

VectorBase

June 14th, 2017

Gloria I. Giraldo-Calderón, Ph.D.
VectorBase Scientific Liaison/Outreach Manager

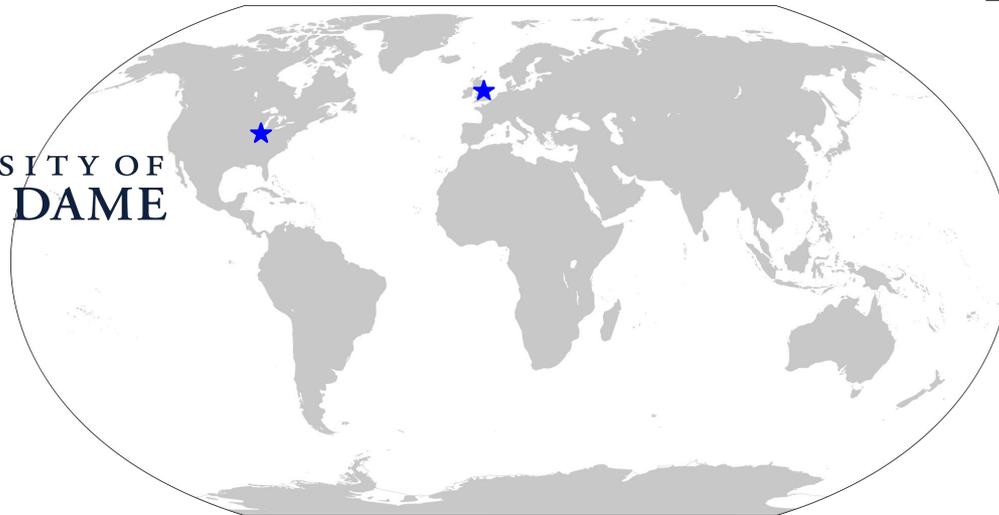


VectorBase

Bioinformatics Resource for Invertebrate Vectors of Human Pathogens

VectorBase

- Personnel: biologist, bioinformaticians, computer scientists and engineers
- Principal investigators: Dr. Frank Collins, Scott Emrich (Univ. of Notre Dame) and Dan Lawson (Imperial College London)



European Molecular Biology
Laboratory - European
Bioinformatics Institute

Imperial College
London

- Scientific Working Group, SWG: 14 senior scientists
- Scientific User Group, SUG: 10 junior scientists
- Genome leaders: 15 scientists
- Financial support: National Institutes of Health, NIH & National Institute of Allergy and Infectious Diseases, NIAID

VectorBase is

One of five funded National Institutes of Health (NIH) - National Institute of Allergy and Infectious Diseases (NIAID) Bioinformatics Resource Centers (BRCs) for infectious diseases



eukaryotic pathogens
including fungi



VectorBase

Bioinformatics Resource for
Invertebrate Vectors of Human Pathogens

invertebrate vectors of
human pathogens



Influenza Research Database
influenza



VIPR
Virus Pathogen Resource
viruses

Funded by the US, used worldwide!

Outline

Present you an overview of VectorBase data, tools and resources:

- website navigation and
- available data types



Welcome to VectorBase!

VectorBase is a National Institute of Allergy and Infectious Diseases (NIAID) Bioinformatics Resource Center (BRC) providing genomic, phenotypic and population-centric data to the scientific community for invertebrate vectors of human pathogens.

DATA



Genomes



Transcripts & Transcriptomes



Proteins & Proteomes



Mitochondrial Sequences



Population Biology

TOOLS & RESOURCES

[Resume](#)



Workshops, Meetings, and Conferences

In the upcoming months VectorBase will be attending and organizing various outreach events, these include:

- 10th Arthropod genomics symposium (AGS) & arthropod bioinformatics workshop
- Mosquito Kolymbari meeting
- Vector behavior in transmission ecology (VectorBiTE) RCN 2nd meeting
- Hands-on VectorBase and Vector Genomics Workshop
- Booth and abstract presentation at ASTMH 66th annual meeting

Follow this link to learn details about these events. Come back again soon as we will post info about a few other events in the upcoming weeks.

Want to see your BLAST, ClustalW and HMMer jobs?

Log in or Register here.

POPULAR ORGANISMS



Aedes aegypti



Anopheles gambiae



Culex quinquefasciatus

RECENT ADDITIONS



Cimex lectularius



Aedes albopictus



Sarcoptes scabiei var. canis

[All organisms](#)

LATEST NEWS

Release VB-2017-04

April 27, 2017

New population biology datasets added

April 23, 2017

Four new differential expression experiments added

April 23, 2017

[More news](#)

DID YOU KNOW?

