

Available data types

Gloria I. Giraldo-Calderón
September 2017




VectorBase

Bioinformatics Resource for Invertebrate Vectors of Human Pathogens

Outline

Data is divided in five categories:

- **Genomes:** 38 species (40 genomes)
- **Transcripts and transcriptomes:** +40 species
- **Proteins and proteomes:** 2 species
- **Mitochondrial sequences:** *Aedes* 96 species, *Anopheles* 334, *Culex* 200 and Order Ixodida 384
- **Population biology:** +100 species



VectorBase data is not limited to species with a genome, it includes many more

Data can be browsed, queried and download:

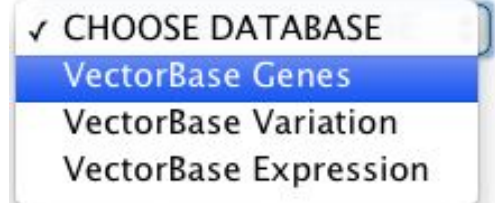
- For big data sets or complete genomes
- For a few genes or small data sets

How to browse, query and download data for big datasets or complete genomes?

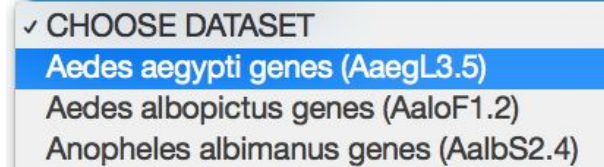


1. Select “BioMart”.

2. Select the database



3. Select the dataset



4. Select the the appropriate “Filters” and “Attributes” for your query:



Dataset
Aedes aegypti genes (AaegL3.5)
Filters
[None selected]
Attributes
Gene stable ID
Transcript stable ID

How to browse, query and download data for a few genes or small data sets?

- Search
- Tools
- Data (two access points: navigation tab and home page icons)

The screenshot shows the VectorBase website interface. At the top left is the VectorBase logo, a green circle containing a mosquito, with the text "VectorBase" and "Bioinformatics Resource for Invertebrate Vectors of Human Pathogens" below it. To the right is a search bar with the placeholder text "Enter search terms" and a "GO" button. Below the search bar is a "LOGIN" button. A navigation bar contains tabs for "ABOUT", "ORGANISMS", "DOWNLOADS", "TOOLS", "DATA", "HELP", "COMMUNITY", and "CONTACT US". The "DATA" tab is highlighted. Below the navigation bar is a "Welcome to VectorBase!" section with a paragraph describing the resource. To the right of this section is a sidebar with a "Want to see your BLAST, ClustalW and HMMer jobs?" section, a "POPULAR ORGANISMS" section featuring images of *Aedes aegypti*, *Anopheles gambiae*, and *Culex quinquefasciatus*, and a "RECENT ADDITIONS" section. The main content area below the welcome message features a "DATA" section with five icons representing Genomes, Transcripts & Transcriptomes, Proteins & Proteomes, Mitochondrial Sequences, and Population Biology. At the bottom is a "TOOLS & RESOURCES" section with a "Pause" button.

VectorBase
Bioinformatics Resource for Invertebrate Vectors of Human Pathogens

Enter search terms **GO**
Advanced Search

LOGIN

ABOUT ORGANISMS DOWNLOADS TOOLS DATA HELP COMMUNITY CONTACT US

Welcome to VectorBase!

VectorBase is an NIAID Bioinformatics Resource Center dedicated to providing data to the scientific community for Invertebrate Vectors of Human Pathogens. We aim to provide a forum for the discussion and distribution of news and information relevant to invertebrate vectors, as well as access to tools to facilitate the querying and analysis of the data sets presented on this site.

DATA

Genomes
Transcripts & Transcriptomes
Proteins & Proteomes
Mitochondrial Sequences
Population Biology

TOOLS & RESOURCES
Pause

Want to see your BLAST, ClustalW and HMMer jobs?
Log in or Register here.

