VectorBase,
a Resource Centre for Invertebrate Hosts of Human Pathogens

http://www.vectorbase.org

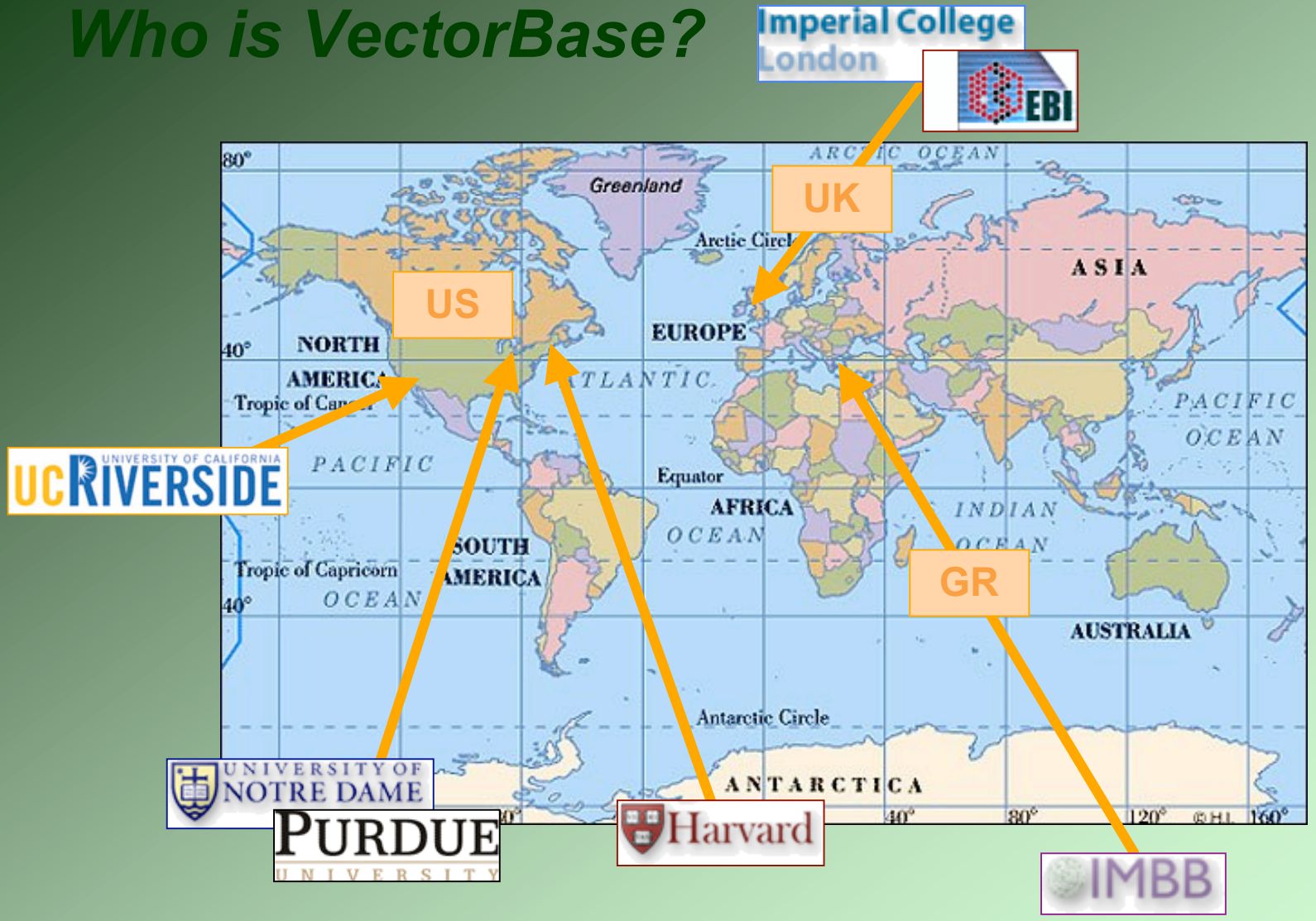
Karyn Mégy
European Bioinformatics Institute, Cambridge
UK

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	**Generate data**
Available to all	Automatic annotation
Display data	Manual annotation
Easy to browse, download	Expression data
Mine data	**Community**
Compare, align	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

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

- How are they selected?
  - US NIAID contract + some ‘orphan’ genomes

Which data do we have?