Q: Phenotype data

A:

Every genome deserves a home



Diptera (mosquitoes -2Ae, 21An, 1Cx,
and flies - 6G, 1M, 1S, 1L, 1P)



Hemiptera (bed bug, kissing bug)



Phthiraptera (body louse)



Acari (tick and mite)

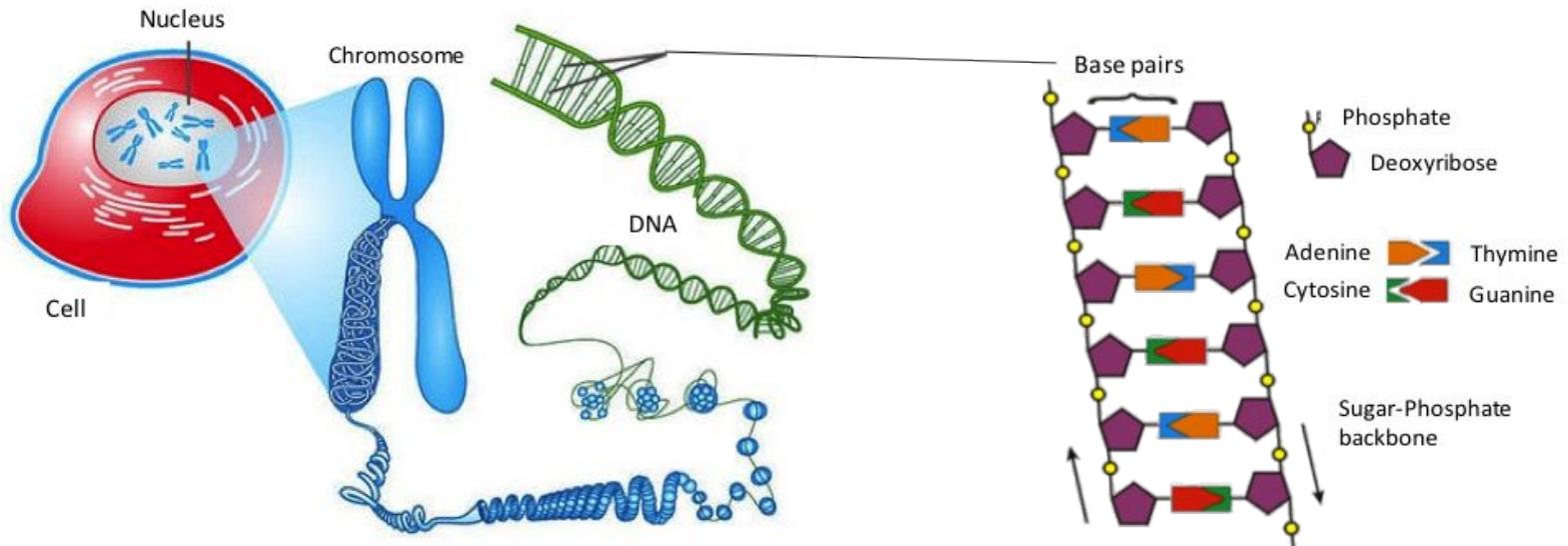


Snail (intermediate host)



**VectorBase is the home of 40 vector genomes
and other invertebrate species of interest**

What is a genome?



Copyright: University of Waikato.
All Rights Reserved. <http://sciencelearn.org.nz>

