from High-Performance Liquid Chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. Biochemistry. 1986;25:5425–5432. PMID: 2430611

Peters B, Sidney J, Bourne P, Bui HH, Buus S, Doh G, Fleri W, Kronenberg M, Kubo R, Lund O, Nemazee D, Ponomarenko JV, Sathiamurthy M, Schoenberger S, Stewart S, Surko P, Way S, Wilson S, Sette A. The immune epitope database and analysis resource: from vision to blueprint. PLoS Biol. 2005 Mar;3(3):e91. PMID: 15760272.

Peters B, Bui HH, Frankild S, Nielson M, Lundegaard C, Kostem E, Basch D, Lamberth K, Harndahl M, Fleri W, Wilson SS, Sidney J, Lund O, Buus S, Sette A., A community resource benchmarking predictions of peptide binding to MHC-I molecules, PLoS Comput. Biol. 2006 Jun 9;2(6):e65. Epub 2006 Jun 9. PMID: 16789818

Peters B, Sette A., "Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method." BMC Bioinformatics 2005 May 31;6(1):132

Peters B, Bulik S, Tampe R, Van Endert PM, Holzhutter HG. Identifying MHC class I epitopes by predicting the TAP transport efficiency of epitope precursors. J Immunol. 2003 Aug 15;171(4):1741-9.

Sturniolo T, Bono E, Ding J, Raddrizzani L, Tuereci O, Sahin U, Braxenthaler M, Gallazzi F, Protti MP, Sinigaglia F, Hammer J. Generation of tissue-specific and promiscuous HLA ligand databases using DNA microarrays and virtual HLA class II matrices. Nat Biotechnol. 1999 Jun;17(6):555-61. PMID: 10385319

Tenzer S, Peters B, Bulik S, Schoor O, Lemmel C, Schatz MM, Kloetzel PM, Rammensee HG, Schild H, Holzhutter HG. Modeling the MHC class I pathway by combining predictions of proteasomal cleavage, TAP transport and MHC class I binding. Cell Mol Life Sci. 2005 May;62(9):1025-37.