Evaluation and Development of GPCR Classifiers for Vectors
Ronald J. Nowing1, Jenica Abrudan2,3, Douglas A. Shoue2,3, Badii Abdul-Wahid1, Mariha Wadsworth2,3, Gwen Stayback2,3, Frank Collins2,3, Mary Ann McDowell2,3, Jesus Izaguirre2,3
1Computer Science & Engineering, University of Notre Dame 2Biological Sciences, University of Notre Dame 3Eck Institute for Global Health, University of Notre Dame

Abstract
We aim to inexpensively develop insecticides for disease vectors such as mosquitoes by incorporating bioinformatics and computational biology into all aspects of the drug development process. Due to the popularity of G-Protein Coupled Receptors (GPCRs) as drug targets, a GPCR classifier that perform well on vector proteomes and provides a prediction confidence score for each identified peptide is of great interest.

Still at the early stages of our project, we seek to identify a set of top drug target candidates from among the G-Protein Coupled Receptors (GPCRs), proteins which are popular drug targets, in the vector proteomes. We have evaluated two existing GPCR classifiers [GPCRHM [1] and PredCouple[4]) on six genomes (Ae. aegypti, An. gambiae, Ap. mellifera, Dr. melanogaster, Ho. sapiens, and Pe. humanus). In addition, we have developed and evaluated an ensemble classifier that provides a probability for each sequence, enabling an intuitive way to control the trade off between sensitivity and accuracy. We show that our ensemble classifier provides greater or equal sensitivity with approximately equal accuracy. Source: http://structbio.vanderbilt.edu/sanders/Research.htm

Results
Ensemble*, GPCRHM, PredCouple, GPCRHM*, and Pfam* were evaluated on six organisms (Ae. aegypti, An. gambiae, Ap. mellifera, Dr. melanogaster, Ho. sapiens, and Pe. humanus). For every organism, Ensemble* was able to identify the most test set GPCRs.

Identification and Validation of Novel GPCRs
Ensemble* was used to identify 52 novels GPCRs from the vectors Ae. aegypti, An. gambiae, and Pe. humanus. The predictions were validated with a pipeline that includes the tools BLAST, ScanPROSITE, and I-TASSER.

Methods
We developed three novel classifiers (GPCRHM*, Pfam*, and Ensemble*). GPCRHM* converts the global scores from GPCRHM to a discrete likelihood score between 0 and 1, while Pfam* converts e-values from the Pfam A GPCR clan Hidden Markov Models (HMMs) to a discrete likelihood score between 0 and 1. Ensemble* combines the likelihood scores of GPCRHM* and Pfam* using a linear weighting.

References
[3] Eck Institute for Global Health, University of Notre Dame

Images:
- Diagram showing the process of developing GPCR classifiers.
- Graphs illustrating the performance of different classifiers on various organisms.
- Bar chart showing the number of sequences found and missed by different classifiers.

Legend:
- Not likely to be a GPCR
- Annotated as GPCRs
- Herdology of known GPCRs
- Present in the training set
- Unconfirmed putative GPCRs