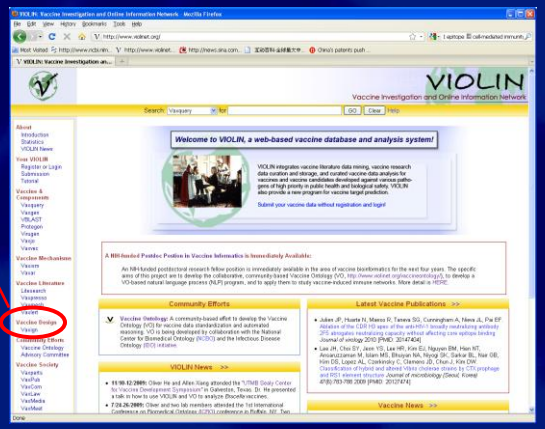
Vaxign Reverse Vaccinology Software Demo Introduction

Zhuoshuang "Allen" Xiang, Yongqun "Oliver" He

Unit for Laboratory Animal Medicine
Department of Microbiology and Immunology
Center for Computational Medicine and Bioinformatics
Comprehensive Cancer Center
University of Michigan Medical School
Ann Arbor, MI 48109

Vaxign is a program in VIOLIN

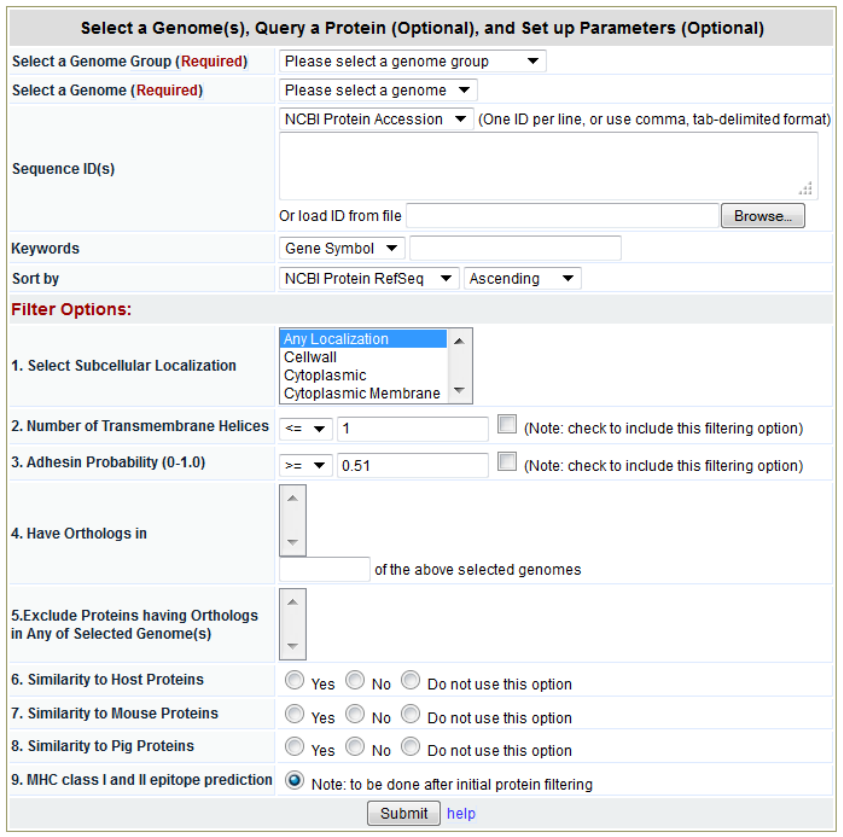
- VIOLIN: Vaccine Investigation and Online Information Network
- A vaccine research database and vaccine data analysis system
- Key tasks:
  - Build a database for vaccines and vaccine candidates in use, clinical trials, and research
  - Vaccine literature mining and curation
  - Vaccine design → **Vaxign**
  - Development and applications of community-based Vaccine Ontology (VO)