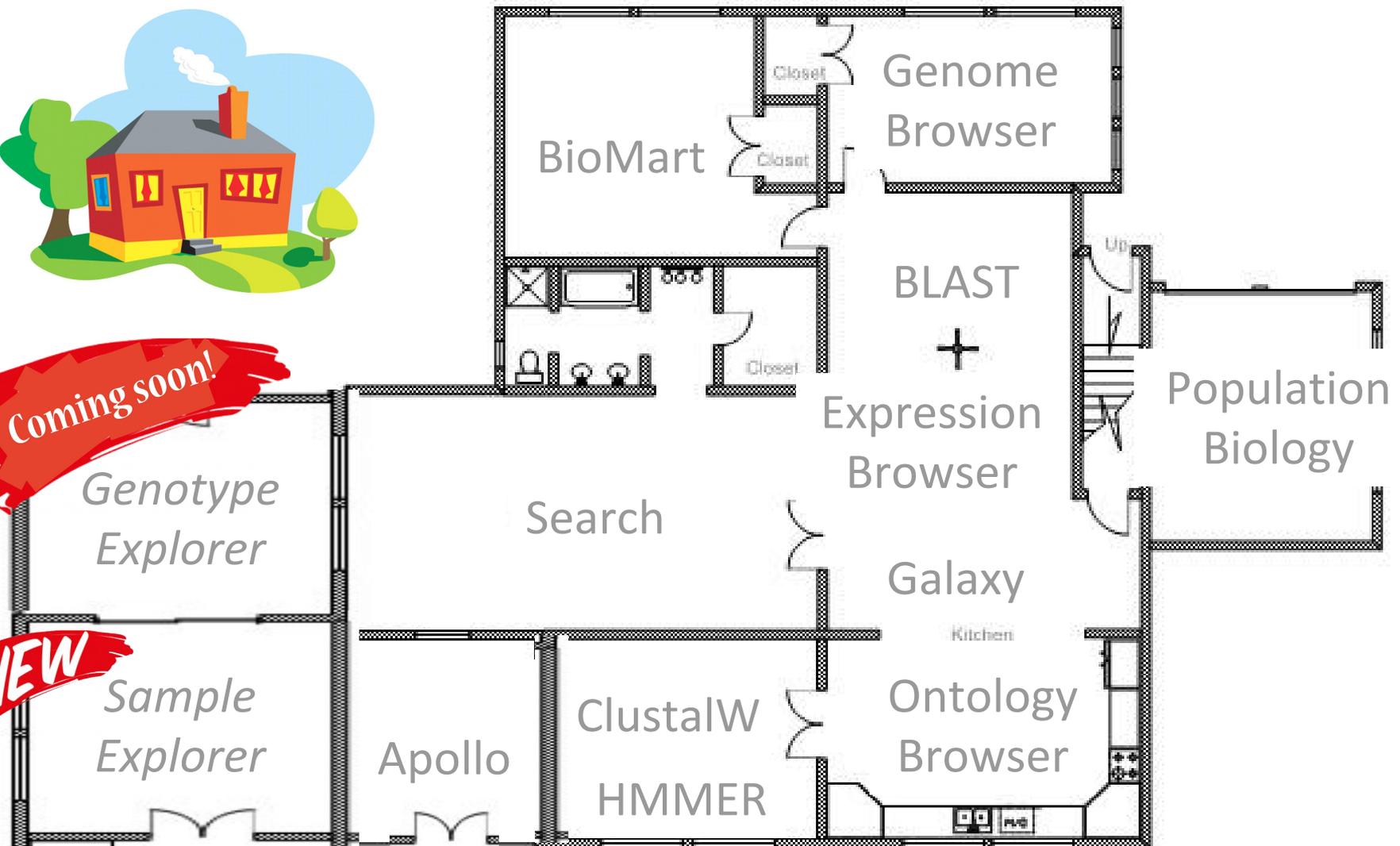
Modified from: Science Learning, Sparking Fresh Thinking, <http://goo.gl/cwWobH>
and EC Biology 1, <http://goo.gl/TEY7TZ>

Anatomy of VectorBase, the home of vector genomes



Coming soon!

NEW



Data

Partial or genome wide



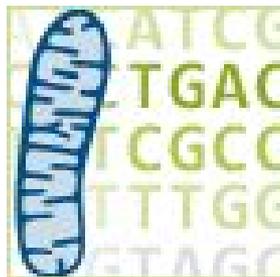
Genes and
Genomes



Transcripts and
Transcriptomes



Proteins and
Proteomes



Mitochondrial Sequences ->
DNA barcodes



Population Data
(also includes other
non-sequence data)

ACTTTGGCCCCATTCTTCGGATGGAACCGATATGTCCCAGAGGGCAACATGTCCGCCTGCGGAACTGATTACCTCACCGATACTTTGTTA
AGCCGCTCATACTTTTGGTGTACTCGATTTTCGTGTACTTCTCACCCTTTTGTGGATCATCTATTACATACGTCTTCATTATTAAGGCA
GTATCTGCTCACGAGAAGAACATGCGCGAACAAAGCGAAGAAAATGAACGTGGCTTCGCTCCGATCCTCGGAGGCTCAGAGTACCAGCACT
GAGATGAAACTCGCCAAGGTTGCCCTGGTCACGATCTCGCTGTGGTTCATGGCCTGGACTCCGTACCTAATCATCAACTACACCGGTATT
TTCAAAGCTGCGCCGATTACCCCACTGGCCACCATCTGGGGATCGCTATTCGCCAAGGCTAATGCCGTCTACAATCCGATTGTATATGGT
ATCAGTCATCCGAAATACCGTGCCGCTCTGTACAAGAAGTTCCTCGCTATCCTGCACGGATGCCACGGATGACAGCCAATCGATGGCC



TCTGGAA
GGCGCA
CTGTGG
ACCCCA
GTGATCA
ATCTGGA
CTGCGC
ACTGCC
CTCACG
GCCTCG
GCCTGG
GCCAAG
TCGTGC
TTCCTG
CAGCAG
ATGAAT
AATCTA
GGTTGC
ATTTCC