VB? Status?	Organisms	Data
VectorBase	*Anopheles gambiae*	Gene set***, ESTs, traces, assembly***
VectorBase	*Aedes aegypti*	Gene set, ESTs, traces, assembly
VectorBase	*Culex pipiens*	Prelim. gene set, ESTs, traces, assembly
VectorBase	*Pediculus humanus*	Prelim. gene set, ESTs, traces, assembly
VectorBase	*Ixodes scapularis*	ESTs, traces
	Glossina morsitans	ESTs
	Rhodnius prolixus	ESTs
	Lutzomia longipalpis	ESTs
	Phlebotomus papatasis	ESTs

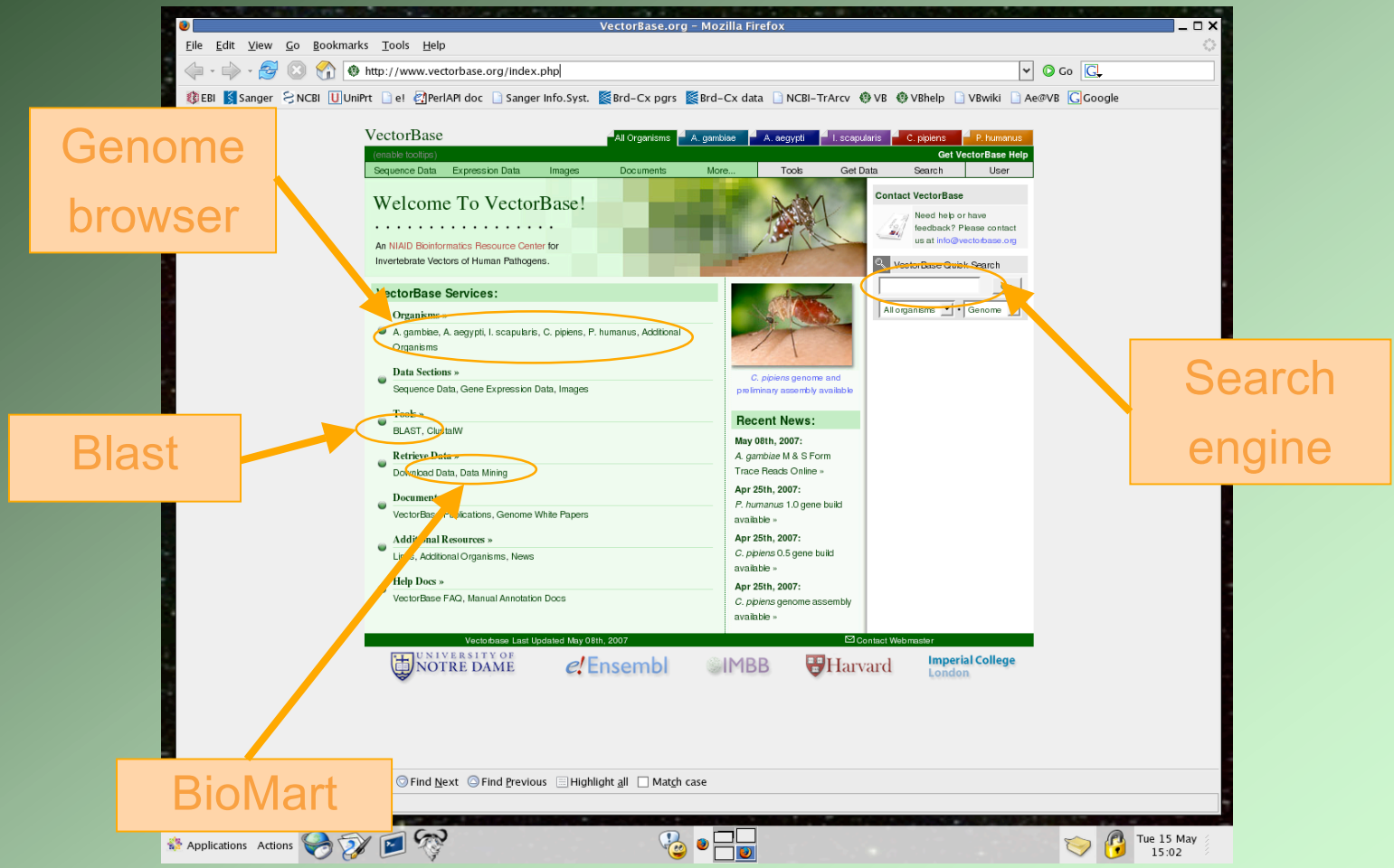
Which information & tools do we have?

