

VectorBase,
a Resource Centre for
Invertebrate Hosts of Human Pathogens

<http://www.vectorbase.org>

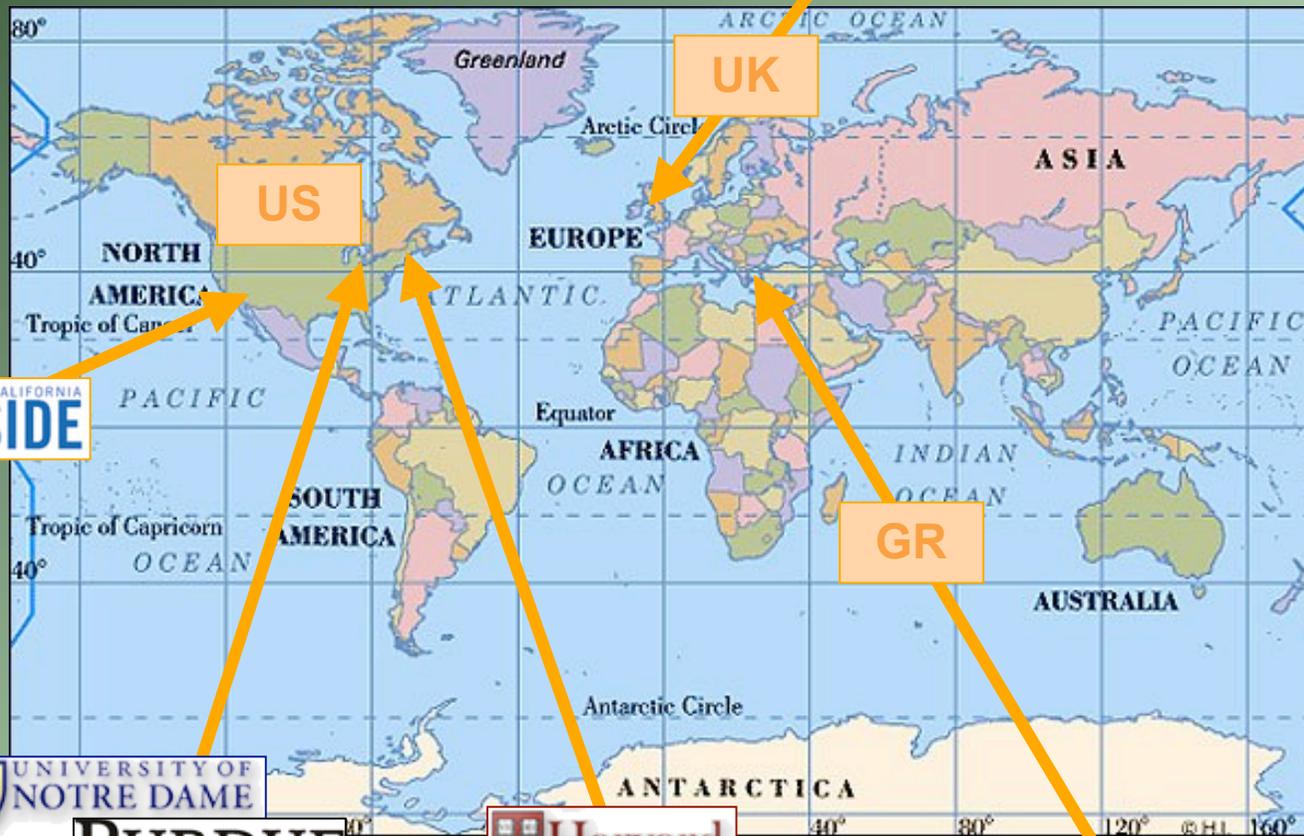
Karyn Mégy
European Bioinformatics Institute, Cambridge
UK



What is VectorBase?

- Aim
 - “A home for invertebrate vectors of human pathogens”
 - Gather and display data
- Funding
 - US NIAID (National Institute for Allergy and Infectious Diseases)
 - via its Bioinformatics Resource Centre (BRC) program

Who is VectorBase?



Oh no, not one more website...

- Browsers: Ensembl, JCVI, BROAD



- Ensembl: no plan to keep mosquitoes for long, or to add any other flying bugs
- JCVI, BROAD etc.: won't add anything if not involved

Yes, but this one ...