Publically available:
http://www.violinet.org/

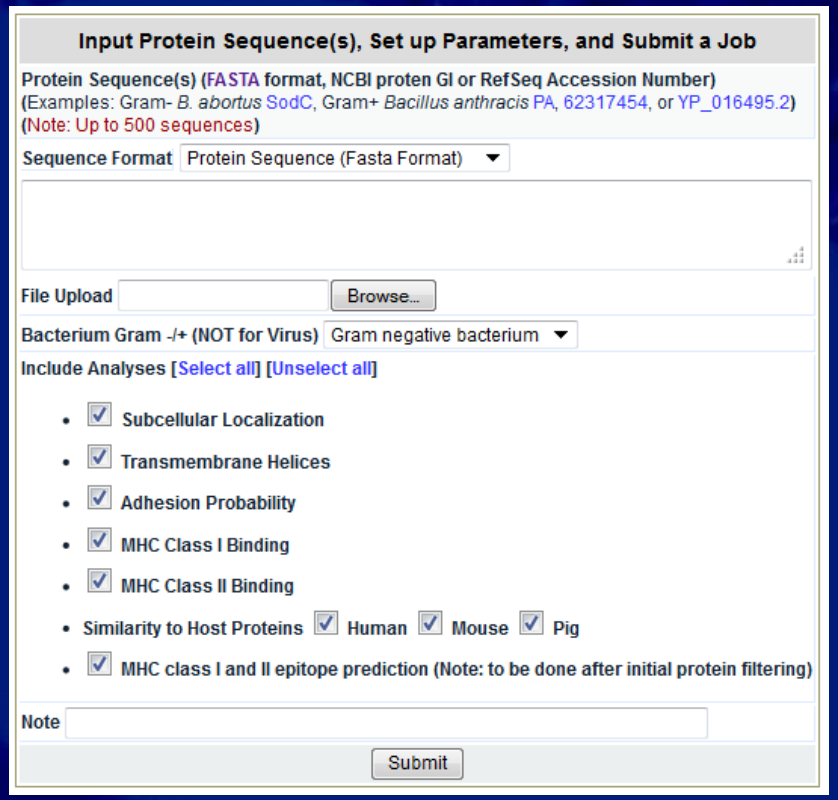
Two Forms of Vaxign Usage

Pre-computed data query



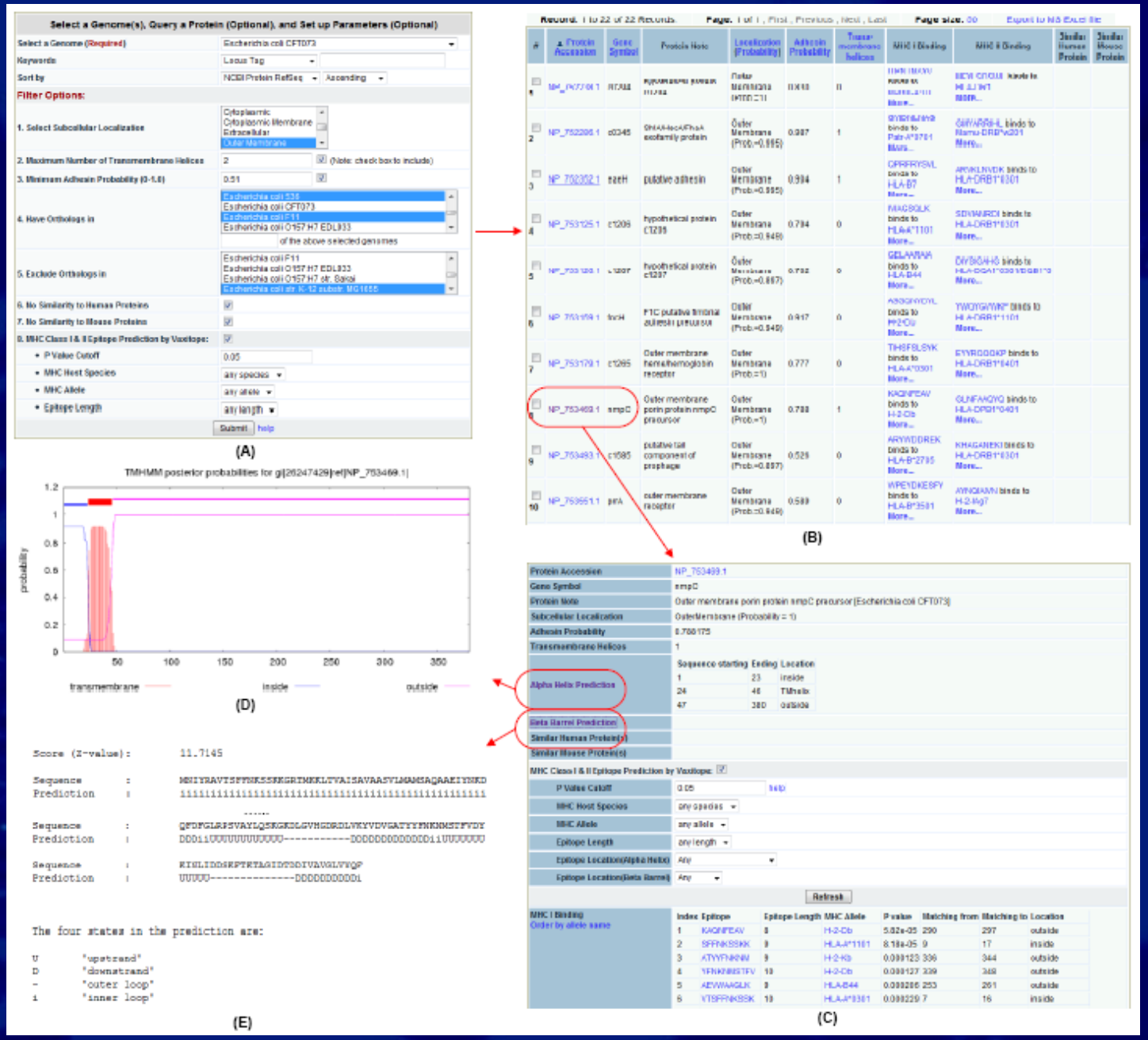
- Pre-computed results
- > 200 genomes in database
- Easy to query