POPULAR ORGANISMS

Aedes aegypti *Anopheles gambiae* *Culex quinquefasciatus*

RECENT ADDITIONS

Search

1. Type an asterisk to find “all” available data, or type a specific query such as a gene ID (*e.g.*, AAEL004325). Click GO.



Advanced Search

Search

Enter terms

► Advanced Search

Filter Results

Domain

Expression

Ontology

Transcriptome

Genome

2. Select a Domain,
e.g., Genome.

3. Select a subdomain
and/or species

Filter Results

Domain (Reset Filter)

Genome

Sub-domain

Gene

Mitochondrial genome

Species

Ixodes scapularis

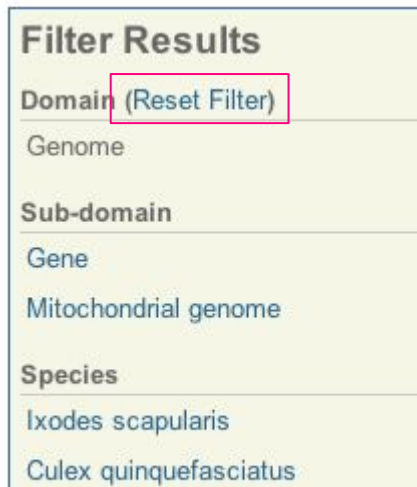
Culex quinquefasciatus

For more details go to the Search video tutorial:
www.vectorbase.org/tutorials/search

Search



A search interface with a light green background. It features a white text input field with a small asterisk icon on the left. To the right of the input field is a blue circular button with the text "GO" in white. Below the input field, the text "Advanced Search" is displayed in a smaller, dark font.



A sidebar titled "Filter Results" with a light yellow background. It contains three sections: "Domain" with a "(Reset Filter)" link, "Sub-domain", and "Species". Each section lists several options.

Filter Results
Domain (Reset Filter)
Genome
Sub-domain
Gene
Mitochondrial genome
Species
Ixodes scapularis
Culex quinquefasciatus



A sidebar titled "Search" with a light yellow background. It features a search input field with the placeholder text "Enter terms" and an asterisk icon. Below the input field is a link labeled "Advanced Search". Below this is another section titled "Filter Results" with a list of options.

Search
Enter terms *
Advanced Search
Filter Results
Domain
Expression
Ontology
Transcriptome
Genome

Note: At any point you can “Reset Filter” or switch to “Advanced Search”

VectorBase genomes

38 species (corresponding to 40 genomes)

Arthropoda

Insecta

- **Diptera**: *Aedes* (2 spp), *Anopheles* (19*), *Culex quinquefasciatus*, *Glossina* (6), *Lutzomyia longipalpis*, *Musca domestica*, *Phlebotomus papatasi*, *Stomoxys calcitrans*
- **Hemiptera**: *Cimex lectularius* and *Rhodnius prolixus*
- **Phthiraptera**: *Pediculus humanus*

Acari

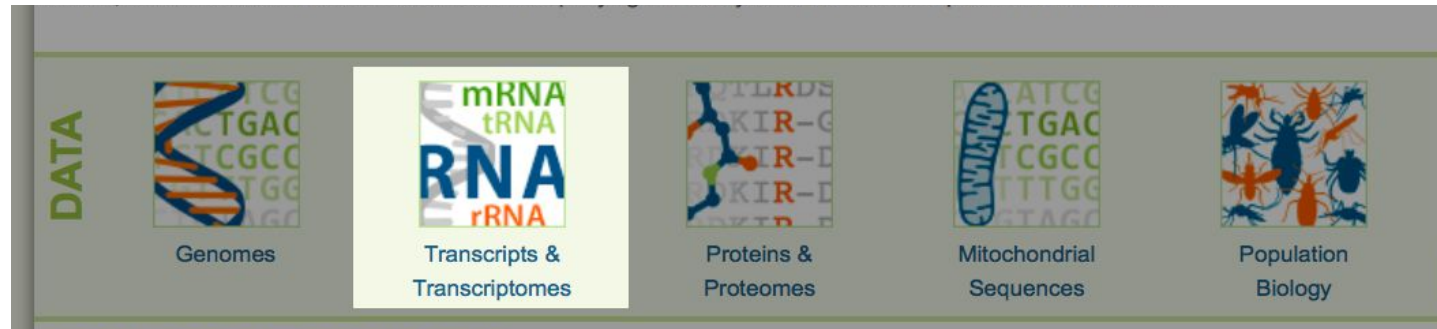
- *Ixodes scapularis* and *Sarcoptes scabiei*