- VectorBase:
 - flying and biting bugs
 - one place with everything -> cross-comparison
 - long term

Store data Available to all	Generate data Automatic annotation
Display data Easy to browse, download	Manual annotation
Mine data Compare, align	Expression data
	Community Use resources
	Add data

Which organisms we don't have (will never have ... and why) ?

- We have
 - Some flying and biting bugs
 - **NOT** drosophila, bee, tribolium or bombyx

=> Only the ones that are vectors
of human diseases



Which organisms do we have?

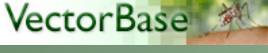
- Organisms

<i>Anopheles gambiae</i> [mosquito]	Malaria
<i>Aedes aegypti</i> [mosquito]	Yellow fever, dengue fever, chikungunya
<i>Culex pipiens</i> [mosquito]	West Nile Virus, filariasis, some encephalitis
<i>Pediculus humanus</i> [body louse]	Pediculosis (=human body infection - not head)
<i>Ixodes scapularis</i> [tick]	Lyme disease
<i>Glossina morsitans</i> [tsetse fly]	Sleeping sickness
<i>Rhodnius prolixus</i> [bug]	Chagas disease
<i>Lutzomia longipalpis</i> [sand fly]	Leishmaniasis (New World)
<i>Phlebotomus papatasi</i> [sand fly]	Leishmaniasis (Old World)

- How are they selected?

- US NIAID contract + some ‘orphan’ genomes

Which data do we have?

VB? Status?	Organisms	Data
 ●	<i>Anopheles gambiae</i>	Gene set ^{***} , ESTs, traces, assembly ^{***}
 ●	<i>Aedes aegypti</i>	Gene set, ESTs, traces, assembly
 ●	<i>Culex pipiens</i>	Prelim. gene set, ESTs, traces, assembly
 ●	<i>Pediculus humanus</i>	Prelim. gene set, ESTs, traces, assembly
 ●	<i>Ixodes scapularis</i>	ESTs, traces
●	<i>Glossina morsitans</i>	ESTs
●	<i>Rhodnius prolixus</i>	ESTs
●	<i>Lutzomia longipalpis</i>	ESTs
●	<i>Phlebotomus papatasi</i>	ESTs

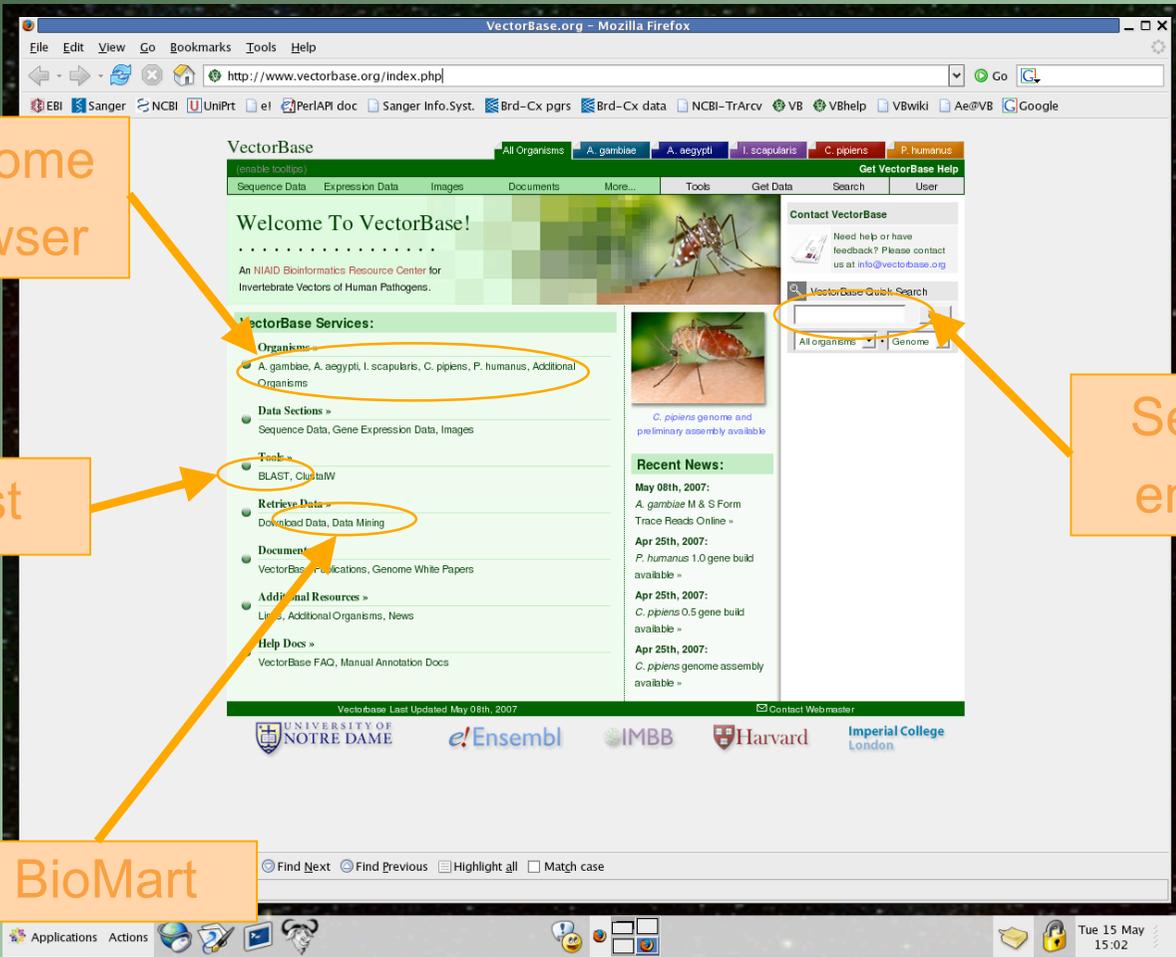
Which information & tools do we have?