Dynamic analysis



- Runtime execution
- Can analyze up to 500 proteins at one time

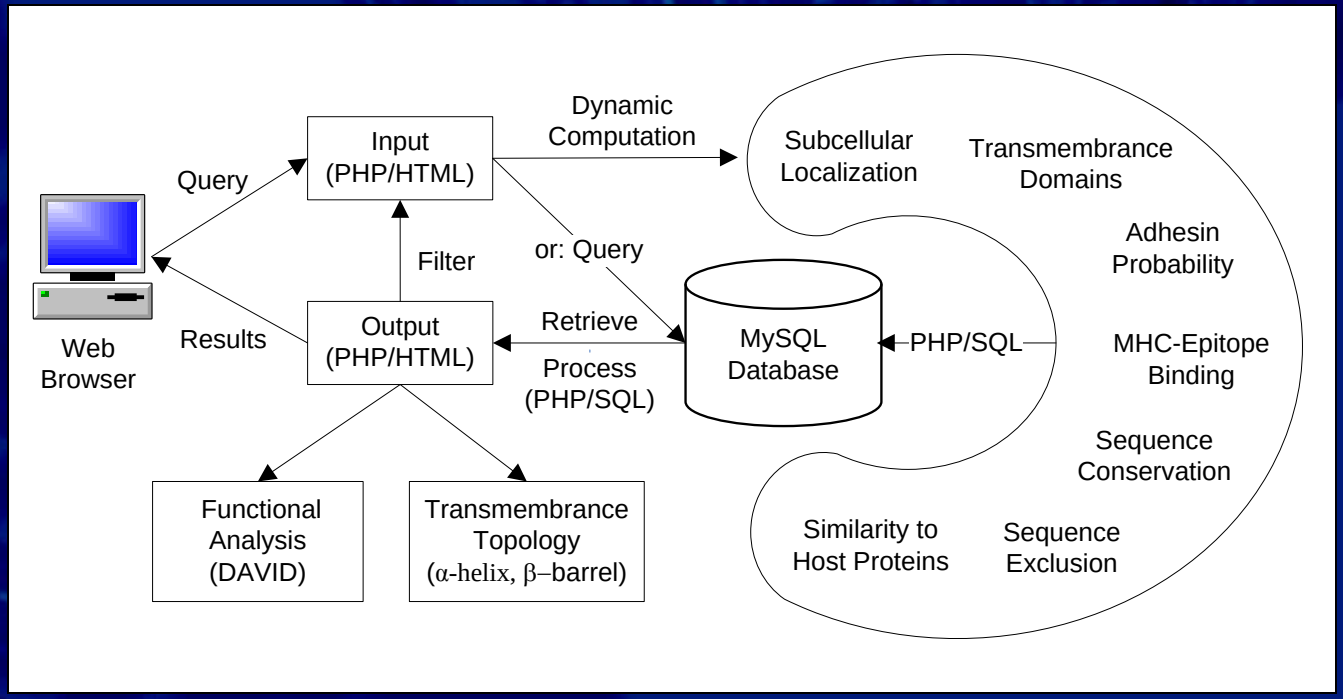
Vaxign analysis pipeline demo



Vaxign
VACCINE DESIGN

Vaxign: Vaccine Design System

- Aim: Vaccine target prediction for reverse vaccinology



- The 1st web-based reverse vaccinology system
- Freely available: http://www.violinet.org/vaxign