Mollusca

- *Biomphalaria glabrata*

* *Anopheles*: 19 spp and 21 genomes

Data: Transcripts and Transcriptomes

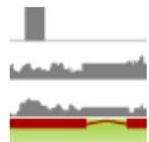


Transcript & Transcriptome Resources

VectorBase currently offers four transcriptome-related resources



Transcript sequence files available on our [download page](#) for organisms with sequenced genomes and annotated gene sets



RNAseq data aligned to and visualised with genomic sequence.



Differential transcriptomics, a.k.a. [gene expression profiling](#), mostly from microarray data at present



Annotated, assembled EST and/or RNAseq transcript sequences.

The Expression Browser

10 species

1. Type the gene ID for a gene of interest →

Browse:

- Experiments
- Microarrays and Genesets
- Hybridisations and Sequencing Runs
- Samples

Go directly to:

Gene or gene symbol:

e.g.: AGAP001111

Probe:

e.g.: NAP1-P97-D-01

Experiment:







-- select one --

Microarray or geneset:

-- select one --

2. In the “Expression Summary” choose the experiment(s) under which you want to analyse the gene differential expression (or absence of it).

Expression summary:

Experiment	P-value	Test	Experimental factor	Summary
Adult tissues (Baker et al., 2011)  Microarray experiment info  Plots and data	0.0	ANOVA	Organism part and sex 	Significant differential expression ↑ Malpighian tubules: male ↓ male accessory gland: male details
MR4 cell-lines (AVCL consortium, 2014)  RNA-Seq experiment info  Plots and data	4.4e-12	ANOVA	Strain or line 	Significant differential expression ↑ MRA-920 L35 ↓ MRA-918 4a-3A details

RNAseq data 30 species

In the RNAseq page you can find brief directions on how to use the data. For detailed step by step instructions, links to the tutorial page are provided.

RNA-Seq tracks



VectorBase has aligned RNA-Seq data experiments against most of its genomes, using data from the [SRA](#). They are available in track hub format and can be displayed in the Genome Browser.

To find relevant tracks, use the [VectorBase website search](#). You can filter the species by using the facets on the left, or use the text search to find tracks based on accession, author name, experimental factor, etc.

For more details on how to search for RNA-Seq tracks see the [transcript data tutorial page](#).

