# Evaluation and Development of GPCR Classifiers for Vectors

Ronald J. Nowling[1], Jenica Abrudan[2,3], Douglas A. Shoue[2,3], Badi' Abdul-Wahid[1], Mariha Wadsworth[2,3], Gwen Stayback[2,3], Frank Collins[2,1,3], Mary Ann McDowell[2,3], Jesus Izaguirre[1,3]

[1]Computer Science & Engineering, Unversity of Notre Dame [2]Biological Sciences, University of Notre Dame [3]Eck Institute for Global Health, University of Notre Dame

## Abstract

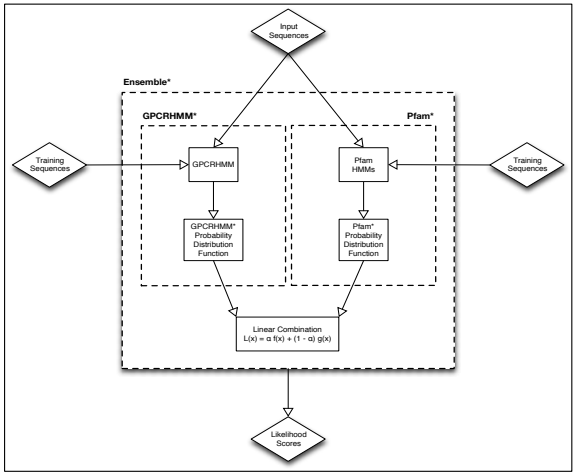
We aim to inexpensively develop insecticides for disease vectors such as mosquitoes by incorporating bioinformatics and computational biology into all aspects of the drug development process. Due to the popularity of G-Protein Coupled Receptors (GPCRs) as drug targets, a GPCR classifier that perform 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 (GPCRHMM [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*

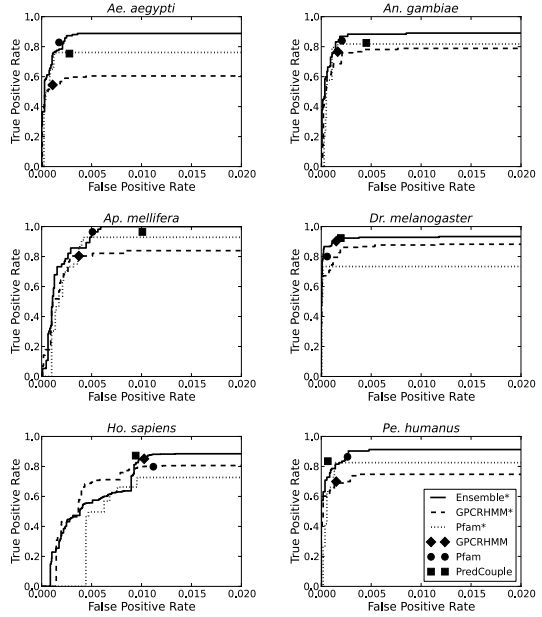
## Methods

We developed three novel classifiers (GPCRHMM*, Pfam*, and Ensemble*). GPCRHMM converts the global scores from GPCRHMM 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 GPCRHMM* and Pfam* using a linear weighting.




## Results

Ensemble*, GPCRHMM, PredCouple, GPCRHMM*, 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.




| Species | Number of Sequences Found / Missed | | | |
|---|---|---|---|---|
| | GPCRHMM | Pfam | PredCouple | Ensemble* |
| *Ae. aegypti* (134) | 73 / 61 | 111 / 23 | 101 / 33 | 122 / 12 |
| *An. gambiae* (137) | 105 / 32 | 115 / 22 | 113 / 24 | 122 / 15 |
| *Ap. mellifera* (56) | 45 / 11 | 54 / 2 | 54 / 2 | 56 / 0 |
| *Dr. melanogaster* (195) | 176 / 19 | 156 / 39 | 180 / 15 | 185 / 10 |
| *Ho. sapiens* (892) | 759 / 133 | 712 / 180 | 778 / 114 | 807 / 85 |
| *Pe. humanus* (103) | 72 / 31 | 89 / 14 | 86 / 17 | 95 / 8 |
| Vectors (374) | 250 / 124 | 315 / 59 | 300 / 74 | 339 / 35 |
| Total (1517) | 1230 / 287 | 1237 / 280 | 1312 / 205 | 1387 / 130 |

## References

[1] Wistrand M, et al. 2006. Protein Sci 15(3): 509-21.
[2] Sgourakis N, et al. 2005. Bioinformatics 21(22):4101-4106.

## 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.