Reference: He Y, Xiang Z, Mobley HLT. Vaxign: the first web-based vaccine design program for reverse vaccinology and an application for vaccine development. *Journal of Biomedicine and Biotechnology*. Volume 2010 (2010), Article ID 297505, 15 pages. [PMID: 20671958]

Vaxign: how to predict subcellular localization?

- Using customized PSORTb 3.0:
  - Source: http://www.psort.org/psortb/
  - Has overall precision (specificity) of 96%
  - GNU General Public Licence (GNU GPL)

- Vaxign customizations:
  - Automatically generate PSORTb input
  - Parse PSORTb output and store in Vaxign relational database

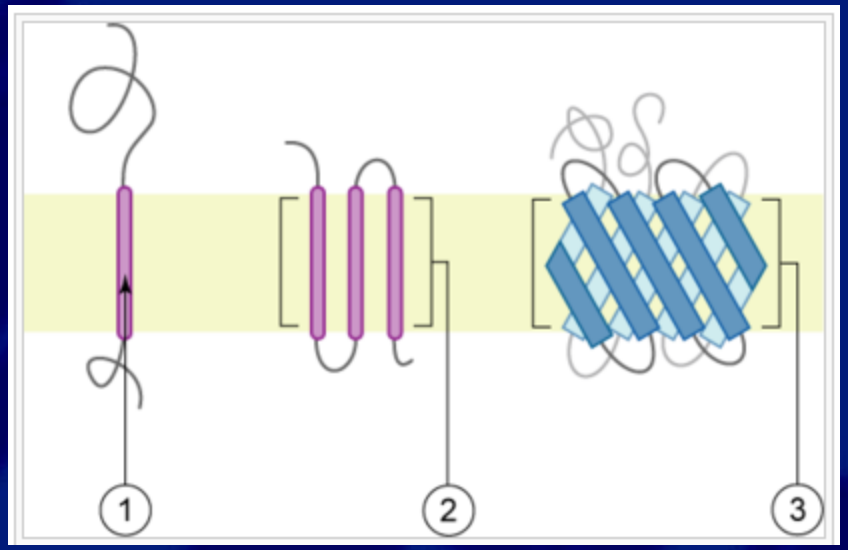
Vaxign: how to predict transmembrane helices?

- Using customized TMHMM:
  - Source: http://www.cbs.dtu.dk/services/TMHMM-2.0/
  - based on a hidden Markov model
  - Academic Software License Agreement

- Vaxign customizations:
  - Automatically generate TMHMM input
  - Parse TMHMM output and store in Vaxign relational database

Vaxign: how to predict β-barrel?

- Using customized PROFtmb:
  - http://www.ncbi.nlm.nih.gov/pubmed/16844988
  - Four-state (up-, down-strand, inner loop, outer loop) accuracy as high as 86%
  - GNU General Public License



http://en.wikipedia.org/wiki/Transmembrane_protein

Vaxign: how to predict adhesin probability?

- Using customized SPAAN:
  - Adhesin: cell-surface components that facilitate bacterial adhesion or adherence to other cells.
  - Source: http://www.ncbi.nlm.nih.gov/pubmed/15374866
  - Has optimal sensitivity of 89% and specificity of 100% on a defined test set
  - Default value is 0.51
  - Freely available

Vaxign: how to predict sequence conservation?

- Using customized BLAST-based OrthoMCL:
  - Source: http://www.orthomcl.org/
  - E-value of 10^{-5} is set as cutoff
  - GNU General Public License

- Alternative internal BLAST method: based on reciprocal best fit principle.

Vaxign: how to do gene functional analysis?

- Link to DAVID:
  - Source: http://david.abcc.ncifcrf.gov/
  - What Vaxign does: parse data in DAVID format and send to DAVID website.