This is how the RNAseq reads look aligned to the gene of interest



Annotated Transcriptomes

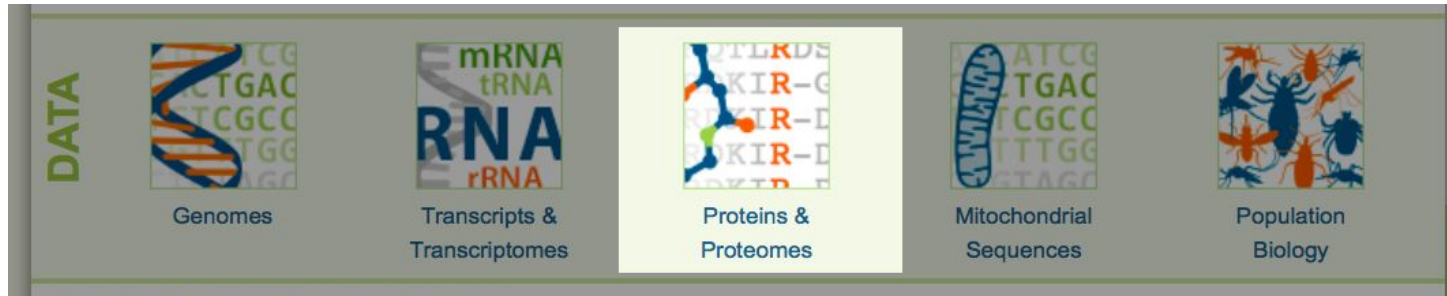
19 species

In the Annotated Transcriptomes page you can find brief directions on how to use the data using [Search](#) and [BLAST](#).

For detailed step by step instructions, links to the tutorial page are provided.

Species	Title
Aedes albopictus	Boes et al (2014) <i>Aedes albopictus</i> transcriptome shotgun assembly
Anopheles albimanus	Hittinger et al (2009) Anopheles RNA-Seq transcriptomes
Anopheles albimanus	Martinez-Barnetche et al (2012) <i>Anopheles albimanus</i> adult female transcriptome
Anopheles aquasalis	Costa-da-Silva, Marinotti, Ribeiro et al (2014) <i>Anopheles aquasalis</i> transcriptome shotgun assembly

Data: Proteins and Proteomes

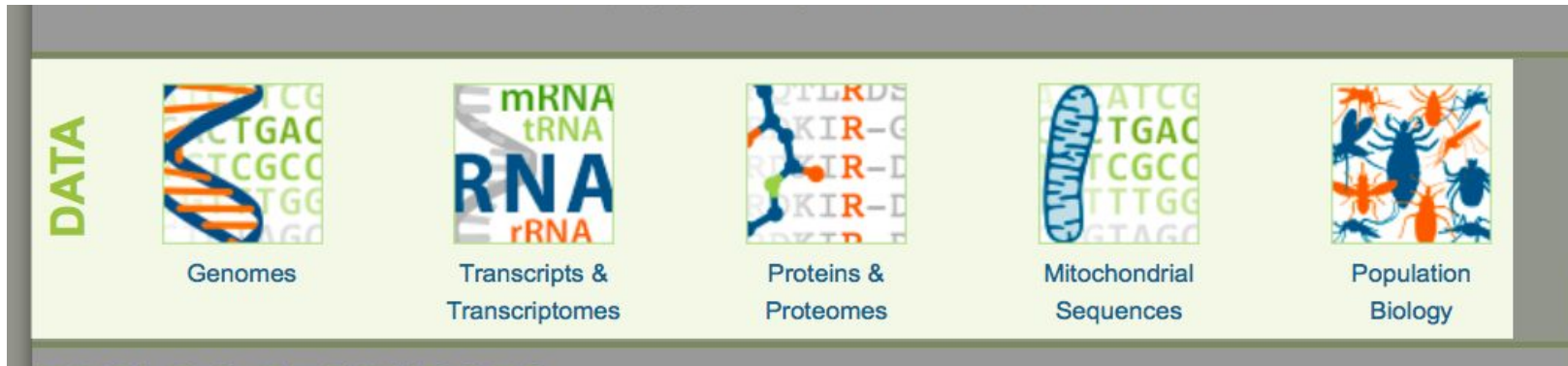


peptide sequences derived from
mass spectrometry experiments - 2 species

In the Proteomes page you can find brief directions on how to use the data using [Search](#) and the [Genome Browser](#).

For detailed step by step instructions, links to the tutorial page are provided.