- Data type
 - Gene sets (structure, Xrefs, expression data)
 - Comparison (genomic, paralogs/homologs)
 - Ontologies, images
 - Sequences (ESTs, traces, assembly)
- Data mining tools
 - Sequence comparison
 - Sequence alignment
- Data querying tool
 - Biomart

Where are the data from?

- Genome annotation
 - automatic annotation (Ensembl pipeline)
 - manual annotation (FlyBase system)
 - community annotation (you!)
- Expression data
 - community data (you again!)
- Ontologies & Anobase
 - Control Vocabulary manually generated
 - Anobase merging

Entry points...



The screenshot shows the VectorBase website interface in a Mozilla Firefox browser window. The address bar displays <http://www.vectorbase.org/index.php>. The page content includes a navigation menu with tabs for 'All Organisms', 'A. gambiae', 'A. aegypti', 'I. scapularis', 'C. pipiens', and 'P. humanus'. Below the navigation, there are sections for 'VectorBase Services', 'Organisms', 'Data Sections', 'Tools', 'Retrieve Data', 'Document', 'Additional Resources', and 'Help Docs'. The 'Organisms' section lists 'A. gambiae, A. aegypti, I. scapularis, C. pipiens, P. humanus, Additional Organisms'. The 'Tools' section includes 'BLAST, ClustalW'. The 'Retrieve Data' section includes 'Download Data, Data Mining'. The 'Search engine' callout points to a search box in the top right corner. The 'Genome browser' callout points to the 'Organisms' list. The 'Blast' callout points to the 'BLAST, ClustalW' link. The 'BioMart' callout points to the 'Download Data, Data Mining' link. The footer of the page features logos for the University of Notre Dame, Ensembl, IMBB, Harvard, and Imperial College London. The browser's status bar at the bottom shows the date and time: 'Tue 15 May 15:02'.

Entry point-1: Walk along the genome...

- Ensembl-like

Chromosome 3R
10,061,488 - 10,076,170

View of Chromosome 3R
Graphical overview
View alongside ...
View region at UCSC

Export data
Export information about region
Export sequence as FASTA
Export EMBL file
Export Gene info in region

Chromosome 3R
Chr. 3R

Overview
Chr. 3R band
DNA(contigs)
Ensembl Genes
ncRNA Genes



Detailed view

Features ▼ Comparative ▼ DAS Sources ▼ Repeats ▼ Decorations ▼ Export ▼ Image size ▼ Help ▼

Jump to region 3R : 10061488 - 10076170 Refresh Band: Refresh

Zoom
 << 5MB < 2MB < 1MB < Window Window > 1MB > 2MB > 5MB >>

Chr. 3R
Length 14.68 Kb
Forward strand

Aaeg trans BLAT
BLAST Drosophila

DNA(contigs)
< AAAB1008964

Ensembl trans.
< TEP15
Ensembl Known Protein Coding
< ENSANGESTT00000004054

EST trans.
< ENSANGESTT00000004056
< ENSANGESTT00000004055
< ENSANGESTT00000004057

BLAST Drosophila
Length 14.68 Kb
Reverse strand

Gene legend
 ■ Ensembl Known Protein Coding
 ■ EST gene
 All tracks are currently switched on
 EnsEMBL Anopheles gambiae version 43.3f (Agamp3) Chromosome 3R 10,061,488 - 10,076,170

