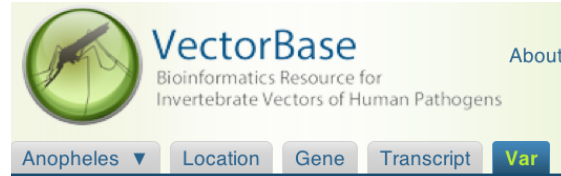


## Browsing genomes I



Gene-based displays	
Gene summary	
Splice variants (1)	
Transcript comparison	
Supporting evidence	
Sequence	
External references	
Regulation	
Gene Expression	
Reporters (20)	
Expression	
Pathways	
PubMed	
Ontology	
GO: molecular function (1)	
Comparative Genomics	
Genomic alignments (image)	
Genomic alignments (text)	

**Gene tab:** from here you can access

- Gene data
- Splice variants
- Expression data
- Comparative genomic data (orthologs, paralogs, gene tree)
- Genetic variations (if available)
- And more ...

Transcript-based displays	
Transcript summary	
Supporting evidence (3)	
Sequence	
Exons (2)	
cDNA	
Protein	
External References	
General identifiers (80)	
Oligo probes (20)	
Ontology	
GO: molecular function (1)	
Genetic Variation	
Variation table	
Variation image	
Population comparison	
Comparison image	
Protein Information	

**Transcript tab:** from here you can access

- Transcript data: summary, supporting evidence
- Sequence
- Link to external databases
- Genetic variations (if available)
- Protein data: summary, domains, variations
- And more ...

Variation displays	
Explore this variation	
Genomic context	
Genes and regulation (4)	
Flanking sequence	
Population genetics	
Individual genotypes (4)	
Linkage disequilibrium	
Phenotype Data	
Phylogenetic Context (1)	
Citations	
External Data	

**Variation tab:** from here you can access

- Variation data
- Genomic context
- Population genetics
- Individual genotypes
- And more ...

The following exercises explore several of the many access points of the gene, transcript and variation tabs.

1. Go to VectorBase search box and type one of the four *A. aegypti* gene IDs below.



Click on “Genome” or in the “top hit” in the results, both links are the same and will take you to the Genome Browser.

The image shows the search results page on the VectorBase website. At the top, there is a 'Search' header. Below it, a search bar contains 'AAEL007601' and a 'GO' button. Under the search bar, there is a link to 'Advanced Search'. The main content area is divided into two columns. The left column is titled 'Filter Results' and contains a table with the following data:

Domain	Hits
Expression	17
Comparative	1
Genome	1
Transcriptome	1

The 'Genome' row is highlighted in yellow. The right column is titled 'Search results' and shows 'Showing 1 to 20 of 20 results found.' Below this, the first result is highlighted in a white box with a blue border. It contains the text 'AAEL007601' and 'Description: trypsin'. Below this, the second result is shown, with the text 'VBGT00740000024168' and 'Description: Protein compara gene tree containing 2'.

Repeat the same steps for the other three genes, each on in a different window or tab.

- AAEL007601
  - AAEL013707
  - AAEL013714
  - AAEL013715
2. According to their metadata what is their putative function?
  3. Where are they located? Give their supercontig.
  4. Where do you click to find out their number of exons?
  5. Where do you click to find out their nucleotide and protein sequence? Is there an option to download these sequences or you can only “copy and paste”?

6. Where do you click to find out if a gene has liked publications?
7. Which genes have microarray evidence? **Hint:** Click on "Gene Expression/Reporters".
  - Follow the link to VectorBase Expression Browser page.
  - Which genes show differential expression?
  - Under which experiments?
8. Click on "External references/Oligo probes". How is these data different from the one about microarrays in the previous question?
9. Proteins.
  - Click on "Protein information/Protein summary". What can you use the protein domains and statistics for?
  - Click on "Protein information/Domains & features? What is the relationship between this and the previous page? Follow one or two links to the InterPro page.
10. Like in the first exercise, use VectorBase search and go to the genome browser for these three genes, open each gene in a different window or tab :
  - AAEL001674
  - AAEL013284
  - AAEL007432
11. According to their metadata what is their putative function?
12. Are these genes paralogs? If yes, how many paralog genes there are in total?
13. Click on "Comparative genomics" and later on "Genomic alignments/image"
  - How many options or species are there to select in this drop down menu?
  - What is the gene ID of the aligned genes?
  - Are these genes orthologs?