Genes and Genomes

- automatic gene predictions
- gene trees & homologous genes

40 species

CTACCCGTTTCAAATTTGGGGTCGATACATTCGGGAAGGTTTCTTACCAGTGTCTCTTTGACTACTTGACTGATGATTCCGGATACT
CGCGTCTTTGTTGGATGCATCTTCACCTGGGCTTACGTTATTCCGATGATTCTTATTTGCTACTTCTACGGACGCTTGTTTGGTCACGTA
AGCAAGCACGAACTTATGCTGAAGAATCAAGCTCGTAAAATGAACGTAGAATCGCTGGCAGCGAACCGTAACGATAAAGCTCAGTCCGTA
GAAATGCGGATTGCCCGGGCTGCCTTCACCATATTCTTCTGTTTGTGGTGCATGGACCCCGTATGCAATCGTCACGATGATCGGAGCC
TTTGGAGACAGGACACTGCTAACACCGTTCCTCACCATGATTCCAGCGGTATGTGCAAAGATAGTGTCTGTTTGGATCCATGGGTGTAC
GCCATAAGTCATCCGAAATACCGACAGGAGCTGGAACGACGACTACCCTGGATGGGTATTAAGGAGCCTGGGGATAATATTAGTACGACA
GATAGCAAAGCAAACAGTCGTTTCAGACTCGGCTCCTATGGGGCCAAACGGAACCGATTGAATGCCCACTCCCAAGTTCCCAAAAAGTTCC
GATCCCATCGGAAACACGGTTCACCAGAATGGGCTGCGGAGCTACTTCAACAATTGCCGGCGATATTC AAGACGAAGACTACGGAGGA

CAGGAAGAGAAAGCCTCAGCTTAAATGGCAGCCTTTGTGGAACTCATTTCGACGCCTGGCAGGCCGCC
GACAAGGTCCCACCGGAGATGTTGCATATGATTCACCCGCACTGGAACCAGTTCACACCGATGAACCCA
TTCGCCATCTTTGTGTTGGGAATGGTCTCGATGATCGGCAATGGCTTCGTGATGTCCATCTTCACCGGC
TCCAACCTGCTAGTGGTCAATCTAGCCTTCTCGGACTTCTGATGATGTTTACGATGGGACCGCCTATG
ACCTGGGTCTTTGGACCATTGCTTGCGAAGTCTACGCTTGCCTTGGATCGCTGTTCCGGCTGTGCCTCG
GCATTGACCGATAACAACGTCATTGTGAAGGGTCTGGCTGCCAAGCCTTTGACCAACAACGGTGCCATG
ATGGAACCGATATGTCCCAGGAAGGAAACATG
CTACGCCATCTTCGTCTACTGGACGCCACTG
GATGCGTGAACAGGCCAAGAAGATGAACGTC
CGCCCTGGTCACCATCTCGCTGTGGTTCATG
CCCCTGGCCACCATCTGGGGATCGCTGTTC
CGCTGCTTTGTACCAGAGACACCCATGGTTG
CACCACCGCCACCGAGGAGAAGGCCTAAATG
CAAGATGCTGGGCTGGAATCTACCTGCAGAC
CCTGCTGCTGGCAATGATCTATTTGCTGCTG
CTTACGCAACGGTTC AACATGTTTCGTGGTG
CTTCGCCGGACATTTGGTTGGCTTCGAAACG
CGCAGTCATCGGTACGACCGATAACCGCACC
CACCTGGCTGTGGGCAATGCCATTACGTTG

ACTTTGGCCCCATTCTTCGGATGGAACCGATATGTCCCAGAGGGCAACATGTCCGCCTGCGGAACTGATTACCTCACCGATACTTTGTTA
 AGCCGCTCATAACATTTTGGTGTACTCGATTTTCGTGTACTTCTCACCCTTTTGTGGATCATCTATTACATACGTCTTCATTATTAAGGCA
 GTATCTGCTCACGAGAAGAACATGCGCGAACAAAGCGAAGAAAATGAACGTGGCTTCGCTCCGATCCTCGGAGGCTCAGAGTACCAGCACT
 GAGATGAAACTCGCCAAGGTTGCCCTGGTCACGATCTCGCTGTGGTTCATGGCCTGGACTCCGTACCTAATCATCAACTACACCGGTATT
 TTCAAAGCTGCGCCGATTACCCCACTGGCCACCATCTG
 ATCAGTCATCCGAAATACCGTGCCGCTCTGTACAAGAA
 TCTGGAACCACAACCGTGGTGCAGGAAGAGAAAAGCCTC
 GCGCGCAACATGACCCTTGTGGACAAGGTCCCACCGGA
 CTGTGGCACTCGATCCTCGGATTCGCCATCTTTGTGTT
 ACCCCAAGCCTCCGCACCCCGTCCAACCTGCTAGTGGT
 GTGATCAACTG
 ATCTGGACCAT
 CTGCGCATTTT
 ACTGCCTGCGG
 CTCACGATCAT
 GCCTCGCTCCG
 GCCTGGACTCC
 GCCAAGGCCAA
 TCGTGCCAGGA
 TTCCTGATAAA
 CAGCAGCACTT
 ATGAATGTGTC
 AATCTAGCCAT
 GGTGCGCTGT
 ATTTCGAACC
 CTACCCGTGTT
 CGCGTCTTTGT
 AGCAAGCACGA
 GAAATGCGGATTGCCCGGCTCCCTTCAACATACTGCTTTTCTTGTGATGGAACCCGATCCCACTTCACGATGATCGGAGCC
 TTTGGAGACAGGACACTGCTAACACCGTTCCTCACCATGATTCCAGCGGTATGTGCAAAGATAGTGCTCTGTTTGGATCCATGGGTGTAC
 GCCATAAGTCATCCGAAATACCGACAGGAGCTGGAACGACGACTACCCTGGATGGGTATTAAGGAGCCTGGGGATAATATTAGTACGACA
 GATAGCAAGCAAACAGTCGTTTTCAGACTCGGCTCCTATGGGGCCAAACGGAACCGATTGAATGCCCACTCCCAAGTTCCCAAAAAGTTCC
 GATCCCATCGGAAACCACGGTTCACCAGAATGGGCTGCGGAGCTACTTCAACAATTGCCGGCGATATTCCAAGACGAAGACTACGGAGGA
 CGGCTGTGCCTCG
 CAACGGTGCCATG
 SGAAGGAAACATG
 CTGGACGCCACTG
 SAAGATGAACGTC
 SCTGTGGTTCATG
 SGGATCGCTGTTC
 ACACCCATGGTTG
 SAAGGCCTAAATG
 TCTACCTGCAGAC
 CTATTTTCGTGCTG
 CATGTTTCGTGGTG
 TGGCTTCGAAACG
 CCGATACCGCACC
 SCCATTACGTTG
 TGATTCCGGATACT
 GTTTGGTACAGTA
 AGCTCAGTCCGTA