Stop for a rest on a gene
Visualise repeats, gene
Xrefs, homologies
on a chromosome

ENSANGG00000017033

Ensembl Gene Report for ENSANGG00000017033

Gene	TEP15 (Anopheles symbol) To view all Ensembl genes linked to the name click here .
Ensembl	ENSANGG00000017033

- Gene information
 - Genomic sequence
 - Genomic sequence alignment
 - Gene splice site image
 - Gene tree info.
 - Gene variation info.
 - ID history
 - Transcript information
 - Exon information
 - Protein information
 - Export gene data
- Chromosome 3R
10,061,489 - 10,076,170
- View of Chromosome 3R
 - Graphical view
 - Graphical overview
 - Export information about region
 - Export sequence as FASTA
 - Export EMBL file
 - Export Gene info in region
 - Export SNP info in region



Orthologue Prediction

The following gene(s) have been identified as putative orthologues:

Species	Type	Gene identifier
<i>Aedes aegypti</i>	1-to-many	AAEL008607 (Novel Ensembl prediction) [MultiContigView] [Align] tep3
	1-to-many	AAEL014755 (Novel Ensembl prediction) [MultiContigView] [Align] tep2
<i>Caenorhabditis elegans</i>	1-to-1	ZK337.1

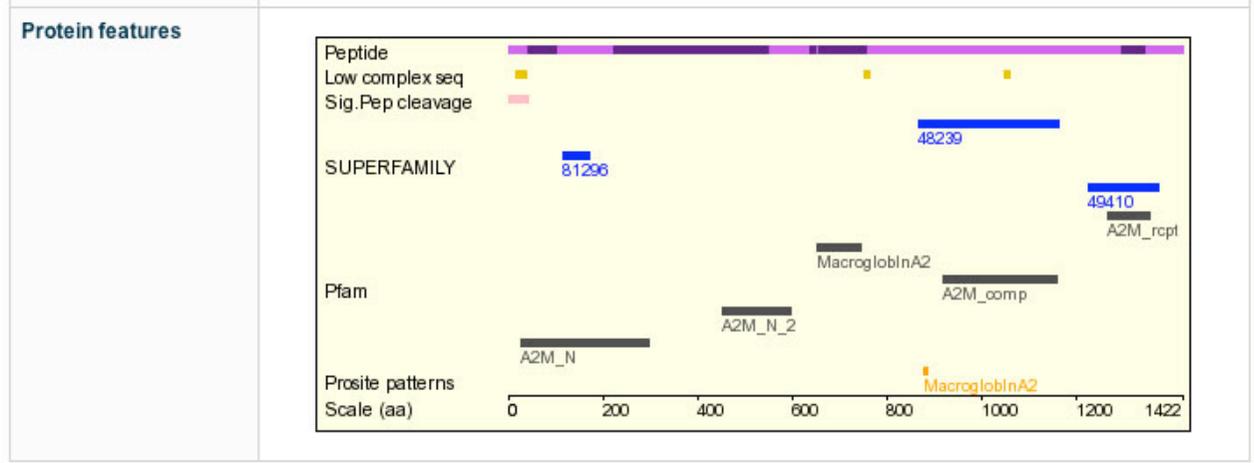
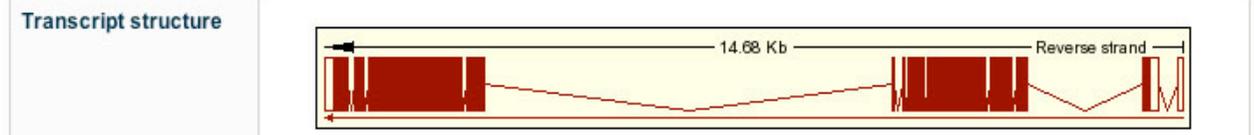
- Gene
- DAS
- Transcript
- Transcript
- Similarity

GO

The following GO terms have been mapped to this entry via UniProt and/or RefSeq:
[GO:0004866 \[endopeptidase inhibitor activity\]](#) IEA