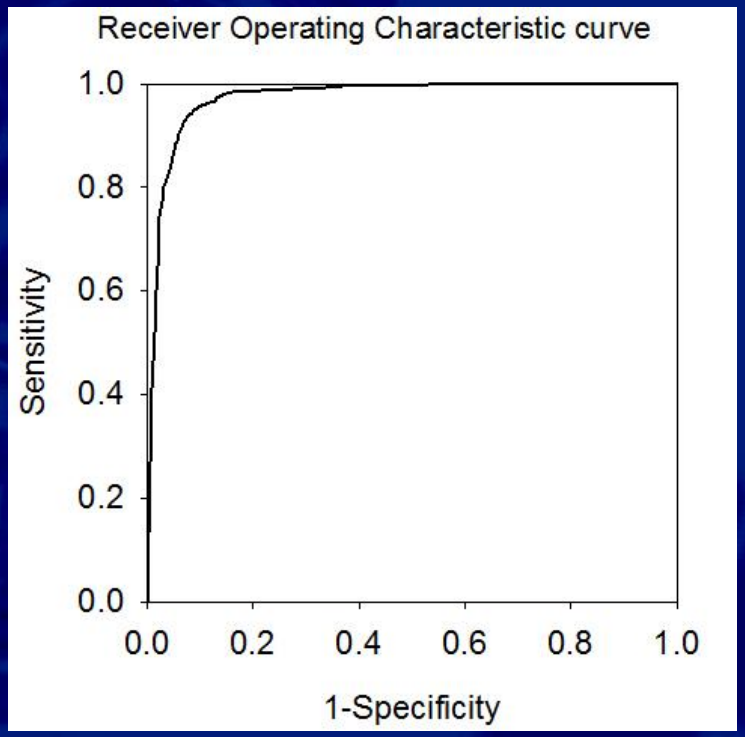
- Develop our own COG-based analysis:
  - Clusters of Orthologous Groups of proteins (COGs): http://www.ncbi.nlm.nih.gov/COG/
  - Classify proteins
  - TO DO: Fisher exact test to find significantly enriched COG groups.

Vaxign Vaxitop program: Epitope Prediction

- An MHC Class I & II binding epitope prediction tool developed internally
- Based on: position specific scoring matrix (PSSM)
- Uniqueness: calculate statistical P value (instead of a percentage or top number) as the cutoff

Vaxitop performance

- Vaxitop is specific and sensitive
- Comparable to known epitope prediction tools
- Example:



Receiver operating characteristic (ROC)

For human MHC class I allele: HLA A*0201

ROC Area Under Curve (AUC)
AUC value: 0.971

Vaxitop P-value of 0.05 has a balanced sensitivity and specificity

#	MHC allele	length	AUC	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity
				(P = 0.01)		(P = 0.05)		(P = 0.1)	
1	HLA-A*01:01	9	0.984	0.962	0.9	1	0.791	1	0.699
2	HLA-A*02:01	9	0.971	0.647	0.979	0.961	0.889	0.986	0.776
3	HLA-A*02:01	10	0.943	0.695	0.961	0.933	0.803	0.978	0.559
4	HLA-B*15:17	9	0.976	0.597	0.996	0.925	0.901	0.984	0.755
5	HLA-B*35:01	9	0.96	0.687	0.978	0.942	0.799	0.987	0.674
6	HLA-A*02:02	9	0.946	0.624	0.96	0.95	0.805	0.988	0.565
7	HLA-A*02:02	10	0.909	0.542	0.946	0.908	0.731	0.988	0.45
8	HLA-A*02:03	9	0.959	0.558	0.99	0.908	0.87	0.981	0.689
9	HLA-A*02:03	10	0.878	0.553	0.941	0.878	0.671	0.97	0.392
10	HLA-A*02:06	9	0.918	0.539	0.968	0.886	0.796	0.964	0.641
11	HLA-A*03:01	9	0.97	0.745	0.974	0.977	0.823	0.994	0.638
12	HLA-A*03:01	10	0.896	0.66	0.909	0.971	0.469	0.996	0.22
13	HLA-A*11:01	9	0.98	0.664	0.991	0.98	0.867	0.998	0.69
14	HLA-A*11:01	10	0.9	0.6	0.942	0.971	0.448	1	0.2
15	HLA-A*24:03	9	0.995	0.799	1	0.991	0.895	0.995	0.763
16	HLA-A*29:02	9	0.932	0.651	0.953	0.929	0.764	0.953	0.608
17	HLA-A*30:01	9	0.964	0.694	0.989	0.91	0.892	0.95	0.806
18	HLA-A*31:01	9	0.968	0.644	0.981	0.957	0.849	0.993	0.725
19	HLA-A*68:01	9	0.837	0.561	0.903	0.966	0.451	0.997	0.11
20	HLA-A*68:01	10	0.875	0.572	0.937	0.955	0.472	0.996	0.141

New feature: Perform IEDB MHC class I epitope prediction and compare with Vaxitop results.

