

Comparative Genomics

Gloria I. Giraldo-Calderón
August 2016



VectorBase

Bioinformatics Resource for Invertebrate Vectors of Human Pathogens

When comparative genomics is helpful?

Sample of comparative genomic queries that can be solved with VectorBase

General features

Table 2. General characteristics of the *A. darlingi* genome

Genome feature	<i>A. darlingi</i>	<i>A. gambiae</i> ^a	<i>A. aegypti</i> ^a
Genome size (Mb)	173.92	278.25	1379
Percent of G + C (%)	48.15	40.9	38.2
Protein coding length (Mb) and (% genome length)	18.2 (10.4)	64.92 (23.3)	224.9 (16.3)
Total number of exons	47 990 ^b	56 210	66 827
Number of protein-coding genes	10 457 ^b	12 670	15 704
Percent genes with introns (%)	91.59	93.6	90.1
Average number of exons per gene	4.6	4.4	4.0
Average gene length (bp)	1735	5124 ^c	14 587 ^c
Total tRNAs	346	450	995

^aStatistics were derived from genome updates for *A. gambiae* AgamP3 (Vectorbase, version 66.3) and *A. aegypti* AegL1 (Vectorbase, version 66.1).

^bIncludes 13 mitochondrial genes.

^cIncludes introns but not untranslated regions.

Marinotti et al. 2013.
Nucleic Acids Res. 41(15): 7387-400.

Sample of comparative genomic queries that can be solved with VectorBase

Homologous genes (infer gene function)

Table 3. Numbers of annotated genes encoding OBP and OR

Protein	<i>Ae. albopictus</i>	<i>Ae. aegypti</i>	<i>An. gambiae</i>	<i>Cx. quinquefasciatus</i>	<i>D. melanogaster</i>
OBP	86	64	58	50	51
OR	158	110	80	88	61

Chen et al. 2013.
Proc Natl Acad Sci U S A. 112(44): E5907-15

Sample of comparative genomic queries that can be solved with VectorBase

Molecular evolution

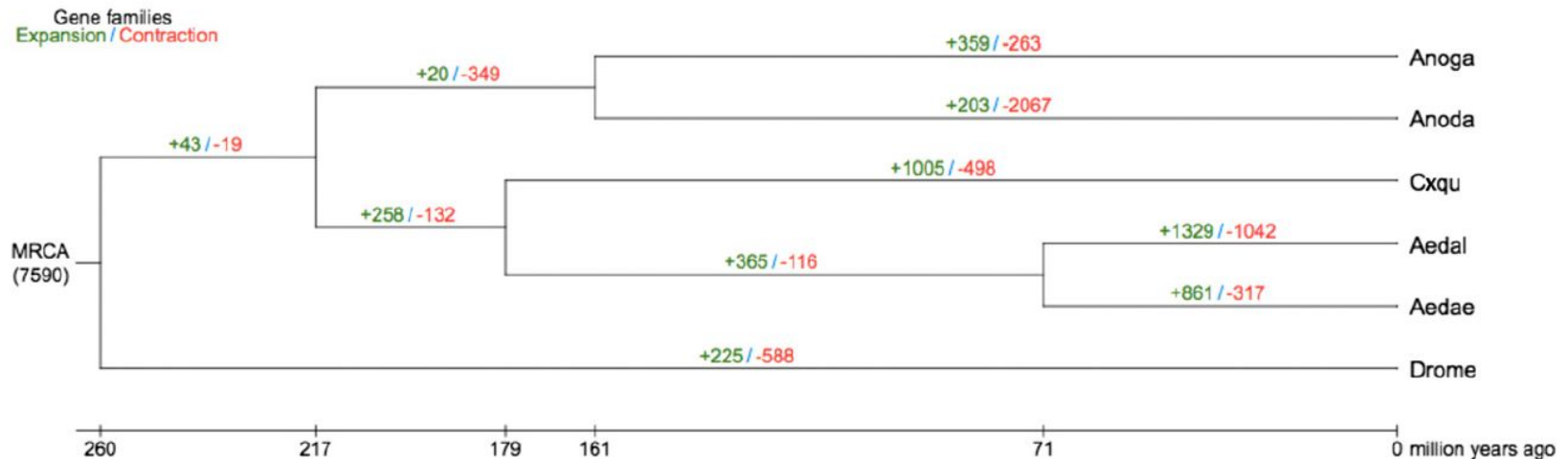
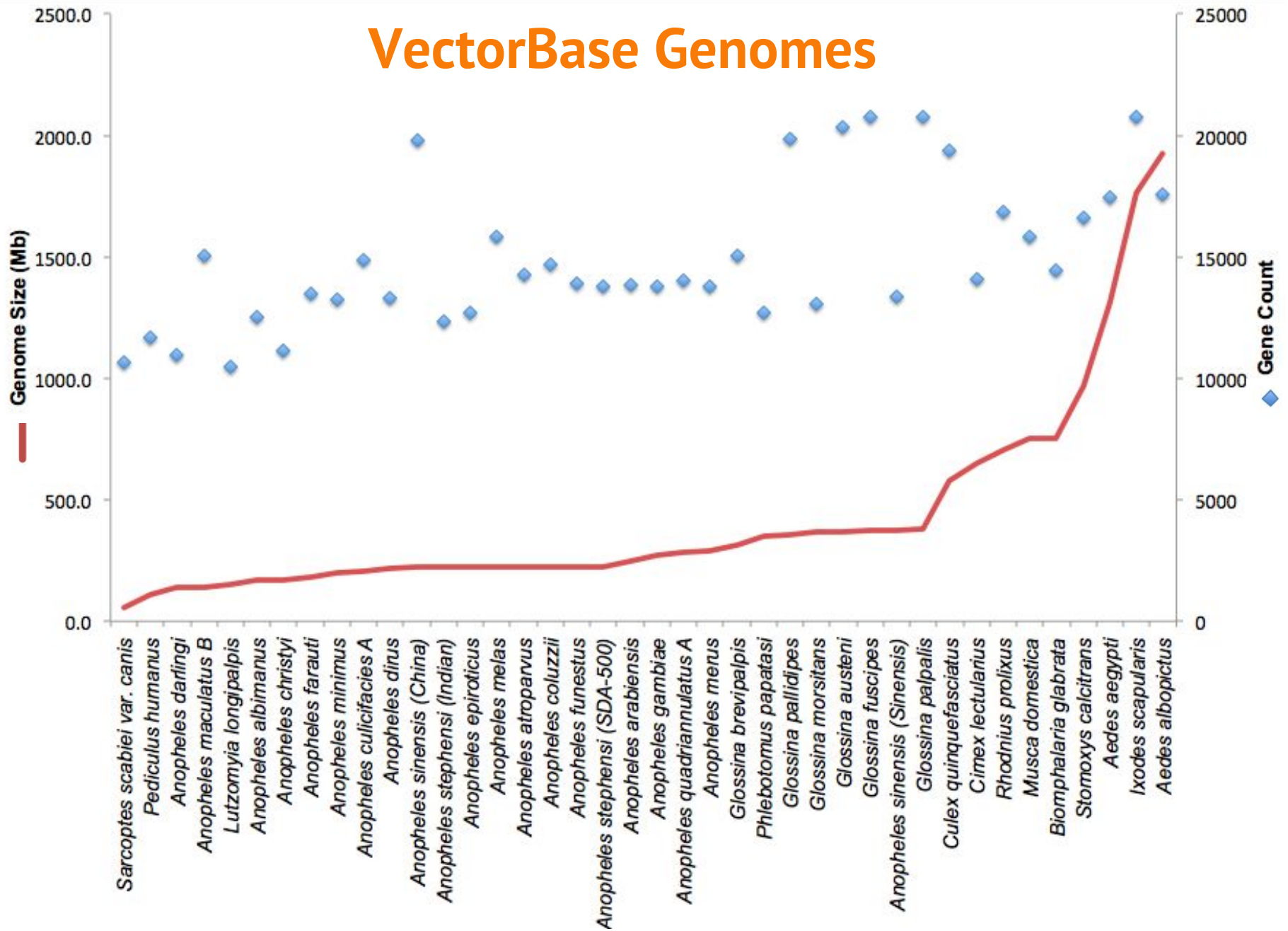
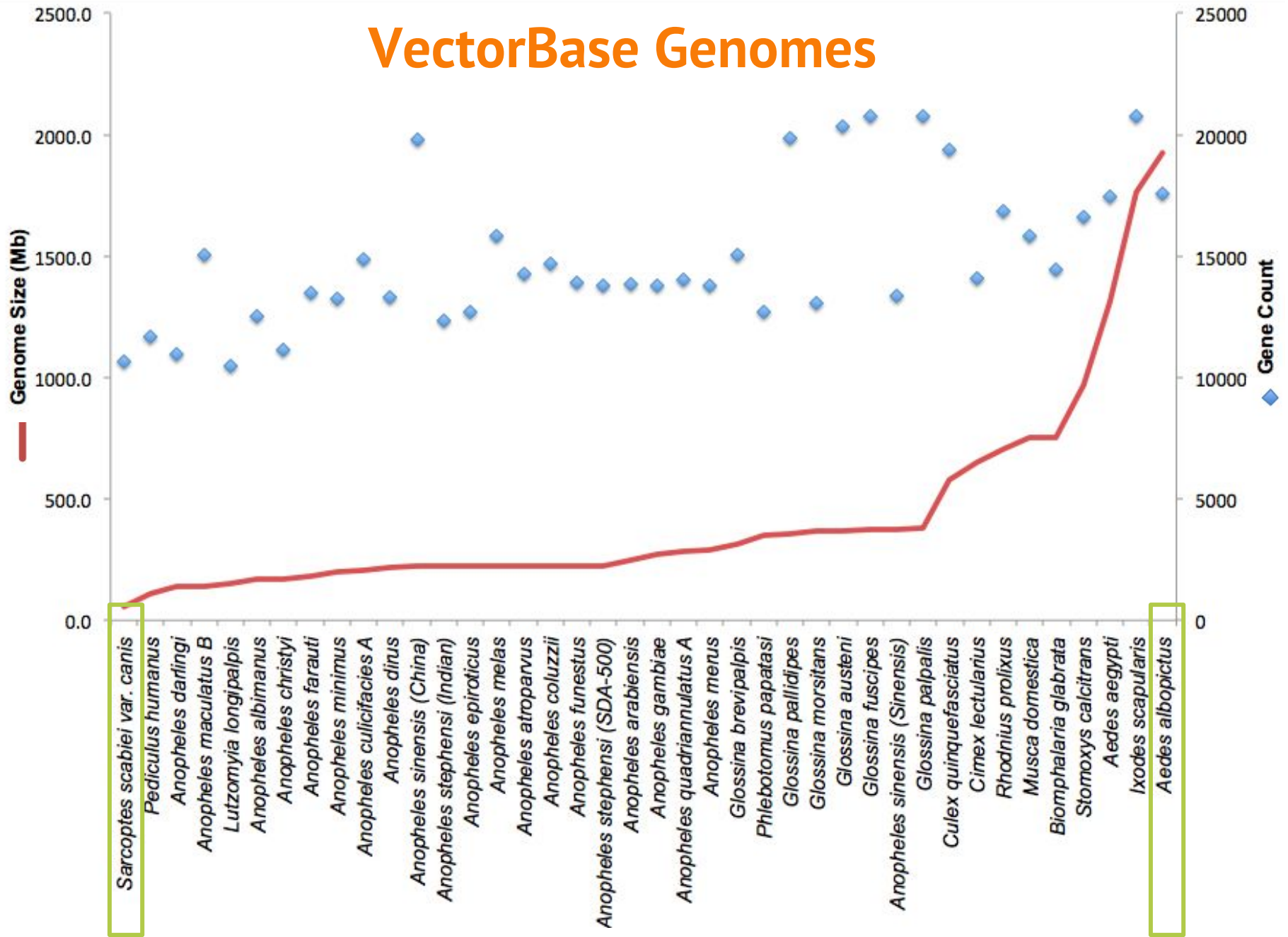


Fig. 2. Expansion and contraction of gene families among mosquito species. The numbers designate the number of gene families that have expanded (green) and contracted (red) since the split from the last common ancestor. The most recent common ancestor (MRCA) has 7,590 gene families. Aedae, *Ae. aegypti*; Aedal, *Ae. albopictus*; Anoda, *An. darlingi*; Anoga, *An. gambiae*; Cxqu, *Cx. quinquefasciatus*; Drome, *D. melanogaster*.

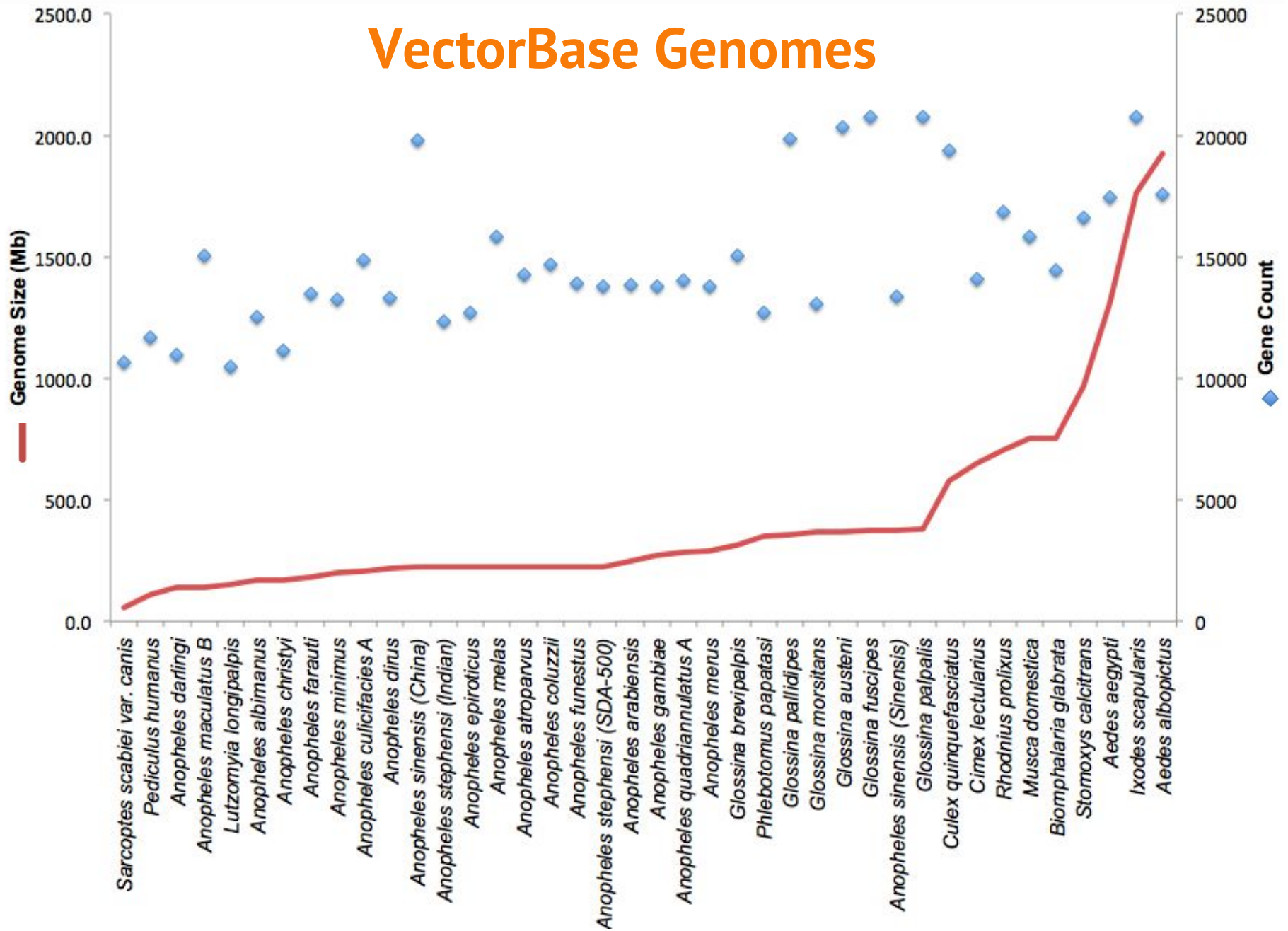
VectorBase Genomes



VectorBase Genomes



VectorBase Genomes



VectorBase Genomes

Arthropoda

Insecta

- Diptera
- Hemiptera
- Phthiraptera

Acari

Mollusca

Gastropoda (snail, *Biomphalaria glabrata*)

Outline

Protein comparisons

- gene trees
- paralogues and orthologues
- the use of BioMart

DNA comparisons

- pairwise alignments

Protein comparisons

- Protein comparison are done for all genes in all VectorBase genomes, <https://www.vectorbase.org/genomes>
- gene alignments (input) ---> gene trees
- gene trees ---> infer homologues (paralogues and orthologues)

Pipeline for protein comparisons

longest protein for every gene in VectorBase



align all proteins vs all proteins



gene trees are inferred

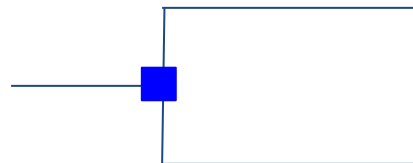
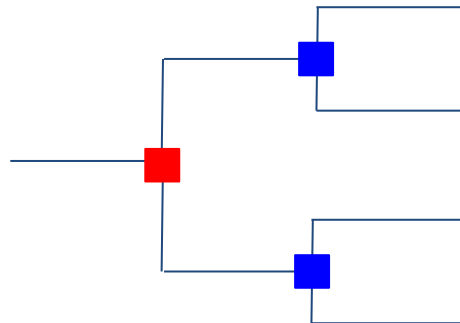
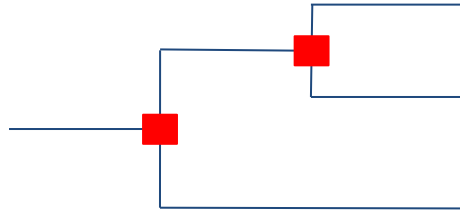


the gene tree is reconciled with the species tree
to infer homologous genes

Gene Trees

■ Speciation

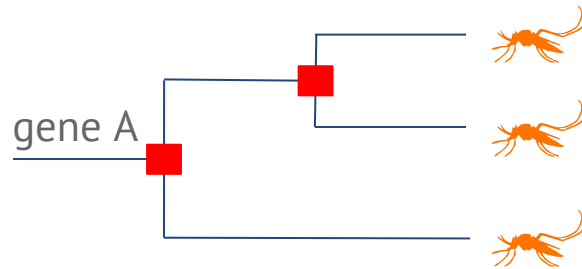
■ Duplication



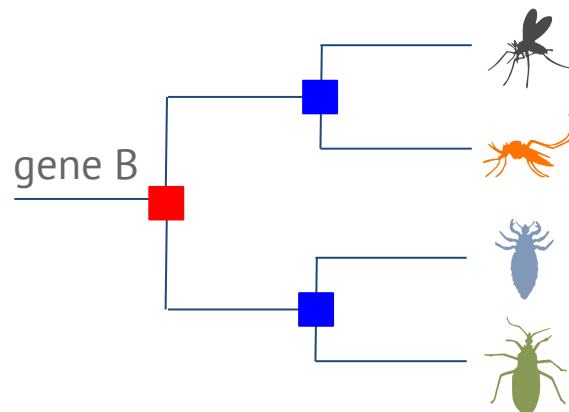
Gene Trees

■ Speciation

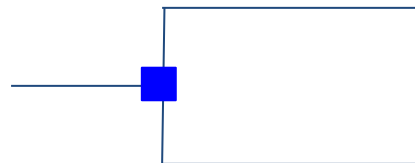
■ Duplication



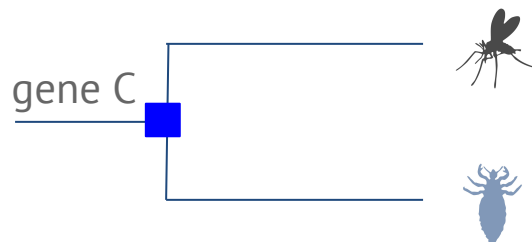
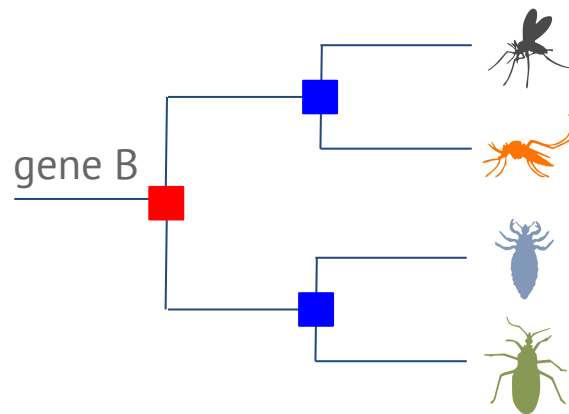
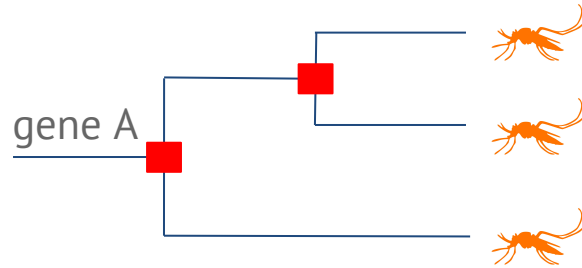
A species with one or multiple homolog gene copies: paralogs



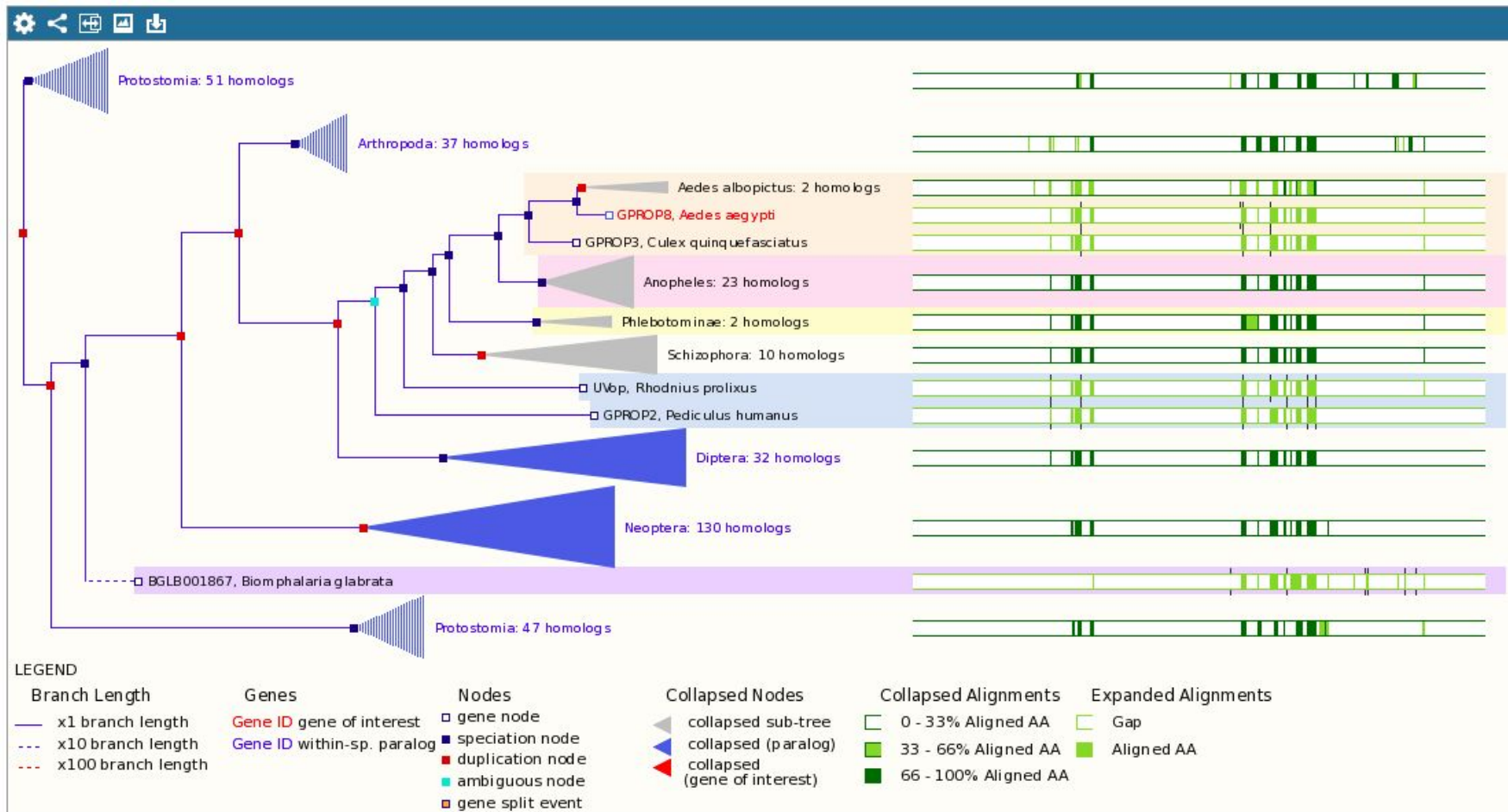
Different species with one or multiple homolog gene copies: orthologs



Gene Trees



These are ortholog genes,
with gene lost and/or gain
in different species



Orthologues



[Download protein sequences](#) [Download DNA sequences](#)

Summary of orthologues of this gene

Click on 'Show details' to display the orthologues for one or more groups of species. Alternatively, click on 'Configure this page' to choose a custom list of species.

Species set	Show details	With 1:1 orthologues	With 1:many orthologues	With many:many orthologues	Without orthologues
All (35 species)	<input type="checkbox"/>	28	13	0	0
Anophellinae (19 species)	<input type="checkbox"/>	15	8	0	0
Brachycera (8 species)	<input type="checkbox"/>	8	0	0	0
Culicinae (2 species)	<input checked="" type="checkbox"/>	1	2	0	0
Eukaryota (1 species)	<input type="checkbox"/>	1	0	0	0
Gastropoda (1 species)	<input type="checkbox"/>	0	1	0	0
Hemiptera (1 species)	<input type="checkbox"/>	1	0	0	0
Phlebotominae (2 species)	<input type="checkbox"/>	2	0	0	0

Selected orthologues

Show All  entries Show/hide columns Filter <input type="text"/> 								
Species	Type	dN/dS	Ensembl identifier & gene name	Compare	Location	Target %id	Query %id	
Aedes albopictus	1-to-many	0.01178	AALF007317 No description	<ul style="list-style-type: none"> Region Comparison Alignment (protein) Alignment (cDNA) Gene Tree (image) 	JXUM01S000183:187269-188006:-1	98	43	
Aedes albopictus	1-to-many	0.11956	AALF007320 No description	<ul style="list-style-type: none"> Region Comparison Alignment (protein) Alignment (cDNA) Gene Tree (image) 	JXUM01S000183:208383-227964:-1	87	52	
Culex quinquefasciatus	1-to-1	n/a	CPIJ009246 GPROP3 ultraviolet wavelength sensitive opsin	<ul style="list-style-type: none"> Region Comparison Alignment (protein) Alignment (cDNA) Gene Tree (image) 	supercont3.237:371599-376427:-1	93	93	

Paralogues

 Download paralogues

Show All entries		Show/hide columns		Filter			
Type	Ancestral taxonomy	Ensembl Identifier & gene name	Compare	Location	Target %id	Query %id	
Paralogues (same species)	Neoptera	AAEL003035 GPROP9 short wavelength sensitive opsin [Source:VB Community Annotation;Acc:AAEL003035]	<ul style="list-style-type: none">Region ComparisonAlignment (protein)Alignment (cDNA)	supercont1.75:1344859-1358612:1	49	49	
Paralogues (same species)	Arthropoda	AAEL005322 GPROP10 unknown wavelength sensitive opsin [Source:VB Community Annotation;Acc:AAEL005322]	<ul style="list-style-type: none">Region ComparisonAlignment (protein)Alignment (cDNA)	supercont1.151:1942356-1943586:1	30	29	
Paralogues (same species)	Arthropoda	AAEL005625 GPROP5 long wavelength sensitive opsin [Source:VB Community Annotation;Acc:AAEL005625]	<ul style="list-style-type: none">Region ComparisonAlignment (protein)Alignment (cDNA)	supercont1.165:1782454-1783587:-1	37	36	

Demo

Protein comparisons:

Go to the gene page in the genome browser,
select the gene tab

Outline

Protein comparisons

- gene trees
- paralogues and orthologues
- the use of BioMart

DNA comparisons

- pairwise alignments

DNA comparisons

- criteria of inclusion: done for closely-related species only
- not ideal for fragmented genomes
- method used:
 - nucleotide alignment for closer species (LASTZ_NET)
 - amino acid alignment for more distant species (TRANSLATED_BLAT_NET)
- only updated when a genome assembly is updated
- helpful to identify conserved coding and non-coding regions
- useful to visualize and compare trouble homologs when annotating genes

VectorBase genomes

Arthropoda

Insecta

- **Diptera**: *Aedes* (2 spp), *Anopheles* (19*) except *A. minimus*, *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

underline: these 8 spp are not included in the alignments

https://www.vectorbase.org/compara_analyses.html

Available alignments

- Click on a species name to expand/collapse its alignment list
- Click on a analysis name to see available statistics

Toggle All ☐ or click a species names to expand/collapse its alignment list

Aedes aegypti

6 alignments

Aedes albopictus	LASTZ_NET example stats
Anopheles darlingi	TRANSLATED_BLAT_NET example stats
Anopheles gambiae	LASTZ_NET example stats
Culex quinquefasciatus	LASTZ_NET example stats
Ixodes scapularis	TRANSLATED_BLAT_NET example stats
Pediculus humanus	TRANSLATED_BLAT_NET example stats

Aedes albopictus

1 alignment

Aedes aegypti	LASTZ_NET example stats
---------------	---

Anopheles gambiae		19 alignments		
Aedes aegypti	LASTZ_NET	example	stats	
Anopheles albimanus	LASTZ_NET	example	stats	
Anopheles arabiensis	LASTZ_NET	example	stats	
Anopheles atroparvus	LASTZ_NET	example	stats	
Anopheles christyi	LASTZ_NET	example	stats	
Anopheles coluzzii	LASTZ_NET	example	stats	
Anopheles culicifacies	LASTZ_NET	example	stats	
Anopheles darlingi	LASTZ_NET	example	stats	
Anopheles dirus	LASTZ_NET	example	stats	
Anopheles epiroticus	LASTZ_NET	example	stats	
Anopheles farauti	LASTZ_NET	example	stats	
Anopheles funestus	LASTZ_NET	example	stats	
Anopheles maculatus	LASTZ_NET	example	stats	
Anopheles melas	LASTZ_NET	example	stats	
Anopheles merus	LASTZ_NET	example	stats	
Anopheles quadriannulatus	LASTZ_NET	example	stats	
Anopheles sinensis	LASTZ_NET	example	stats	
Anopheles stephensi	LASTZ_NET	example	stats	
Culex quinquefasciatus	LASTZ_NET	example	stats	
Anopheles maculatus		1 alignment		
Anopheles gambiae	LASTZ_NET	example	stats	
Anopheles melas		5 alignments		
Anopheles arabiensis	LASTZ_NET	example	stats	
Anopheles coluzzii	LASTZ_NET	example	stats	
Anopheles gambiae	LASTZ_NET	example	stats	
Anopheles merus	LASTZ_NET	example	stats	
Anopheles quadriannulatus	LASTZ_NET	example	stats	

Demo

DNA comparisons:

Go to the gene page in the genome browser,
select the location tab

In summary

How to search for more information or help?

E-mail us at
info@vectorbase.org

Thank you!