Transcripts and Transcriptomes

- genomes ---> *in silico*
- different experimental conditions ---> *in vivo*
- microarrays and RNAseq
- partial and genome wide

Most of the 40 genome species plus others

GAAATGCGGATTGCCCGGCTCCCTTCAACATACTGCTTTTCTTGTGATGGAACCCGATCCCACTTCACGATGATCGGAGCC
 TTTGGAGACAGGACACTGCTAACACCGTTCCTCACCATGATTCCAGCGGTATGTGCAAAGATAGTGCTCTGTTTGGATCCATGGGTGTAC
 GCCATAAGTCATCCGAAATACCGACAGGAGCTGGAACGACGACTACCCTGGATGGGTATTAAGGAGCCTGGGGATAATATTAGTACGACA
 GATAGCAAGCAAACAGTCGTTTTCAGACTCGGCTCCTATGGGGCCAAACGGAACCGATTGAATGCCCACTCCCAAGTTCCCAAAAAGTTCC
 GATCCCATCGGAAACCACGGTTCACCAGAATGGGCTGCGGAGCTACTTCAACAATTGCCGGCGATATTCCAAGACGAAGACTACGGAGGA

ACTTTGGCCCCATTCTTCGGATGGAACCGATATGTCCCAGAGGGCAACATGTCCGCCTGCGGAACTGATTACCTCACCGATACTTTGTTA
 AGCCGCTCATAATTTTGGTGTACTCGATTTTCGTGTACTTCTCACCCTTTTGTGGATCATCTATTTCATAAAGGCA
 GTATCTGCTCACGAGAAGAACATGCGCGAACAAGCGAAGAAAATGAACGTGGCTTCGCTCCGATCCTCGGAAGCACT
 GAGATGAACTCGCCAAGGTTGCCCTGGTCACGATCTCGCTGTGGTTCATGGCCTGGACTCCGTACCTAATGGTATT
 TTCAAAGCTGCGCCGATTACCCCACTGGCCACCATCTGGGGATCGCTATTCGCCAAGGCTAATGCCGTCTATATGGT
 ATCAGTCATCCGAAATACCGTGCCGCTCTGTACAAGAAGTTCCTCGCTATCCTGCACGGATGCCACGGAATGGCC



Proteins and Proteomes

TCT
GGC
CTG
ACC
GTG
ATC
CTG
ACT
CTC
GCC
GCC
GCC
TCG
TTC
CAG
ATG
AAT
GGT
ATT
CTA
CGC
AGC

genomes ---> *in silico*

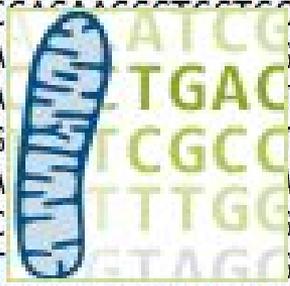
Anopheles gambiae (in vivo):

- different experimental conditions
- partial and genome wide for live organisms

Ixodes scapularis (in vivo):