Vaxign Use Case Studies

Two use cases published:

- For *Brucella*:

Citations: -- Xiang Z, He Y. 2009. Vaxign: a web-based vaccine target design program for reverse vaccinology. *Procedia in Vaccinology*. Volume 1, Issue 1, Pages 23-29.

-- He Y, Xiang Z. Bioinformatics analysis of *Brucella* vaccines and vaccine targets using VIOLIN. *Immunome Res*. 2010 Sep 27;6 Suppl 1:S5.

- For uropathogenic *E. coli:*

Citation: He Y, Xiang Z, Mobley HLT. Vaxign: the first web-based vaccine design program for reverse vaccinology and an application for vaccine development. *Journal of Biomedicine and Biotechnology*. Volume 2010 (2010), Article ID 297505, 15 pages.

Two use cases for this afternoon's exercises:

- For human herpesvirus 1 (*i.e.*, Human simplex virus 1)
- For enterohemorrhagic *E. coli* strain O157:H7

Use Case 1: Human Herpesvirus 1 Vaccine Design

- Downloaded from NCBI RefSeq database:
  - Three HHV-1 genomes
  - 12 HHV genomes
  - 52 herpesvirus genomes

- Vaxign pre-computed analysis ready for query
  1) 77 proteins in human hepsevirus 1
  2) All present in three human hepsevirus 1 genomes
  3) Only 19 conserved in 12 human hepsevirus genomes
  4) 17 removed due to high transmembrane domains
  5) One with high adhesin probability: NP_044628.1, which is a capsid scaffold protein (UL26.5)

Results and discussion:

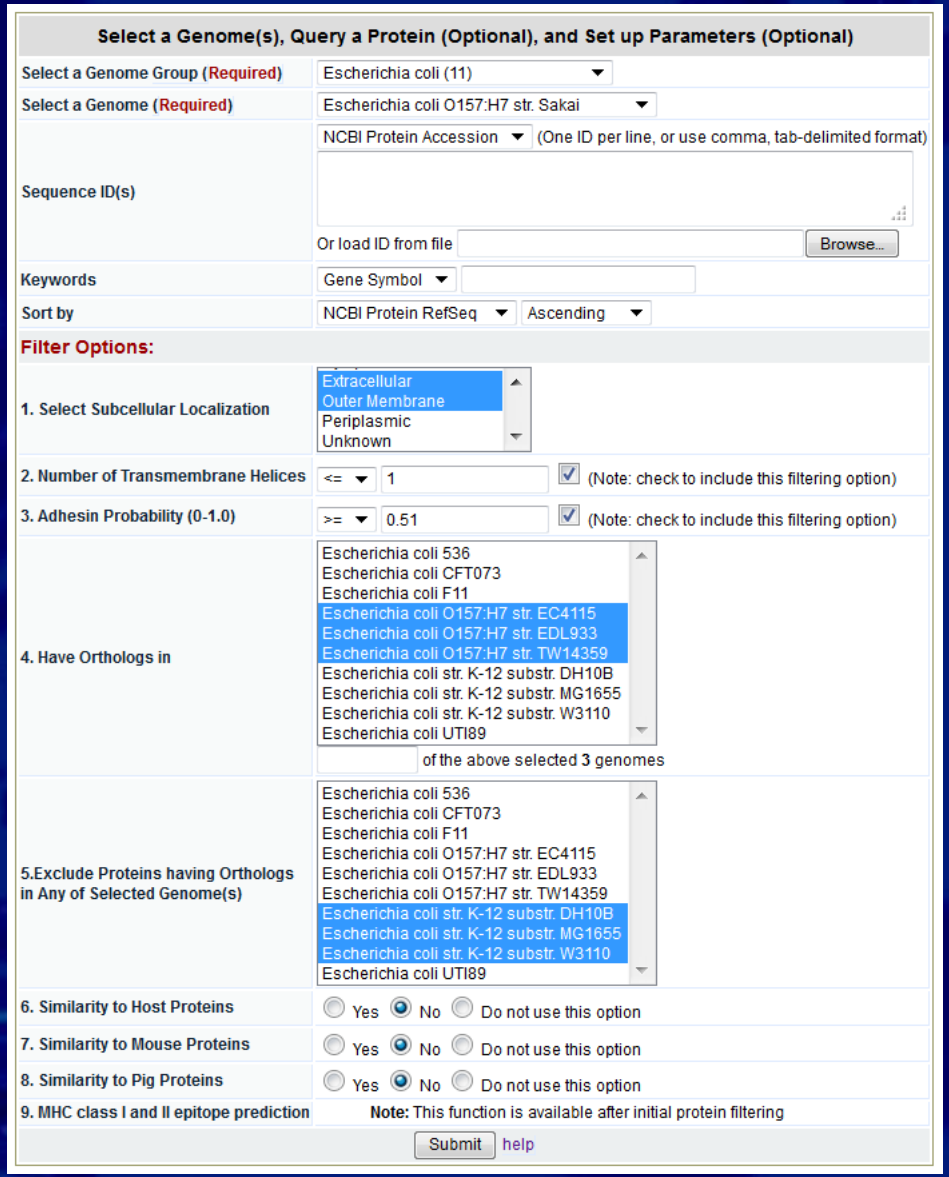
See Oliver’s presentation for details.