14. How many more orthologs each of these genes has?
15. For gene AAEL001674 click on "Ontology/Ontology table".
- Do these ontology terms agree with the putative function of these genes?
  - Click on "Biological process/GO:0006508". Follow the link to EMBL-EBI page. What is this term definition?
16. Like in the first exercise, use VectorBase search and go to the genome browser for the gene AGAP004786
- What is this gene putative function based on its metadata?
  - Click on Pathway. Click on the top KEGG ID. Follow this link to KEGG. This gene is a member of which pathway or pathways?
17. Click on "help" and understand the notation of KEGG pathway diagrams:
- What do the rectangles with numbers represent?
  - What do the circles represent?
  - What do the colors mean?
  - Hover over the two orange boxes of the pathway? What information are these showing?
18. Click on "Pathway entry". What information can you obtain in this page?
19. Click on "show description". Was this information contained in the previous link?
20. This is a very conserved pathway. Do you believe that some of these enzymes are not present in *A. gambiae*? What are some possibilities?
21. Compare completeness of this pathway with *Anopheles gambiae* vs. *Aedes aegypti*. What are the differences and similarities?
22. You can use BioMart to ask about KEGG attributes of your genes of interest.
- Go to the TOOLS tab and select "BioMart".
  - Chose Database: VectorBase genes
  - Chose dataset: *Anopheles gambiae*
  - **Filters:** GENE / Select "ID list limit" / Type "AGAP004786"
  - **Attributes:** EXTERNAL / Select "KEGG ID" and "KEGG description"
  - Click on count

- Click on results

Dataset
Anopheles gambiae genes (AgamP3 (AgamP3.7))
Filters
Gene stable ID(s): [ID-list specified]
Attributes
Gene stable ID
Transcript stable ID
KEGG ID
KEGG description

**Note:** If you are not yet familiar with BioMart, we suggest you follow its tutorial and practice exercises following this link: [www.vectorbase.org/tutorials/biomart](http://www.vectorbase.org/tutorials/biomart) .

23. Click on the KEGG ID “aga00010”. Where is this link taking you?

24. You can repeat this same BioMart KEGG search with one or multiple genes simultaneously. Try these four:

- AAEL007599
- AAEL006563
- AAEL005188
- AAEL008876

25. Like in the first exercise, use VectorBase search and go to the genome browser for these two genes, open each gene in a different window or tab :

- AAEL001806
- AAEL002467

26. Use VectorBase genome browser to give a putative function to these two genes.

27. Like in the first exercise, se VectorBase search and go to the genome browser for this gene:

- AGAP004707

28. Once in the genome browser:

- Click on transcript –RA (table in the center of the page.
- Click on “Variation table” (link in the left hand side menu)
- Click on “show” for missense (or non-synonymous) variant (table called: Summary of variation consequences)
- Click on the variation link ID “2L.2422652” (table called: Missense variant consequences)

- Click on the “Individual genotypes” link (link in the left hand side menu)
- Click on “show” and the “plus sign” to display data (table called: summary of genotypes by population)
- What is this page showing you? Genotypes per individual!
- Take note of this sample ID: **AD0026-C**.
- If you click on each individual, to which VectorBase page are you going? Notice this page has three headings: sample, assay and project. What is the icon/link that will take you back to the genome browser?

29. Go to VectorBase search box and type **AD0026-C**. Go to the population biology browser, PopBio. What information is this page telling you?

30. Go to the tutorials page [www.vectorbase.org](http://www.vectorbase.org) and click on “Variation data: How to visualize SNPs?”. Try to follow and replicate this tutorial example.

**Note:** Remember that answers may change because VectorBase its updated every two months ([www.vectorbase.org/releases](http://www.vectorbase.org/releases)).