

Annotation of vector genomes in Ensembl/VectorBase – recent developments

(Abstract for EMBO workshop, July 2005, Kolymbari, Crete)

Martin Hammond & Ewan Birney

European Bioinformatics Institute, Hinxton, Cambridge, CB10 1SD,
UK (email: mhammond@ebi.ac.uk)

The *Anopheles gambiae* genome sequence and its annotation are currently curated by the Ensembl group as part of the VectorBase project. This presentation will summarize recent developments in the annotation and presentation of the *A. gambiae* genome, and discuss plans for Ensembl contributions to other vector genome projects.

An improved assembly (AgamP3) of the *A. gambiae* scaffolds into chromosomes has been prepared (primarily by the Collins group at Notre Dame). This assembly is being made available through Ensembl and VectorBase, and provides placement of additional scaffolds on chromosomes and identification of scaffolds that may represent alternative assemblies of the same region.

http://www.ensembl.org/Anopheles_gambiae/

<http://www.vectorbase.org/>

Ensembl is producing a new genome-wide set of gene predictions based on the new assembly. The majority of the gene models are similar to those in the previous release, and identifiers can be tracked between releases. Ensembl's automatic gene building system has been improved so as that it can find more complete coding sequences. The new gene set also benefits from gene models curated manually by the Harvard FlyBase group, and from sequences submitted by researchers directly to the *A. gambiae* gene symbol database.

Other developments in the Ensembl annotation will also be discussed, including improved presentations of orthologues across a range of species, genome-genome comparisons with other insects, information about transposons, and information about the likely effects of SNPs. Options for presenting your own data sets within the Ensembl gene browser will also be described.

As genome sequence becomes available for more vector species, Ensembl plans to contribute to their annotation as part of VectorBase. The current status of this process will be reviewed.

Acknowledgments:

We would like to thank the other members of the Ensembl team (at EBI & at the Sanger Institute), the other groups in the VectorBase project (at Notre Dame, Harvard, IMMB & EMBL), and members of the *Anopheles* research community who have made their data available.