VectorBase

June 14th, 2017

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

• 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

invertebrate vectors of human pathogens

Funded by the US, used worldwide!
Outline

Present you an overview of VectorBase data, tools and resources:
- website navigation and
- available data types
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

Anatomy of VectorBase, the home of vector genomes

Coming soon!

Genotype Explorer
Sample Explorer
Apollo
ClustalW
HMMER
Search
Galaxy
Ontology Browser
Expression Browser
BLAST
Genome Browser
BioMart
Population Biology

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)
Genes and Genomes
- automatic gene predictions
- gene trees & homologous genes

40 species
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
Proteins and Proteomes

genomes ---\(\rightarrow\) *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
Mitochondrial Sequences

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

DNA barcodes: species identification
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
Data

Genomes
Browse by Organism, Strain, Assembly, or Gene set

Transcriptomes
Transcript sequence files, gene expression profiling, RNAseq data and RNAseq transcript

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

Mitochondrial Sequences
Mitochondrial genomes

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

Release Notes
List of changes and the state of current versions
VectorBase updates: Releases
What is search and how it is organized?

Is an efficient way to find and understand data.

Its organized in:

<table>
<thead>
<tr>
<th>Species</th>
<th>Hits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ixodes scapularis</td>
<td>1,007,295</td>
</tr>
<tr>
<td>Biomphalaria glabrata</td>
<td>785,783</td>
</tr>
<tr>
<td>Aedes albopictus</td>
<td>563,966</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sub-domain</th>
<th>Hits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genomic sequence assembly</td>
<td>3,737,967</td>
</tr>
<tr>
<td>Transcript</td>
<td>619,374</td>
</tr>
<tr>
<td>Gene</td>
<td>599,966</td>
</tr>
<tr>
<td>Translation</td>
<td>592,621</td>
</tr>
<tr>
<td>Mitochondrial genome</td>
<td>47,271</td>
</tr>
</tbody>
</table>
Where is the Search box located?

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)
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!