

# The VectorBase Newsletter

## Issue 8 – February 2010



### Driving Biological Projects

VectorBase invites proposals for Driving Biological Projects (DBPs) that use high-throughput experimental technologies to functionally characterize the genome, proteome or metabolome of vector species. We invite letters of intent (submission deadline 28th February 2010) with selected full proposals by 31st March. Full details are available from the [VectorBase website](http://vectorbase.org) and queries and proposals should be emailed to [vector@nd.edu](mailto:vector@nd.edu).

### Call for white papers

The Arthropod Genomics Consortium has launched a website to aid coordination between researchers with a view of writing white papers and fostering future collaborations. You can find the site at <http://arthropodgenomes.org>

One recent development is the BGI proposal to sequence 1000 plant and animal genomes and are inviting white paper submission for genomes of interest. More details are available on the arthropodgenomes site or from the BGI.

### Data updates:

#### *Culex quinquefasciatus*

*Culex pipiens quinquefasciatus* was recently considered as a species in itself and was renamed *Culex quinquefasciatus*. VectorBase has updated the name of this organism to follow the NCBI taxonomy. There is no change to the underlying data.

#### *Ixodes scapularis*

The Xref data for *Ixodes scapularis* have been updated linking predictions to various external databases such as RefSeq, UniProt, EntrezGene, etc.,

#### Mosquito RNA-Seq data

RNA-Seq data for *Anopheles gambiae* and *Aedes aegypti* (SRA experiments SRX014416 and SRX014417) have been aligned back to the reference genomes and can be displayed using the EBI DAS server. You can find details of how to attach DAS sources to your browser in the tutorial section of the website. We invite groups producing next-gen transcriptomics data to contact us for help in displaying their data through VectorBase.

### Tools:

VectorBase has various ways of accessing and mining its data. Most are in the 'Tools' and the 'Retrieve data' sections but here is a reminder of what we offer.

#### Accessing the data

**Search box.** The most commonly used way to access the data is by using the search box, to search for genes by their identifiers or their function, in any of the VectorBase organisms.

Specific data sets like Ontologies and Control Vocabulary, Insecticide Resistance and Expression Data, are searchable by independent interfaces.

**BioMart** is the tool by excellence if what you want is a list of genes, corresponding to certain criteria (e.g. on chromosome 2 of *Anopheles*), with various information (such as positions, function, homology data), in a format easy to parse automatically.

**Download.** Most of the data are also available for download (sequences, expression). If you can't find what you want, ask us and we will be happy to give you the data. Warning: it might be as a raw MySQL database!

#### Mining the data

**BLAST** is the program that allows you to compare a protein or DNA sequence to the genome, the gene set or the protein set of all the VectorBase organisms. We recently added a functionality for blasting against individual EST libraries.

**ClustalW** produces a multiple alignment from a set of sequences given by the user. The result can also be used as input for HMMER.

**HMMER** generates an HMM from a multiple alignment. A useful tool if you want to search a genome for a degenerated sequence (e.g. representing a gene family) rather than from a plain sequence.

#### Twitter

For up to the minute news on VectorBase, follow us on Twitter: <http://twitter.com/vectorbase>

**The VectorBase Editors.**  
[news@vectorbase.org](mailto:news@vectorbase.org)