- one experimental condition
- genome wide for a cell line

GAAATGCGGATTGCCCGGGCTGCCTTCACCATATTCTTCCTGTTTGTGGTGCATGGACCCCGTATGCAATCGTCACGATGATCGGAGCC
 TTTGGAGACAGGACACTGCTAACACCGTTCCTTCACCATGATTCCAGCGGTATGTGCAAAGATAGTGCTCTGTTTGGATCCATGGGTGTAC
 GCCATAAGTCATCCGAAATACCGACAGGAGCTGGAACGACGACTACCCTGGATGGGTATTAAGGAGCCTGGGGATAATATTAGTACGACA
 GATAGCAAAGCAAACAGTCGTTTCAGACTCGGCTCCTATGGGGCCAAACGGAACCGATTGAATGCCCACTCCCAAGTTCCCAAAAAGTTCC
 GATCCCATCGGAAACCACGGTTCACCAGAATGGGCTGCGGAGCTACTTCAACAATTGCCGGCGATATTC AAGACGAAGACTACGGAGGA



Mitochondrial Sequences

- 96 species *Aedes*
- 334 species of *Anopheles*
- 200 species of *Culex*
- 384 species of ticks

DNA barcodes: species identification

ACTTTGGCCCCATTCTTCGGATGGAACCGATATGTCCCAGAGGGCAACATGTCCGCCTGCGGAACTGATTACCTCACCGATACTTTGTTA
AGCCGCTCATACTTTTGGTGTACTCGATTTTCGTGTACTTCTCACCCTTTTGTGGATCATCTATTACATACGTCTTCATTATTAAGGCA
GTATCTGCTCACGAGAAGAACATGCGCGAACAAGCGAAGAAAATGAACGTGGCTTCGCTCCGATCCTCGGAGGCTCAGAGTACCAGCACT
GAGATGAAACTCGCCAAGGTTGCCCTGGTCACGATCTCGCTGTGGTTCATGGCCTGGACTCCGTACCTAATCATCAACTACACCGGTATT
TTCAAAGCTGCGCCGATTACCCCACTGGCCACCATCTGGGGATCGCTATTCGCCAAGGCTAATGCCGTCTACAATCCGATTGTATATGGT
ATCAGTCATCCGAAATACCGTGCCGCTCTGTACAAGAAGTTCCCCTCGCTATCCTGCACGGATGCCACGGATGACAGCCAATCGATGGCC
TCTGGAACGACAAGGCTGCTCCAGGAAGAGAAAAGCCTCAGCTTAAATGGCAGCCTTTGTGGAACCTCATTTTGACGCCTGGCAGGCCGCC
GGCGGCAAACAAGGTCACCAAGGTCCCACCGGAGATGTTGCATATGATTCACCCGCACTGGAACCAGTTCACACCGATGAACCCA
CTGTGGCACTCGCCATCTTTGTGTTGGGAATGGTCTCGATGATCGGCAATGGCTTCGTGATGTCCATCTTCACCGGC
ACCCCAAGTCCAACTGCTAGTGGTCAATCTAGCCTTCTCGGACTTCCTGATGATGTTTACGATGGGACCGCCTATG
GTGATCAAACCTGGGTCTTTGGACCATTGCTTGGCAAGTCTACGCTTGCCTTGGATCGCTGTTCCGGCTGTGCCTCG
ATCTGGACGTCATTGACCGATAACAACGTCATTGTGAAGGGTCTGGCTGCCAAGCCTTTGACCAACAACGGTGCCATG
CTGCGCATGTCATTGCTCTTTCTGGACTTTGGCCCCATTCTTCGGATGGAACCGATATGTCCCAGGAAGGAAACATG
ACTGCCTGCGGAACTGACTACCTGACCCAGACCTGGTTGAGCCGTTTCGTACATCATCTACGCCATCTTCGTCTACTGGACGCCACTG
CTCACGATCGATGCGTGAACAGGCCAAGAAGATGAACGTC
GCCTCGCTCGCCCTGGTCACCATCTCGCTGTGGTTCATG
GCCTGGACGCCCGCTGGCCACCATCTGGGGATCGCTGTTCC
GCCAAGGCICGCTGCTTTGTACCAGAGACACCCATGGTTG
TCGTGCCAICACCACCGCCACCGAGGAGAAGGCCTAAATG
TTCCTGATICAAGATGCTGGGCTGGAATCTACCTGCAGAC
CAGCAGCAICCTGCTGCTGGCAATGATCTATTTTCGTGCTG
ATGAATGTICTTACGCAACGGTTC AACATGTTTCGTGGTG
AATCTAGCICCTTCGCCGGACATTTGGTTGGCTTCGAAACG
GGTTGCGCICGCAGTCATCGCGTACGACCGATAACCGCACC
ATTTCGAAICACCTGGCTGTGGGCAATGCCATTACGTTG
CTACCCGTICTTTGACTACTTGACTGATGATTCCGGATACT
CGCGTCTTICTACTTCTACGGACGCTTGTGGTTCACGTA
AGCAAGCAIAGCGAACCGTAACGATAAAGCTCAGTCCGTA
GAAATGCGICCCCGTATGCAATCGTCACGATGATCGGAGCC
TTTGGAGAIAGATAGTGTCCTGTTTGGATCCATGGGTGTAC
GCCATAAGI CATCCGAAA IACCGACAGGAGC IGGAAACGALGAL IACCC IGGATGGG IAI TAAGGAGCCTGGGGATAATATTAGTACGACA
GATAGCAAGCAAACAGTCGTTTTAGACTCGGCTCCTATGGGGCCAAACGGAACCGATTGAATGCCCACTCCCAAGTTCCCAAAAAGTTCC
GATCCCATCGGAAACCACGGTTCACCAGAATGGGCTGCGGAGCTACTTCAACAATTGCCGGCGATATTCCAAGACGAAGACTACGGAGGA