- 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?

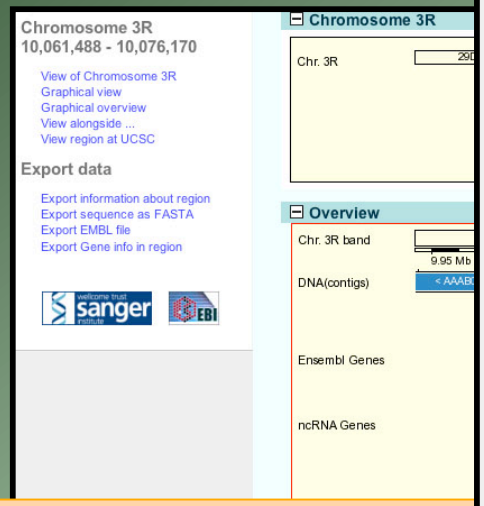
- Genome annotation
  - automatic annotation (Ensembl pipeline)
  - manual annotation (FlyBase system)
  - community annotation (you!)
- Expression data
  - community data (you again!)
- Ontologies & Anobase
  - Control Vocabulary manually generated
  - Anobase merging

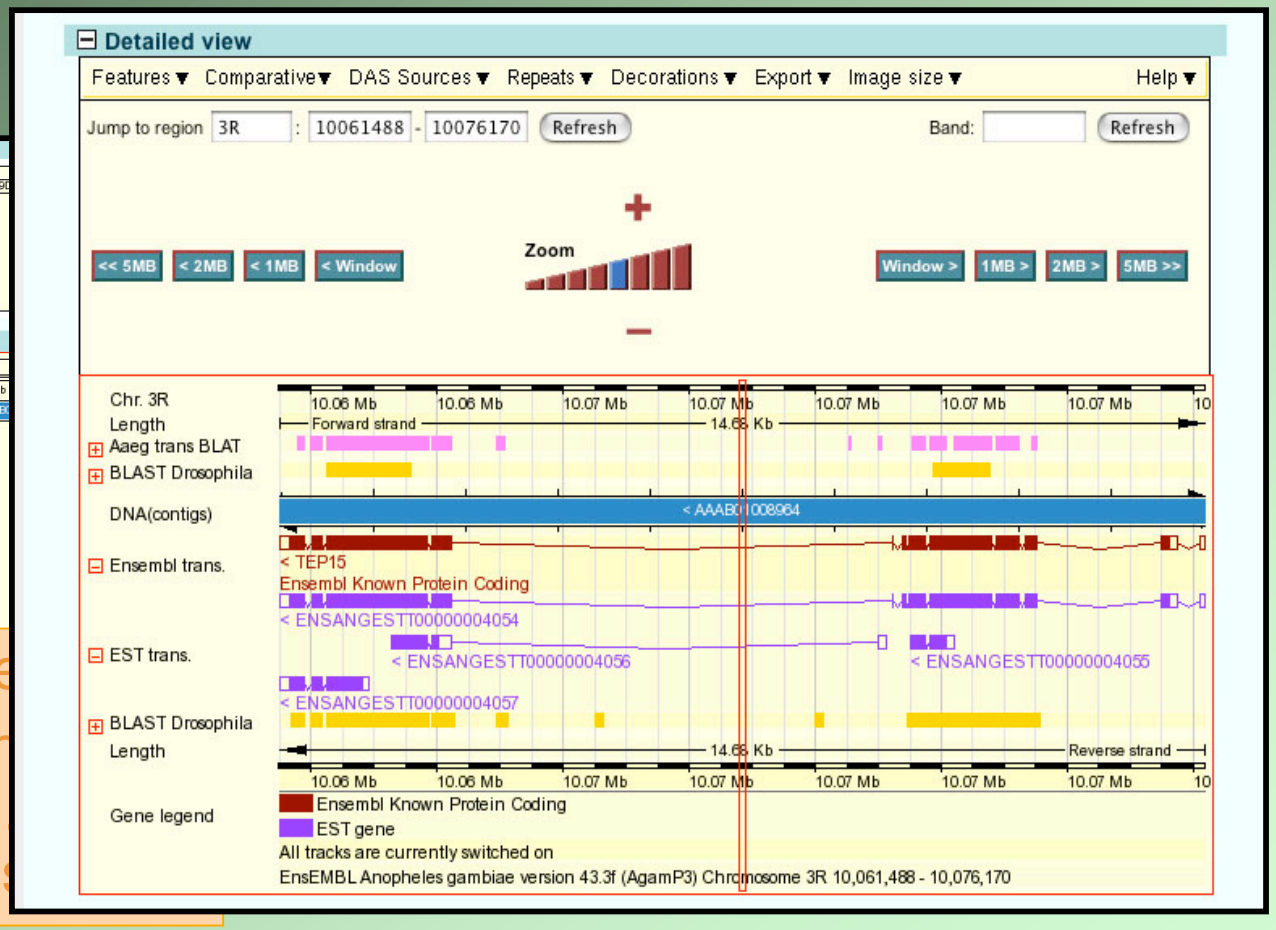
Entry points…



Entry point-1: Walk along the genome…

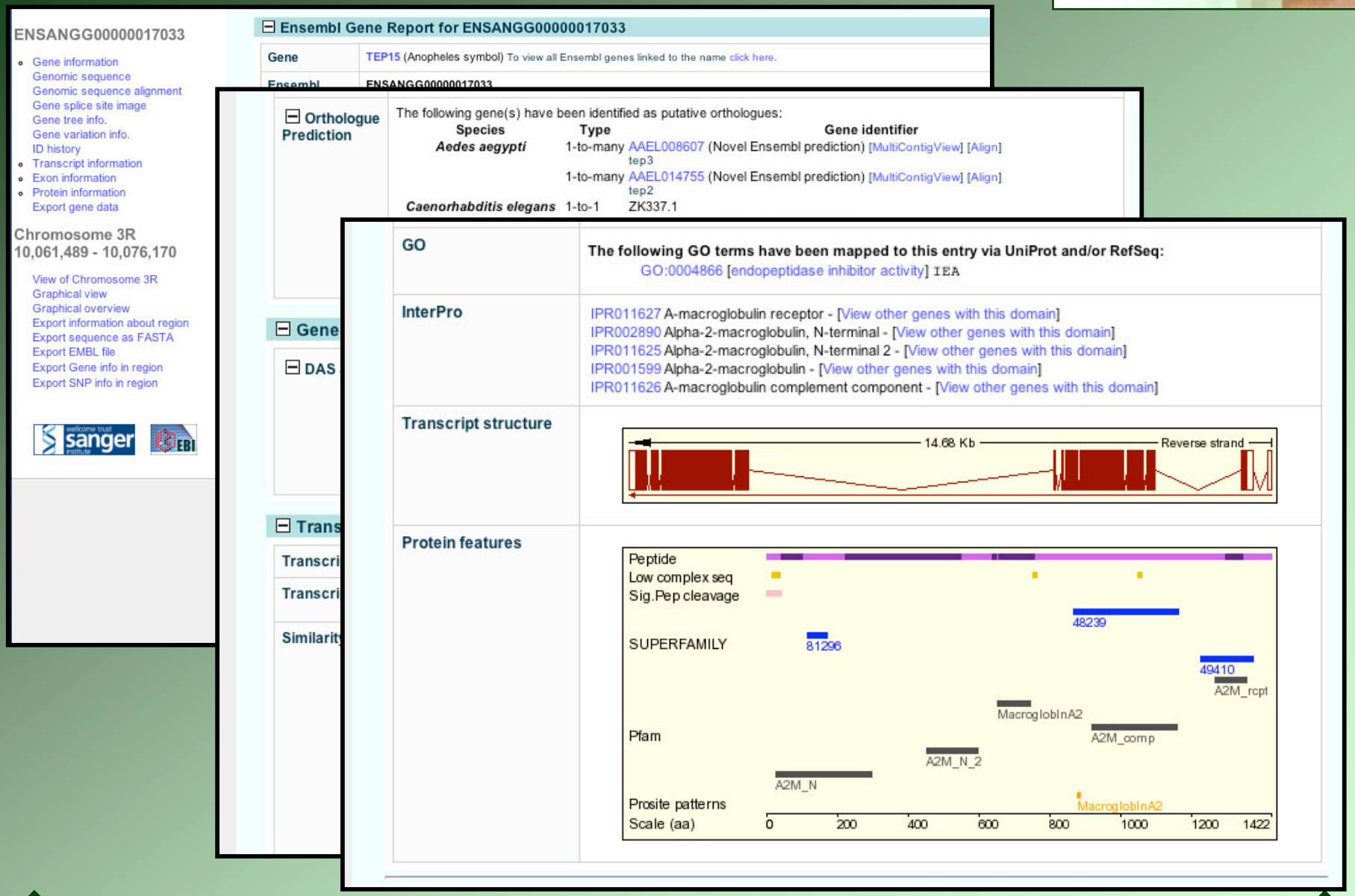
- Ensembl-like





Stop for a rest on a ge

Visualise repeats, ger
Xrefs, homologies
on a chromosome


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
Ensembl Gene Report for ENSANGG00000017033
Gene
TEP15 (Anopheles symbol) To view all Ensembl genes linked to the name click here.
Ensembl
ENSANGG00000017033
Orthologue Prediction
The following gene(s) have been identified as putative orthologues:
Species
Type
Gene identifier
Aedes aegypti
1-to-many AAEL008607 (Novel Ensembl prediction) [MultiContigView] [Align]
tep3
1-to-many AAEL014755 (Novel Ensembl prediction) [MultiContigView] [Align]
tep2
Caenorhabditis elegans
1-to-1 ZK337.1
Gene
DAS
Trans
Transcri
Transcri
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]
Transcript structure
14.68 Kb
Reverse strand
Protein features
Peptide
Low complex seq
Sig.Pep cleavage
SUPERFAMILY
48239
81296
49410
A2M_rcpt
MacroglobInA2
Pfam
A2M_comp
A2M_N_2
A2M_N
Prosite patterns
MacroglobInA2
Scale (aa)
0
200
400
600
800
1000
1200
1422