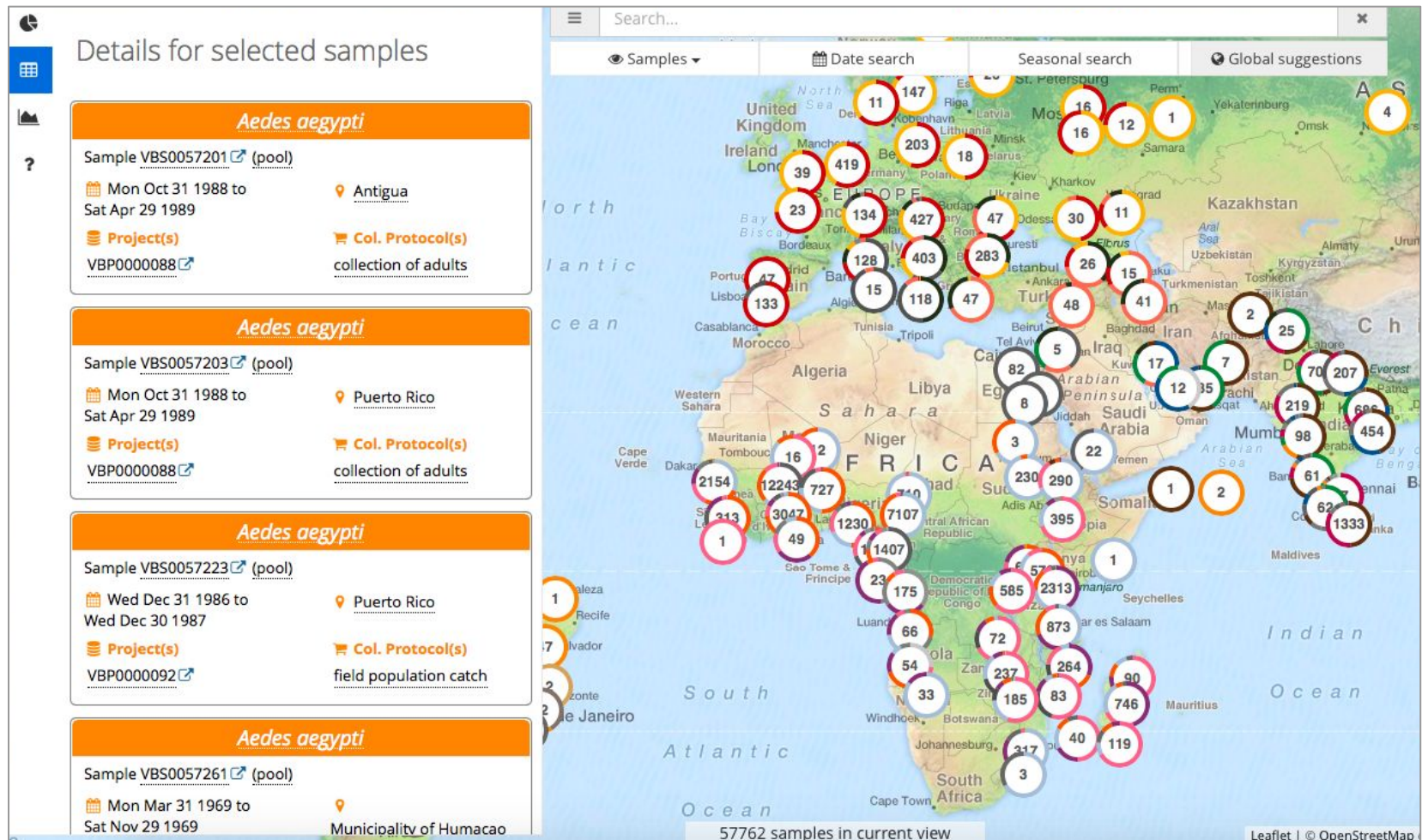
Genomes, Transcriptomes and Proteomes



- Previous slides show *in vivo* data.
- There is also *in silico* data, predicted transcriptomes and proteome sequences, based on the genome data

Population Biology (PopBio)

PopBio data includes: genotypes, insecticide resistance and other phenotypes, and field collection metadata for +100 species.



How to search for more information or help?

E-mail us at
info@vectorbase.org