ACTTTGGCCCCATTCTTCGGATGGAACCGATATGTCCAGAGGGCAACATGTCCGCCTGCGGAACTGATTACCTCACCGATACTTTGTTA
AGCCGCTGACTCGATTTTCGTGTACTTCTCACCACCTTTTGTGGATCATCTATTACATACGTCTTCATTATTAAGGCA
GTATCTGATGGCGAACAAGCGAAGAAAAATGAACGGTGGCTTCGCTCCGATCCCTCGGAGGGCTCAGAGTACCCAGCACT
GAGATG
TTCAAA
ATCAGT
TCTGGA
GGCGGC
CTGTGG
ACCCCA
GTGATC
ATCTGG
CTGCGC
ACTGCC
CTCACG
GCCTCG
GCCTGG
GCCAAG
TCGTGC
TTCCTG
CAGCAG
ATGAAT
AATCTA
GGTTGC
ATTTCC
CTACCC
CGCGTC
AGCAAG
GAAATG
TTTGGA
GCCATA
GATAGC
GATCCCATCGGAAACCACGGTTCACCAGAATGGGCTGCGGAGCTACTTCAACAATTGCCGGCGATATTCCAAGACGAAGACTACGGAGGA



Population Data

Field data and lab colonies

Non-sequence data (but could be link to sequence data):

- insecticide resistance, IR
- taxonomy, biology, ecology and behavior
- geographic distribution
- *new*: mosquito abundance

Sequence data:

- Single Nucleotide Polymorphisms, SNPs
- Genotyping
- Taxonomy

115 species

[Genomes](#)[Transcriptomes](#)[Proteomes](#)[Mitochondrial sequences](#)[Release notes](#)

Data



Genomes

Browse by Organism, Strain, [Assembly](#), or [Gene set](#)



Proteomes

Simple databases of peptide sequences, both predicted and derived from experiment



Population Biology (PopBio)

Integration of genomic, phenotypic and population data



Transcriptomes

[Transcript](#) sequence files, gene expression profiling, RNAseq data and RNAseq transcript



Mitochondrial Sequences

Mitochondrial genomes



Release Notes

List of changes and the state of current versions

VectorBase updates: Releases

The screenshot shows the VectorBase website interface. At the top, there is a navigation bar with buttons for ABOUT, ORGANISMS, DOWNLOADS, TOOLS, DATA, HELP, COMMUNITY, and CONTACT US. The 'DATA' button is highlighted, and a dropdown menu is open, listing Genomes, Transcriptomes, Proteomes, Mitochondrial sequences, and Release notes. Below the navigation bar, the breadcrumb 'Home » Data' is visible. The main content area is titled 'Data' and features several categories: Genomes (Browse by Organism, Strain, Assembly, or Gene set), Proteomes (Simple databases of peptide sequences, both predicted and derived from experiment), Population Biology (PopBio) (Integration of genomic, phenotypic and population data), Transcriptomes (Transcript sequence files, gene expression profiling, RNAseq data and RNAseq transcript), and Mitochondrial Sequences (Mitochondrial genomes). A 'Tutorials' icon is also present. A white box in the bottom right corner highlights the 'RELEASE' section, titled 'Release Notes', which is described as a 'List of changes and the state of current versions'.

Home » Data

Data

Genomes
Browse by Organism, Strain, [Assembly](#), or [Gene set](#)

Proteomes
Simple databases of peptide sequences, both predicted and derived from experiment

Population Biology (PopBio)
Integration of genomic, phenotypic and population data

Transcriptomes
[Transcript](#) sequence files, gene expression profiling, RNAseq data and RNAseq transcript

Mitochondrial Sequences
Mitochondrial genomes

RELEASE

Release Notes
List of changes and the state of current versions

Tutorials

What is search and how it is organized?