InterPro

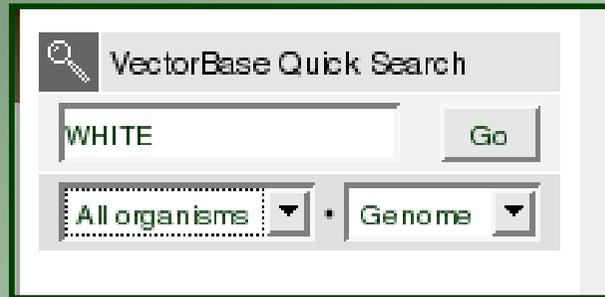
IPR011627 A-macroglobulin receptor - [[View other genes with this domain](#)]
 IPR002890 Alpha-2-macroglobulin, N-terminal - [[View other genes with this domain](#)]
 IPR011625 Alpha-2-macroglobulin, N-terminal 2 - [[View other genes with this domain](#)]
 IPR001599 Alpha-2-macroglobulin - [[View other genes with this domain](#)]
 IPR011626 A-macroglobulin complement component - [[View other genes with this domain](#)]



Entry point-2 & 3: Hunt for genes ...

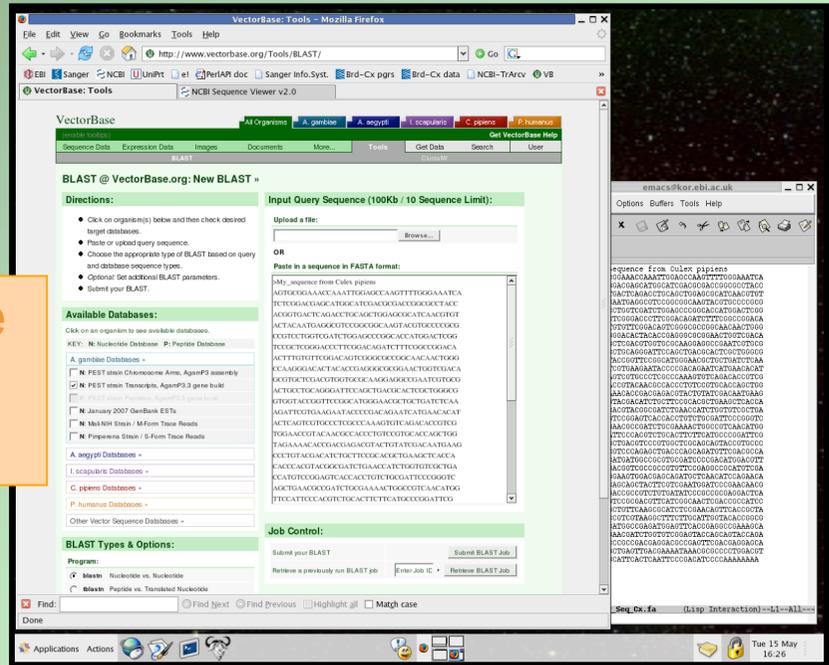
- Search engine

Try with a gene name, a chromosome, an organism...



