

The VectorBase Newsletter

Issue 1 – September 2007



Launch of the Newsletter

Welcome to the first issue of the VectorBase Newsletter. The newsletter highlights new data sets and tool developments available from VectorBase as well as giving advanced notice of upcoming changes. The newsletter will be sent quarterly to those subscribed to the newsletter@vectorbase.org mail list.

New Anopheles identifiers

We have changed the identifiers of genes, transcripts and proteins for *Anopheles gambiae* to bring them in line with the VectorBase nomenclature¹ rather than that of Ensembl. For those of you who are familiar with the *Drosophila* gene identifiers you'll recognise the usage of -R/P suffixes for transcripts and proteins and alternative isoforms defined by a letter (e.g. -RA, -RB, -RC etc.). A major advantage of this system is that there is now an implicit connection between a gene and its encoded transcripts and peptides as they all share the same identifier (AGAPxxxxxx).

	Ensembl	VectorBase
Genes	ENSANGGxxxxxxxxxx	AGAPxxxxxx
Transcripts	ENSANGTxxxxxxxxxx	AGAPxxxxxx-RA
Proteins	ENSANGPxxxxxxxxxx	AGAPxxxxxx-PA

Help with converting old-style identifiers² to the new set is provided through the IDHistory View or using the mapping files available for download.

New data

• *Anopheles gene set*

A new *Anopheles* gene set (AgamP3.4), was released in July containing approx. 12,000 protein-coding gene predictions. Manual appraisal of predictions on chromosome arm 2L has been completed and significant volumes of community annotation have been integrated. The VectorBase nomenclature for gene identifiers (AGAPxxxxxx) is employed throughout. Further details are available on the *Anopheles gambiae* homepage.

Please contact the VectorBase team if you have queries about this gene set. We are especially interested to hear from researchers who think we have deleted genes or made retrograde predictions for their favourite genes.

• *Tick anatomy*

An application ontology covering the anatomy of ticks (TADS) has been developed in collaboration with Prof. Dan Sonenshine (Old Dominion University, USA). The TADS ontology, as well as the Mosquito anatomy ontology (TGMA), are now compliant with CARO (Common Anatomy Reference Ontology). These two Controlled Vocabularies (CVs), together with two additional ones covering the disease transmission physiology that are now under development, provide the basis for the annotation of gene expression experiments.

Ontologies can be accessed at VectorBase via the CVSearch tool.

- **Tiling array**

Anopheles gambiae tiling array data from Halasz *et al* (2006) is available as a DAS track on the genome browser (ContigView). This shows expression levels in adult male and female mosquitoes for 76,782 36-mer probes designed against exons from an old Ensembl gene set (release 15) and 94,469 probes targeted at regular intervals where there were no predicted genes. Enable the DAS track using the "DAS Sources" menu in the "Detailed view" (click on the box for "Tiling Expression") and then zoom out all the way (200Kb) to see the inter-gene probes. The 36-mer probes were designed to be unique. The mapping to the current repeat-masked VectorBase assembly used Exonerate in an ungapped alignment mode where only exact matches are shown.

Highlight tool

- **Manual curation tool**

VectorBase will integrate community annotations into the canonical gene sets wherever possible. We have developed a submission interface for the community using a simple Excel spreadsheet form. Submitters will need to sign up for a user account and can then proceed to submit gene predictions, CV terms, citations and general comments. Submissions are moderated by VectorBase community representatives and displayed using DAS. All submitted gene curation will be integrated into the canonical set during a subsequent gene build.

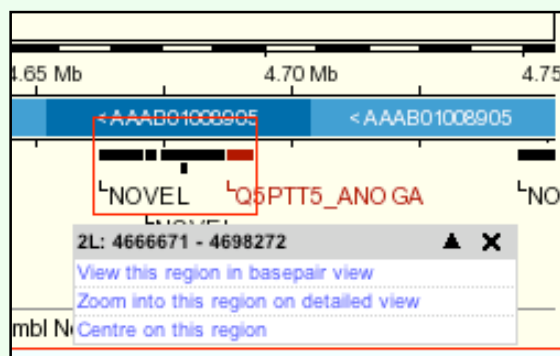
Coming soon

Two new gene sets are expected by the end of this year: the mosquito *Culex pipiens* and the body louse *Pediculus humanus*.

The ontology for insecticide resistance is almost complete and will soon be available in VectorBase via the CVSearch³ tool.

VectorBase top tip

Did you know you could use the mouse to draw a rectangular region on the genome browser windows and then zoom into that exact region. This is very useful when you want to focus on a particular feature (gene, exon, microarray probe, etc). It works in all the graphical display panels of ContigView.



The VectorBase Editors

newsletter@vectorbase.org

¹ http://www.vectorbase.org/Help/Gene_nomenclature

² http://www.vectorbase.org/Help/AgamP3.4/New_IDs_for_Anopheles

³ <http://www.vectorbase.org/Search/CVSearch>