Search

Filter Results		<input type="checkbox"/> Show/Hide Category
Domain(Reset Filter)	◆	Hits ▼
Genome		5,597,199
Sub-domain	◆	Hits ▼
Genomic sequence assembly		3,737,967
Transcript		619,374
Gene		599,966
Translation		592,621
Mitochondrial genome		47,271
Species	◆	Hits ▼
Ixodes scapularis		1,007,295
Biomphalaria glabrata		785,783
Aedes albopictus		563,966

Is an efficient way to find and understand data



Its organized in:



Where is the Search box located?



The screenshot shows the top section of the VectorBase website. On the left is a circular logo featuring a mosquito. To its right is the text "VectorBase" in a large blue font, with "Bioinformatics Resource for Invertebrate Vectors of Human Pathogens" in a smaller black font below it. In the top right corner, there is a search bar with the placeholder text "Enter search terms" and a blue "GO" button. Below the search bar is the text "Advanced Search". To the right of the search bar is a green "LOGIN" button. Below these elements is a horizontal navigation menu with buttons for "ABOUT", "ORGANISMS", "DOWNLOADS", "TOOLS", "DATA", "HELP", "COMMUNITY", and "CONTACT US". Below the navigation menu is a "Welcome to VectorBase!" section with a brief description of the site's purpose. To the right of this section is a green box with the text "Want to see your BLAST, ClustalW and HMMer jobs? Log in or Register here."

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

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

Try Search using the wildcard, * to see:

- all VB data and how is organized
- help to learn familiarity with VB
- *e.g.*, data sets for a species of interest

Keywords to look for a gene

- ✓ VectorBase gene ID (use BLAST to find ID -seq homolog)
- ✓ Gene metadata (name and description) - **Caveats**
- ✓ Terms from external databases such as InterPro or Gene Ontology (GO)
- ✓ Links from the archival databases, such as GenBank, to VectorBase

What can you do with VectorBase?

Search, analyze, visualize and compare organisms genomes and various data types (called tracks)

Aedes aegypti (AeegL3) Location: supercont1.114:1,986,883-1,991,252 **Gene: RpL5** Transcript: AAEL004325-RA

Gene-based displays

- Summary
 - Splice variants
 - Transcript comparison
- Sequence
 - Secondary Structure
- Literature
- Comparative Genomics
 - Genomic alignments
 - Gene tree
 - Gene gain/loss tree
 - Orthologues
 - Paralogues
- Ontologies
 - GO: Biological process
 - GO: Molecular function
 - GO: Cellular component
- Phenotypes
- Genetic Variation
 - Variant table
 - Variant image
 - Structural variants
- Expression report
- External references
- Supporting evidence
- ID History
 - Gene history

Gene: RpL5 AAEL004325

Description 60S ribosomal protein L5 [Source:UniProtKB/Swiss-Prot;Acc:[Q1HQU2](#)]

Location [SuperContig supercont1.114: 1,986,883-1,991,252](#) forward strand.
AeegL3:CH477299.1

About this gene This gene has 1 transcript ([splice variant](#)) and [38 orthologues](#).

Transcripts [Hide transcript table](#)

Show/hide columns (1 hidden)		Filter					
Name	Transcript ID	bp	Protein	Biotype	UniProt	RefSeq	Flags
Novel	AAEL004325-RA	1250	297aa	Protein coding	Q1HQU2	XM_001648846.1 XP_001648896.1	

Summary

Name RpL5 (VB Community Annotation)

UniProtKB This gene has proteins that correspond to the following UniProtKB identifiers: [Q1HQU2](#)

Gene type Protein coding

Annotation method Genes were annotated by merging Ensembl and TIGR prediction sets ([see details](#)).

What can you do with VectorBase?

Available tracks include:

- Single Nucleotide Polymorphisms, SNPs
- RNAseq (reads), microarrays (probes)
- mass spectrometry peptides
- DNA and protein alignments

Available data types are different for each species

The species with the most resources are *Anopheles gambiae* and *Aedes aegypti*

Use case: identify patterns of differential expression for gene(s) of interest for different experimental conditions.

What can you do with VectorBase?

Download simple or complex data sets (single genes, few genes and genome wide) for one or more species

Possible queries include:

- nucleotide and protein sequences
- GO term enrichment
- genes with available transcript and protein expression

All available data is available for download.

What can you do with VectorBase?

Manually annotate genes, submit the new intron exon boundaries and their metadata (gene name and function)

- split
- delete
- merge
- create new ones

Use VectorBase predicted homology relationships, confirm or calculate new ones based on improved gene manual annotations.

What can you do with VectorBase?

Map visualization of population data:

- blood meal analyses
- KDR genotypes
- abundance
- genotypes
- microsatellite
- pathogen collection
- ...

Galaxy: Available servers for high computing tasks

- SNP calling
- RNAseq differential expression

VectorBase, 13 years and counting



Thank You!

For more information or help

E-mail us at
info@vectorbase.org

We also will like to hear *feedback* from
you about any current or new
feature or data set

Thank you!