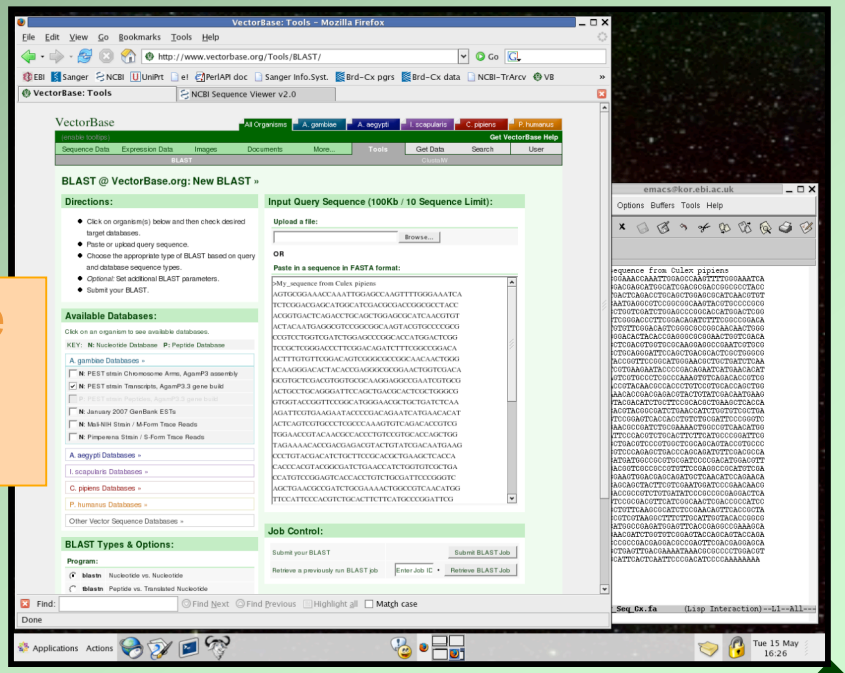
Entry point-2 & 3: Hunt for genes …

- Search engine

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

- Blast

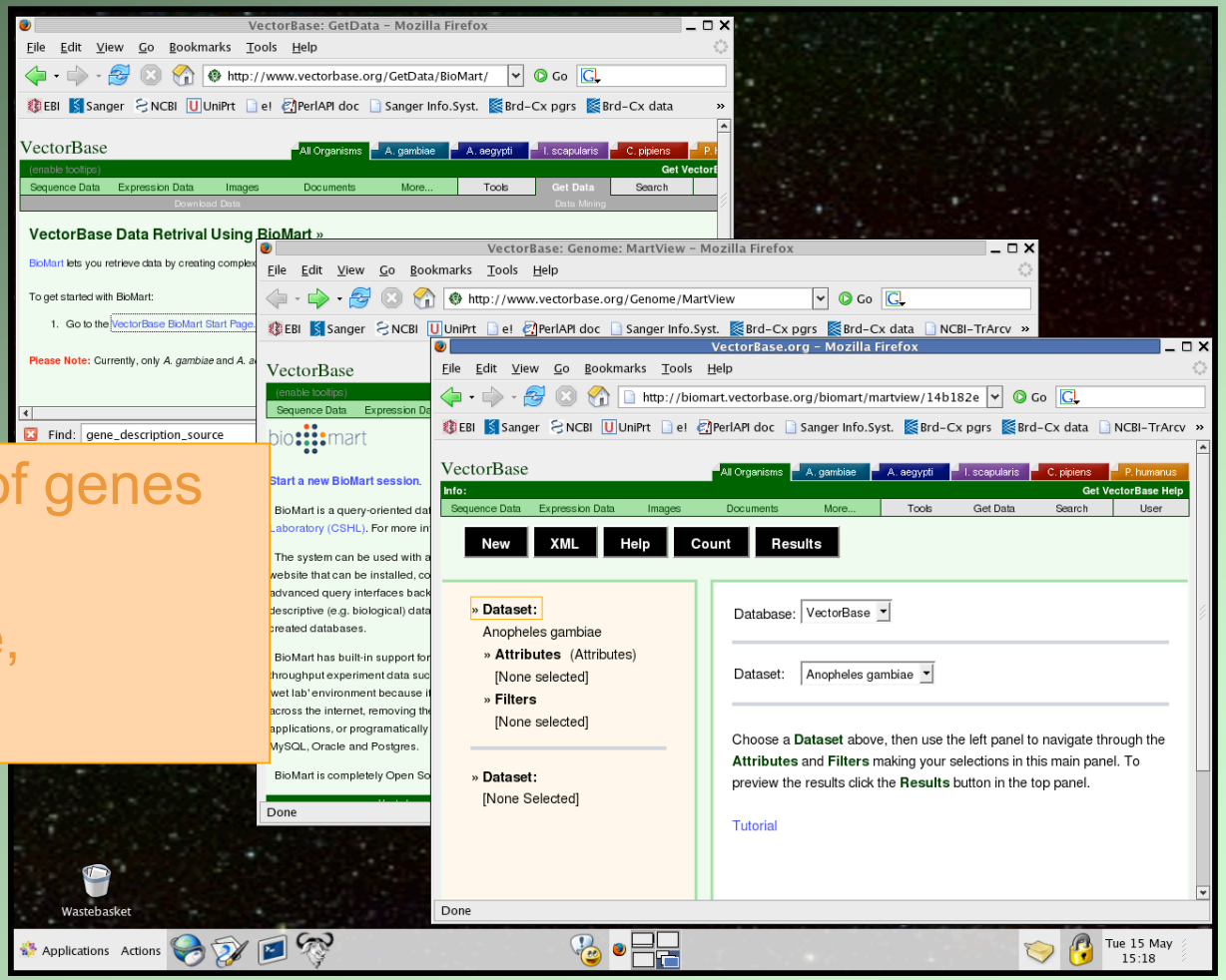
Compare your favorite sequence
to the transcript/protein sets
or to the genome



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: