

## Exploring the Gene Page

**Note:** this exercise uses **VectorBase.org** as an example database, but the same functionality is available on all VEuPathDB resources.

### What is a Gene Page on VectorBase?

- The gene page conveniently consolidates links to all the data available on VectorBase for a particular gene into a single page

### Learning objectives

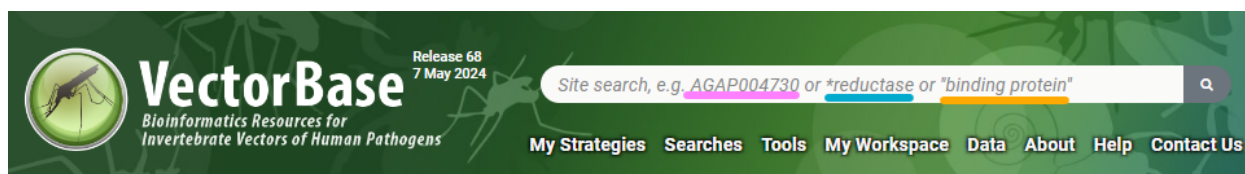
#### Gene pages:

- Become familiar with the information on gene pages
- Navigate to and from the gene pages
- Use the contents section of the gene page
- Interact with gene page subsections

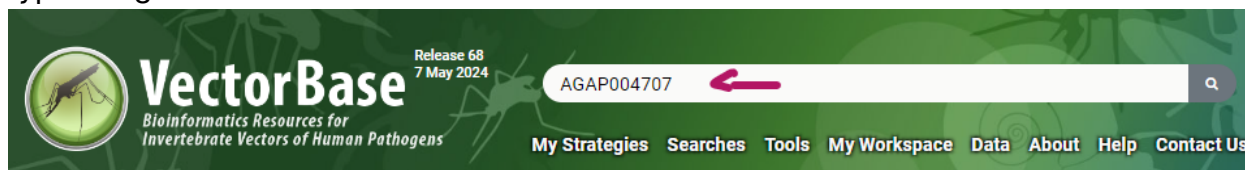
#### 1. Navigation to the Gene pages

For this exercise visit the gene page for AGAP004707 (voltage-gated sodium channel). How can you do it? Hint: Go to the **Site Search** box, at the top of the website. Notice you can use different types of keywords:

- gene ID
- gene name or function, with a wild card/asterisk
- gene name or function, with quotation marks



Type the gene ID: AGAP004707



Click on the gene ID on the results page

Genes matching **AGAP004707**

1 - 1 of 1

**Filter results** ☒ Hide zero counts

Genome  
Genes 1

**Filter Gene fields**  
select all | clear all

☐ Gene ID 1  
☐ Names, IDs, and aliases 1

**Filter organisms**  
select all | clear all | expand all | collapse all

**Gene - AGAP004707** voltage-gated sodium channel

Gene name or symbol: para

Organism: Anopheles gambiae PEST

► Fields matched: Gene ID; Names, IDs, and aliases

Gene - AGAP004707 voltage-gated sodium channel

Gene name or symbol: para

Organism: Anopheles gambiae PEST

► Fields matched: Gene ID; Names, IDs, and aliases

## 2. Explore the layout of the page

- What information is in the top left (section A in the image below)?
  - Can you easily find which chromosome this gene is located on?
  - Is this gene protein coding?
- What do the shortcuts do? (section B)
- Where can you find the contents of this page? (Section C)
- Scroll down and examine what section D contains.

**VectorBase** Release 68  
Bioinformatics Resources for Invertebrate Vectors of Human Pathogens  
7 May 2024

Site search, e.g. AGAP004730 or \*reductase or "binding protein"

My Strategies Searches Tools My Workspace Data About Help Contact Us

My Organism Profile

Add to basket Remove from favorites Download Gene

**AGAP004707** voltage-gated sodium channel

**Name:** para

**Gene Type:** protein coding gene

**Biotype Classification:** protein\_coding

**Chromosome:** 2L

**Location:** AgamP4\_2L:2,358,158..2,431,617(+)

**Species:** *Anopheles gambiae*

**Strain:** PEST

**Status:** Reference Strain

Add the first user comment

View and update community annotations in Apollo

**Model Organism Database(s)**

**FlyBase:** FBgn0085434, FBgn0285944

**Shortcuts**

Synteny Alignments Transcriptomics Protein Features Proteomics

Also see AGAP004707 in the Genome Browser or Protein Browser

**AGAP004707**

expand all | collapse all

Search section names...

► 1 Gene models ☒

► 2 Annotation, curation and Identifiers ☒

► 3 Link outs ☒

► 4 Genomic Location ☒

**1 Gene models**

# Exons in Gene 39

# Transcripts 13

▼ Gene Models

This gene is available in **Apollo** for community annotation. To find out more about Apollo, please visit this help page

**3. Explore the gene model section.** Scroll down to the gene model section of the gene page.

- What direction is the transcript relative to the chromosome?
- Does the gene have UTRs?
- How many exons does the gene have?
- Does this gene have any available community annotation from Apollo?

## 1 Gene models

[Collapse all sections for better performance](#)

# Exons in Gene ? 39

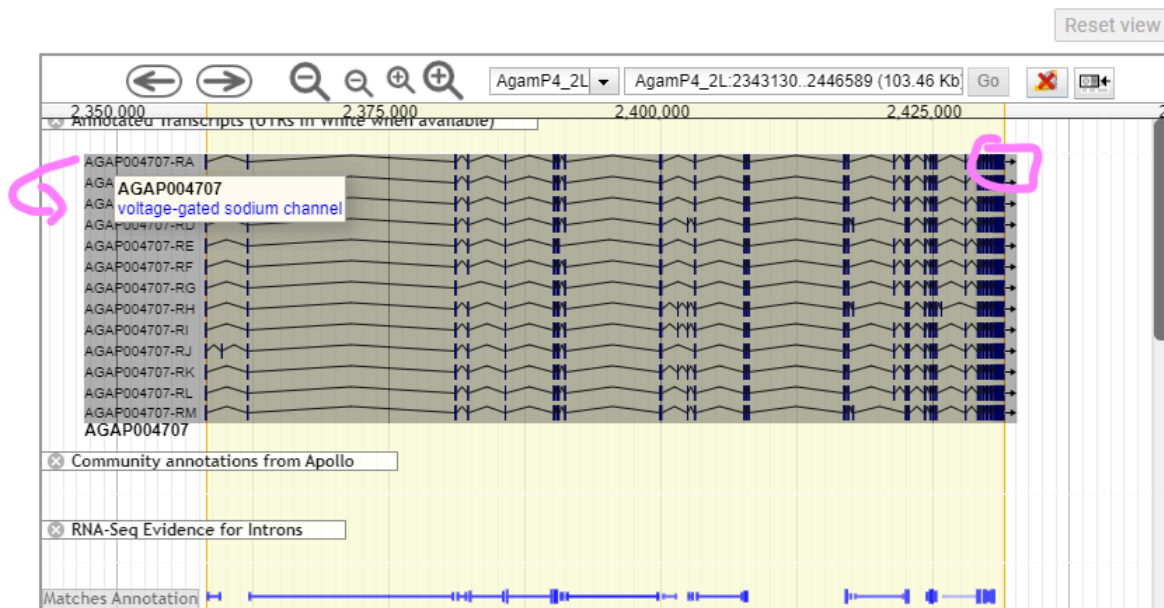
# Transcripts ? 13

▼ Gene Models ?

This gene is available in **Apollo** for community annotation. To find out more about Apollo, please visit [this help page](#).

[View in JBrowse genome browser](#)

[Annotate in Apollo](#)



- How long is the longest transcript? You can determine transcript length by expanding the Transcripts table. Click on the appropriate column header to sort the rows

GFF format of gene and transcript features [Click to open GFF in a new tab](#)

▼ **Transcripts** [Download](#) [Data sets](#)

Search this table...

| Transcript | # exons | Transcript length | Protein length | Transcript Type |
|------------|---------|-------------------|----------------|-----------------|
|            |         |                   |                |                 |

### 3. Content navigation

How do you find/navigate to the different sections of the page? Use the “Table of contents” menu on the left side. Gene page content is organized by data type and the section titles serve as links to data within the page. When expanded, each section reveals more navigation links. The content menu can also be filtered using the search function as shown below. Begin typing the ‘synteny’ in the filter to collapse the content menu.

AGAP004707

expand all | collapse all

Search section names...

- 1 Gene models ☒
- 2 Annotation, curation and identifiers ☒
- 3 Link outs ☒
- 4 Genomic Location ☒
- 5 Literature ☒
- 6 Taxonomy ☒
- 7 Orthology and synteny ☒
- 8 Phenotype ☒
- 9 Transcriptomics ☒
- 10 Sequence analysis ☒
- 11 Sequences ☒
- 12 Structure analysis ☒
- 13 Protein features and properties ☒
- 14 Function prediction ☒
- 15 Pathways and interactions ☒
- 16 Proteomics ☒
- 17 Immunology ☒

expand all | collapse all

AGAP004707

syn

7 Orthology and synteny ☒

Orthologs and Paralogs within VectorBase

- Click to navigate to the **Synteny** section. Does *Anopheles gambiae* share synteny in this region? If you get an error message reload the page
- Navigate to the **Transcript Expression** table and open the row for the experiment called “Chemosensory appendages, male and female” (Notice the filter box at the top of the table)

- i. What data type was used to produce this data (microarray or RNAseq)?
- ii. Is the expression level equal in male and female maxillary palps? What about male and female bodies and antennae?
- iii. Open the Data table for this experiment. What is the TPM expression value for unique reads mapped to female antenna?

#### 4. Running an alignment of selected sequences

- a. Expand the “**Orthologs and Paralogs within VectorBase**” section.
- b. Select a few genes from the table using the checkbox.
- c. Scroll to the bottom of the table and click on the Run Clustal Omega button.

AGAP004707

7 Orthology and synteny

|                                     | Gene       | Description  | Source                   | yes | no  | no |
|-------------------------------------|------------|--|--------------------------|-----|-----|----|
| <input type="checkbox"/>            | RPRC012818 | unspecified product  | Rhodnius prolixus CDC    | yes | no  | no |
| <input type="checkbox"/>            | SSCA009908 | Sodium channel protein [Source:UniProtKB/TrEMBL;Acc:A0A131ZZN1]          | Sarcoptes scabiei Arlian | yes | yes | no |
| <input checked="" type="checkbox"/> | SCAU005827 | sodium channel protein para2 [Source:RefSeq gene name;Acc:106082698]     | Stomoxys calcitrans USDA | yes | yes | no |
| <input checked="" type="checkbox"/> | SCAU006354 | sodium channel protein para-like [Source:RefSeq gene name;Acc:106087858] | Stomoxys calcitrans USDA | yes | yes | no |
| <input checked="" type="checkbox"/> | SCAU012130 | sodium channel protein 60E [Source:RefSeq gene name;Acc:106088042]       | Stomoxys calcitrans USDA | yes | no  | no |

Check All Uncheck All

Select sequence type for Clustal Omega multiple sequence alignment:

Please note: selecting a large flanking region or a large number of sequences will take several minutes to align.

☒ Protein ☐ CDS (spliced) ☐ Genomic

Output format: Mismatches highlighted

Run Clustal Omega for selected genes

#### 5. Explore other sections of the gene page.

Feel free to scroll around the gene page and ask questions for clarification. Here are some questions you may want to ask about the following two genes:

##### Gene AGAP001633

- a. Is there any mass spec evidence for this gene? (hint: go to the Mass Spec. Expression Evidence table in the **Proteomics** section).
- b. In the Mass Spec. Expression Evidence in Protein Browser (**Proteomics** section), What do the different color lines mean in the protein browser?

##### Gene AAEL012109

- c. Is there any microarray or RNA-seq evidence for AAEL012109?
- d. Search for “alphaFold” within the section names navigation pane. What data can you obtain in this section? (hint: it will take you to the **Structure Analysis** section)