

The VectorBase Newsletter

Issue 2 – January 2008



New data

- ***Culex pipiens***

The genome of the mosquito [Culex pipiens](#)¹, vector of the West Nile Virus, has been sequenced and annotated and is now available at VectorBase.

It complements the existing genomes of *Anopheles gambiae* and *Aedes aegypti*, completing the "mosquito triad", with one representative in each of the three main mosquito families being sequenced (Anopheline, Aedine and Culicine).

The genome has been sequenced jointly by The Broad Institute and J. Craig Venter Institute (JCVI), and funded by NIAID. The shotgun sequencing was completed at the end of June 2006 and a first assembly has been produced by the Broad Institute using their whole genome assembly package ARACHNE, consisting of 3,171 supercontigs, CpipJ1.

The 540 Mb assembly has been jointly annotated by the Broad Institute, JCVI and VectorBase. A total of 20,316 protein-coding genes were identified. Fasta files of the gene, transcript and protein sequences are available in the download section of VectorBase and the genome can be browse via our genome browser, with comparative data linking the three mosquito genomes.

Come and have a look at our new mosquito!

- ***Micro-array probe drug resistance***

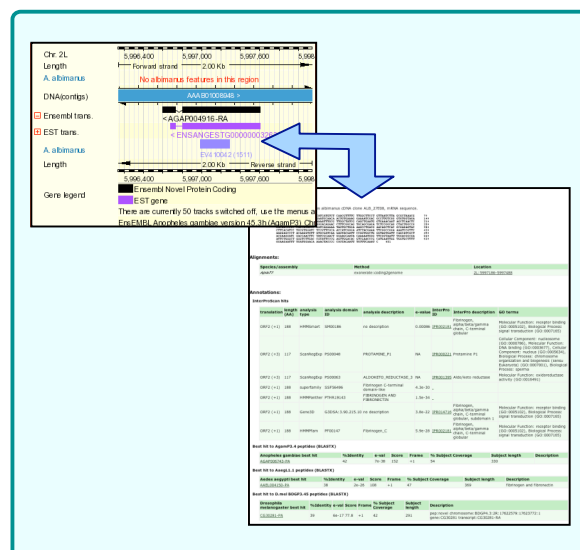
A new detox chip has been designed by the Dr. Hilary Ranson and Dr. Clare Strode, from the Liverpool School of Tropical Medicine, for a set of *Aedes* genes, including cytochrome P450's, glutathione S-transferase (GSTs) and carboxylesterases. The probes have been mapped to the *Aedes* genome and are accessible on the genome browser and in the Base expression database.

New tools

- ***GDAV***

As a result of feedback from within the community, we recognised a need for a method to disseminate annotation for species with data (e.g. EST sets) that lack genomic sequence. To meet this need, VectorBase has developed [GDAV](#)² (Genome De-linked Annotation Viewer) - a lightweight web application that provides display and search facilities for sequence and text-based annotation. GDAV also includes a DAS server so that, if the data can be mapped to one or more VectorBase genomes, these alignments can be viewed in the VectorBase genome browser.

The first use of GDAV has been to publish the [annotation](#)³ of around 5,000 *Anopheles albimanus* ESTs at the Institute of Public Health, Mexico. This GDAV instance contains InterPro annotations and alignments to *A. gambiae*, *A. aegypti* and *D. melanogaster*.



- **HMMER tools**

Providing multiple sequence alignments, the HMMER tools (HMMER Build & HMMER Search) create a probabilistic model that is then used to search the VectorBase databases. HMMER is similar to BLAST but is less likely to be affected by noise due to conservative or semi-conservative substitutions.

Look for HMMER under the Tools section at [VectorBase](http://www.vectorbase.org)⁴!

Coming soon

The *A. gambiae* and *A. aegypti* gene set updates are expected during the first quarter of 2008.

An assembly of the *Ixodes scapularis* genome has recently been completed and submitted to GenBank, and should be publicly available very soon.

We are working on generating a new Blast interface, allowing blasting against individual EST libraries or quickly selecting databases based on phylogeny.

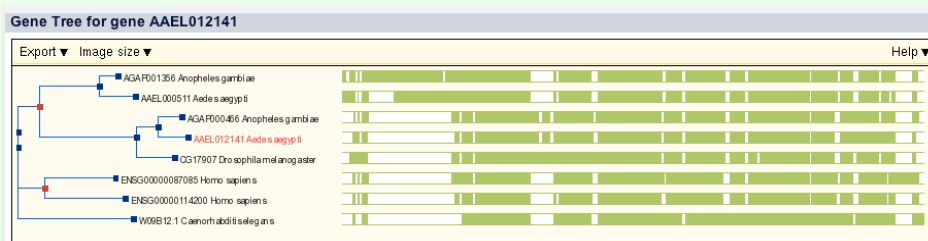
To (un)subscribe to the newsletter, contact newsletter@vectorbase.org

The VectorBase Editors.

newsletter@vectorbase.org

VectorBase top tip

Did you know that you could visualize the Gene Tree between a gene and its homologs? It is reachable from the left-hand tab of the geneview pages, under 'gene tree info.'



The Gene Trees represent the evolutionary history of gene families. Red squares represent duplications nodes, blue squares represent speciation nodes. Clicking on an internal node gives a pop-up-menu with indications on its taxonomy and clicking on a gene gives its gene and protein IDs (linking to the cognate page), and various information. The green bars at the right of the tree are a schematic representation of the multiple alignment of the peptides made using MUSCLE. Full boxes indicate matches/mismatches, open boxes indicate gaps in the alignment.

- ¹ <http://cpipiens.vectorbase.org/index.php>
- ² <http://base.vectorbase.org:8080/gdav/>
- ³ http://imm.insp.mx:8008/gdav_albimanus_v1/
- ⁴ <http://www.vectorbase.org/Tools/HMMER/>