Use Case 2: enterohemorrhagic *E. coli* strain O157:H7

- Genomes available for four pathogenic O157:H7 strains:
  - str. EC4115
  - str. EDL933
  - str. Sakai (NOTE: used as seed genome since it comes from RefSeq database)
  - str. TW14359

- Control: three *E. coli* str. K-12 genomes
  - substr. DH10B
  - substr. W3110
  - substr. MG1655

A Vaxign setting for *E. coli* strain O157:H7 vaccine design



- Choose str. Sakai as seed genome
- Outer membrane or secreted proteins
- Conserved in all O157:H7 strains
- Absent from nonpathogenic K-12 strains
- Others …

Vaxign results for *E. coli* strain O157:H7 vaccine design

- 35 genes found
- DAVID gene functional enrichment results

Rerun Using Options | Create Sublist

30 chart records

Download File

Sublist	Category	Term	RT	Genes	Count	%	P-Value	Benjamini
☐	GOTERM_CC_FAT	outer membrane	RT		30	50.0	1.1E-17	1.6E-16
☐	GOTERM_CC_FAT	cell outer membrane	RT		24	40.0	2.5E-14	1.8E-13
☐	GOTERM_BP_FAT	biological adhesion	RT		13	21.7	5.4E-14	7.0E-13
☐	GOTERM_BP_FAT	cell adhesion	RT		13	21.7	5.4E-14	7.0E-13
☐	GOTERM_CC_FAT	external encapsulating structure part	RT		24	40.0	3.5E-10	1.7E-9
☐	GOTERM_CC_FAT	cell envelope	RT		24	40.0	5.0E-8	1.8E-7
☐	GOTERM_BP_FAT	pathogenesis	RT		7	11.7	5.6E-8	3.6E-7
☐	COG_ONTOLOGY	Intracellular trafficking and secretion / Extracellular structures	RT		4	6.7	1.8E-6	7.1E-6
☐	GOTERM_CC_FAT	pilus	RT		13	21.7	1.7E-5	4.8E-5
☐	GOTERM_CC_FAT	envelope	RT		24	40.0	3.3E-5	7.7E-5
☐	KEGG_PATHWAY	Pathogenic Escherichia coli infection	RT		3	5.0	6.1E-5	3.0E-4
☐	KEGG_PATHWAY	Pathogenic Escherichia coli infection	RT		3	5.0	7.3E-5	1.8E-4
☐	GOTERM_CC_FAT	external encapsulating structure	RT		24	40.0	8.5E-5	1.7E-4
☐	GOTERM_CC_FAT	cell projection	RT		13	21.7	2.4E-4	4.1E-4

Conclusion

- Vaxign is a specific and sensitive predictor of vaccine targets.
- It is web-based, user-friendly, and free.

http://www.violinet.org/vaxign/

Discussion

Challenges:

How to better rank predicted vaccine targets?

What's unique about protective antigens?

Directions: integrated with microarray, proteomics, literature mining, 3D structure, and VIOLIN components.

Acknowledgements

He Laboratory / U of Michigan (UM)

Funding: NIH-NIAID Grant 1R01AI081062