- Blast

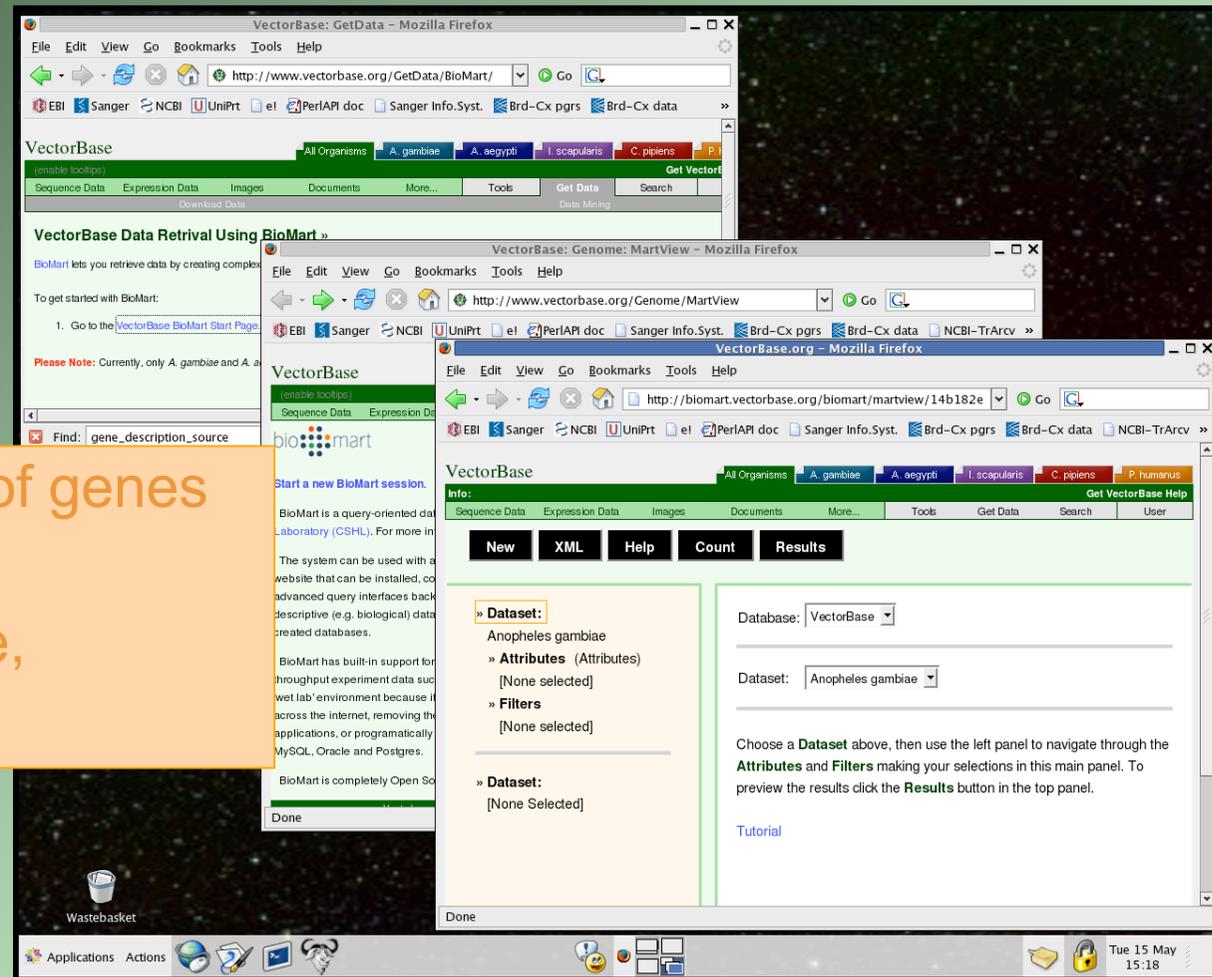
Compare your favorite sequence to the transcript/protein sets or to the genome



Entry point-4: Collect large scale data...

- BioMart

Query whole set of genes
By organisms,
by chromosome,
or by IDs



What's next?

- More organisms
 - Annotate (*Pediculus*, *Culex*) and reannotate/update (*Anopheles*, *Aedes*)
 - Data repository (add data as they are generated)
 - ‘orphan’ organisms (Sand fly)
- Involve the community
 - Tutorial, documentation, help
 - Data submission by the users

Acknowledgments

- **Automatic Annotation, EBI, UK**
Ewan Birney
Martin Hammond
Dan Lawson
Karyn Megy
- **Manual Annotation, Harvard, MA**
Bill Gelbart
Kathy Campbell
- **Website, NotreDame University, IN**
Frank Collins (Anopheles)
Dave Severson (Aedes)
Greg Madey
Rob Bruggner
E.O. Stinton
Ryan Kennedy
- **AnoBase & Ontologies, IMBB, GR**
Kitsos Louis
Emmanuel Dialynas
Pantelis Topalis
- **Expression data, Imperial College, London**
Fotis Kafatos, George Christophidis
Bob Mc Callum
Seth Redmond
- **University of Riverside, CA**
Peter Atkinson (Culex)
- **Community**
Neil Lobo (Anopheles & Aedes)
Peter Arensburger (Culex)
Catherine Hill (Ixodes)
YOU!



Ensembl team

For allowing us to parasite them

Questions?

<http://www.vectorbase.org>

NAR DataBase issue 2007

‘VectorBase, a home for invertebrate vector of human pathogens’
Lawson D. *et al.*

E-mail

Everybody: info@vectorbase.org

...or just me: kmegy@ebi.ac.uk

We speak:

