

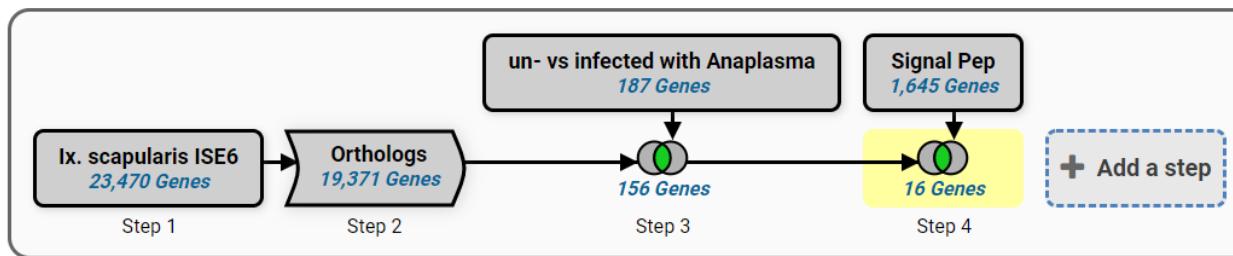
Strategies Tutorial

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

Learning objectives

- Build a multistep strategy
- Use the organism genes, orthologs, RNA-Seq, and signal peptide searches
- Combine search results using Boolean operators
- Transform genes of one genome into their orthologs in another genome
- Infer expression for genes of interest based on *in vivo* data

This is an example of a VectorBase search strategy: find genes expressed in the tick *Ixodes scapularis*, when infected with the bacteria *Anaplasma phagocytophilum*, that are probably secreted. The infection experiments were done with the ISE6 tick cells, a genome with limited functional data (approximately 30% genes with functional annotation), for this reason the genes will be transformed to its Wikle tick orthologs (approximately 90%). The strategy combines four steps (or Searches) with the intersect boolean operation:



Strategies Overview

The strategy system offers approximately 100 searches. As shown in the example above, these can be combined to produce multi-step strategies. Each search queries a specific data set and **returns a list of IDs** that share the biological characteristic defined by the data.

Searches are accessible from the ‘Search For...’ menu on the home page and from the ‘Searches’ dropdown menu in the header of every page.

VectorBase Bioinformatics Resources for Invertebrate Vectors of Human Pathogens Release 57 21 Apr 2022 Site search e.g. AGAP004730 or *reductase or "binding protein"

My Strategies **Searches** Tools My Workspace Data About Help

Search for...

expand all | collapse all

Filter the searches below...

Genes

Overview of Resources and Tools

Take a Tour Getting Started Search Strategies Genome Browser Transcriptomic Resources

Searches listed under Genes will return a list of gene IDs, while searches listed under 'Metabolic Pathways' will return metabolic pathways, respectively.

Search for...

expand all | collapse all

Filter the searches below...

Genes

- Annotation, curation and identifiers
- Function prediction
- Gene models
- Genomic Location
- Immunology
- Orthology and synteny
- Pathways and interactions
- Protein features and properties
- Protein targeting and localization
- Proteomics
- Sequence analysis
- Structure analysis
- Taxonomy
- Text
- Transcriptomics

Genes

- Annotation, curation and identifiers
- Function prediction
- Gene models
- Genomic Location
- Immunology
- Orthology and synteny
- Pathways and interactions
- Protein features and properties
- Protein targeting and localization
- Proteomics
- Sequence analysis
- Structure analysis
- Taxonomy
- Text
- Transcriptomics

Search for...

expand all | collapse all

Filter the searches below...

Genes

Organisms

Genomic Sequences

Genomic Segments

ESTs

Metabolic Pathways

Compounds

Genes

Identifiers (pathway, gene, compound, etc.)

Pathway Name/ID

Compounds

The 4 searches you will use in this tutorial are:

Search for...

expand all | collapse all

Filter the searches below...

Genes

- Annotation, curation and identifiers
- Function prediction
- Gene models
- Genomic Location
- Immunology
- Orthology and synteny
- Pathways and interactions
- Protein features and properties
- Protein targeting and localization
- Proteomics
- Sequence analysis
- Structure analysis
- Taxonomy
- Text
- Transcriptomics

Genes

- Annotation, curation and identifiers
- Function prediction
- Gene models
- Genomic Location
- Immunology
- Orthology and synteny
- Pathways and interactions
- Protein features and properties
- Protein targeting and localization
- Proteomics
- Sequence analysis
- Structure analysis
- Taxonomy
- Text
- Transcriptomics

Organism

2 **3** **4** **5**

Transform 23,470 Genes into...

Transform into related records

2 **Organism 23,470 Genes** Step 1 → Step 2

Orthologs

Metabolic Pathways

Compounds

1. Identify Genes based on Organism – Find all genes from one or more species/organism.
2. Transform by Orthology – VectorBase integrates ortholog profiles from OrthoMCL (<https://orthomcl.org/orthomcl/app>). The OrthoMCL algorithm clusters proteins into ortholog groups based on BLAST similarity across hundreds of genomes (709 genomes as of the last

release, 21 April 2022), that span the tree of life. The transform we perform here will convert a list of genes in one organism to their orthologs in a different organism.

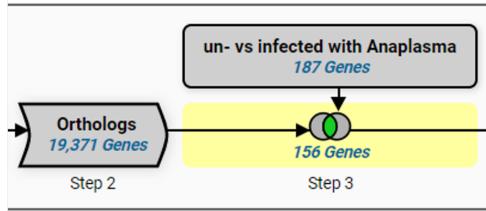
3. Identify Genes based on RNA Seq Evidence – VectorBase integrates raw RNA sequencing data from many different experiments and analyzes all data according to the same workflow to produce expression values. This search returns genes based on their transcript expression as measured by RNA sequencing.
4. Identify genes based on predicted signal peptide - Find genes that are predicted to encode a secretory signal peptide. These predictions are made with the SignalP program.

Before we get started... a few words about combining search results:

Each search returns a list of IDs. When two searches are combined, the two result sets (list of IDs) are merged. The table shows the 5 options for combining search results.

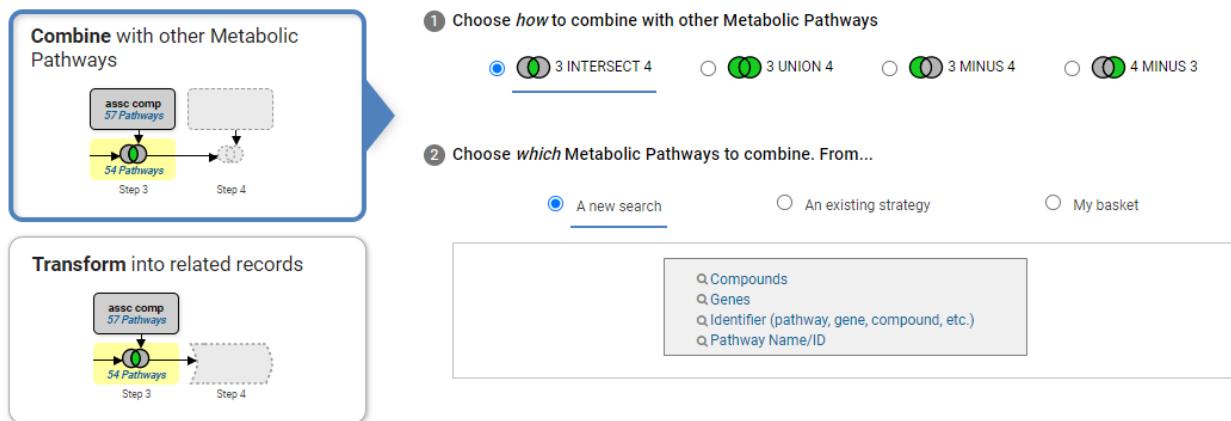
Operator	:	Combined Result will contain:
 1 INTERSECT 2	:	IDs in common between the two lists
 1 UNION 2	:	IDs from list 1 and list 2
 1 MINUS 2	:	IDs unique to 1
 2 MINUS 1	:	IDs unique to 2
 1 Relative to 2	:	IDs whose features are near each other (collocated) in the genome

If the searches return the same type of genomic feature they can be combined using any of the 5 operators, i.e. search 1 returns genes, search 2 returns genes as in image below:



orthologs list: 19,371 genes	differential expression list: 187 genes	<u>intersection of the two lists:</u> 156 genes
Gene ID ISCW000001 ISCW000002 ISCW000003 ISCW000240 ISCW000241 ISCW000242 ISCW000243 ISCW000315 ISCW000316 ISCW000317 ISCW000319 ISCW000320 ISCW000660 ISCW000661 ISCW000662 ISCW000665 ISCW000666	Gene ID ISCW000242 ISCW001360 ISCW001361 ISCW001378 ISCW000317 ISCW000666 ISCW001103 ISCW001293 ISCW002644 ISCW002854 ISCW002895 ISCW001432 ISCW001433 ISCW001644 ISCW001655 ISCW001764 ISCW001655 ISCW001764 ISCW001875	Gene ID ISCW000242 ISCW000317 ISCW000666 ISCW001103 ISCW001361 ISCW001378 ISCW001432 ISCW001433 ISCW001644 ISCW001655 ISCW001764 ISCW001935 ISCW002565 ISCW002588

For this reason, the site interface will only show you compatible searches, which can return the same type of genomic features when combined with intersect, union or minus operators. This is illustrated in the image below. Using a Metabolic Pathways step, the only other 4 options available are: compounds, genes, identifier and pathway name.



Build the Strategy:

Find genes expressed in the tick *Ixodes scapularis*, when infected with the bacteria *Anaplasma phagocytophilum*, that are probably secreted. The strategy combines four steps (or Searches).

- Step 1: returns all *Ix. scapularis* ISE6 genes
- Step 2: transforms the ISE6 genes into *Ix. scapularis* Wikle genes

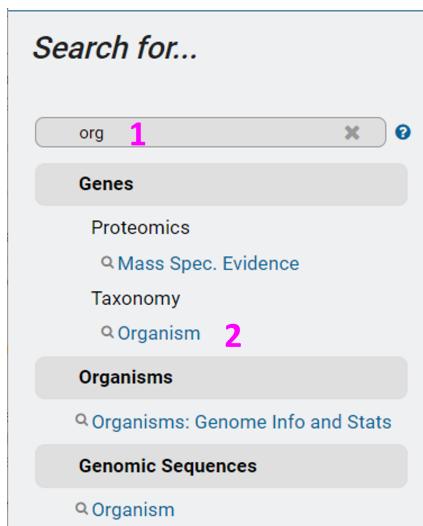
- Step 3: is the intersect combination of orthogs with RNAseq differential expressed genes (tick un- vs infected w/*Anaplasma*, up-regulated, 2x fold change, p <= 0.1)
- Step 4: is the intersection of the list from the previous step with genes that have a singal peptide

Steps 3 and 4 are combined using the intersect operator to produce a list of genes that have BOTH biological properties. Step 3: these genes are orthologs with evidence for expression with *Anaplasma*. Step 4: the results list from the previous step with genes that have a signal peptide. This results in a set of 16 *Ix. scapularis* genes that when infected with Anaplasma are probably upregulated and secreted.

Step by Step Instructions

1. Run an organism search to obtain all the genes from *Ix. scapularis* ISE6.

Navigation and parameters: Using the Organism Search, obtain all *Ix. scapularis* ISE6 genes. To reach the text search, go to VectorBase.org, click on the link in the home page ‘Search For...’ menu. Notice there are several organism options, select the one that returns genes.



The page opens showing a list of parameters that are needed to query the data. Every search is loaded with default parameters so that you can click Get Answer and run the search. Reset the values if you have run a search before, and set the parameters as shown below.

Identify Genes based on Organism

[Reset values](#) **3**

Organism

1 selected, out of 57

[select only these](#) | [add these](#) | [clear these](#)

ixodes **4** [?](#)

- Arthropoda
- Arachnida
- Ixodida
- ixodes

5 Ixodes scapularis ISE6
 Ixodes scapularis Wikle [Reference]

[select only these](#) | [add these](#) | [clear these](#)

6 [Get Answer](#)

The search results are displayed in the My Strategies section which consists of a strategy panel followed by a filter table and a result table.

Unnamed Search Strategy * [edit](#)

Strategy box showing your one-step strategy

Organism 23,470 Genes [Add a step](#) Step 1

23,470 Genes (15,126 ortholog groups) [Revise this search](#)

[Gene Results](#) [Genome View](#) [Analyze Results](#)

Rows per page: 1000

[Download](#) [Add to Basket](#) [Add Columns](#)

Filter table showing the distribution of hits across the VectorBase organisms.

Organism Filter	Gene ID	Transcript ID	Gene Name or Symbol	Product Description	Organism	Ortholog count
<input type="checkbox"/> Arthropoda 23,470	ISCI025258	SCI025258-RA	SSU_rRNA_eukarya	Eukaryotic small subunit ribosomal RNA	Ixodes scapularis ISE6	0
<input type="checkbox"/> Mollusca 0	ISCI027729	SCI027729-RA	tRNA-Ser	tRNA-Ser f		
<input type="checkbox"/> select all clear all expand all collapse all			SSU_rRNA_eukarya	Eukaryotic small subunit ribosomal RNA	Ixodes scapularis ISE6	0
<input type="checkbox"/> Hide zero counts						

Link to gene page

Results list showing all hits from the search

Result: You created a one-step strategy by running the organism search. The strategy returns 23,470 genes. You can analyze this result by exploring the hits. Look at the data in the columns of the result

table. You can add more data with the Add Columns button. Clicking a gene ID in the first column will take you to that gene's record page. Explore your results.

2. Add a step to transform your list of genes into orthologs.

Navigation and parameters: Add a step > Transform into related records > Orthlogs > *Ix. scapularis* Wikle > Rus step

*Unnamed Search Strategy **

1

Organism
23,470 Genes

+
Add a step

Step 1

23,470 Genes (15,126 ortholog groups) Revise this search

2

Transform into related records

Organism
23,470 Genes
Step 1
Step 2

Transform 23,470 Genes into...

3

Orthologs

Metabolic Pathways

4

Organism

1 selected, out of 57

[select only these](#) | [add these](#) | [clear these](#)

ixodes 4 ?

- Arthropoda
- Arachnida
- Ixodida
- Ixodes
 - Ixodes scapularis* ISE6
 - Ixodes scapularis* Wikle [Reference]

5

[select only these](#) | [add these](#) | [clear these](#)

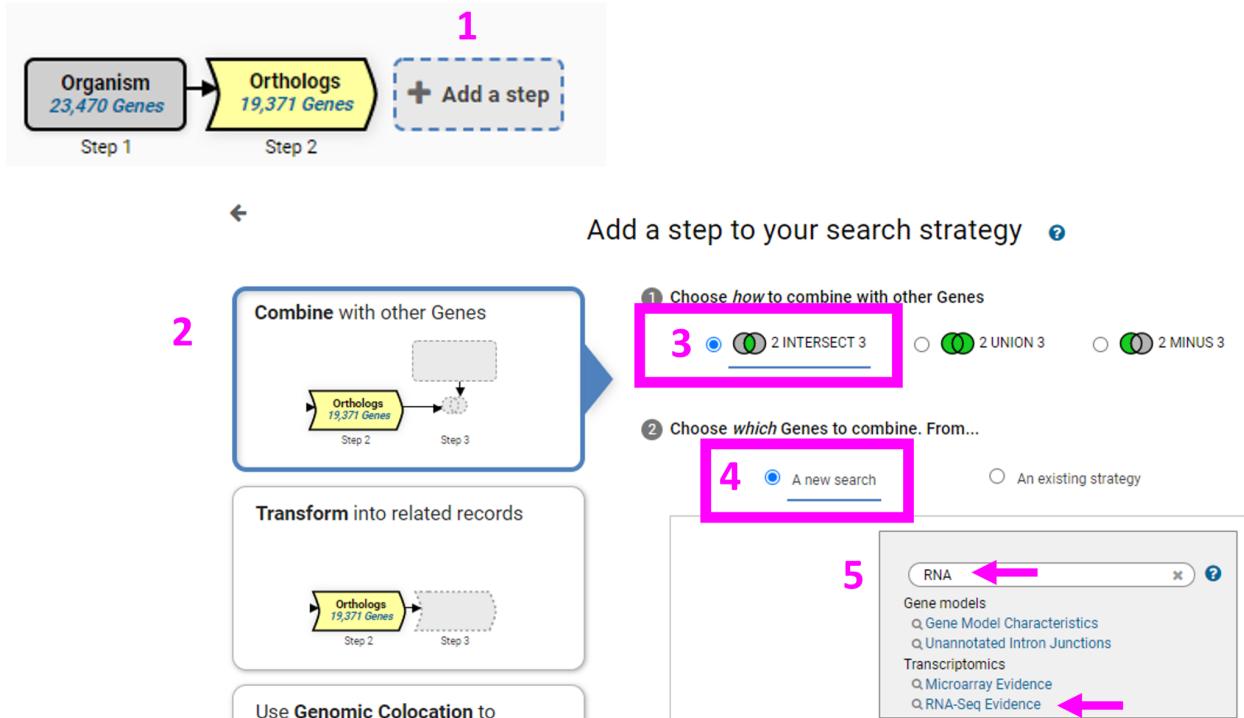
6

Result: The ortholog search returned 19,371 genes.



3. **Add a step choosing to run a search for genes based on Transcript Expression using RNA-Seq Evidence.** Since VectorBase has integrated several RNA sequencing data sets (334 total in the last release, 21 April 2022), you must first choose what data set (experiment) to search before you are taken to the search form to choose parameters. Use the Filter Data set tool to choose the “*I. scapularis* ISE6 cells uninfected (control) vs infected with *Anaplasma phagocytophilum* human NY18 isolate (aligned to Wikle strain) (Villar et al 2015)”. This data set contains the RNA sequencing analysis of a tick cell line infected with the bacteria *Anaplasma phagocytophilum* human NY18 isolate.

Navigation: Add Step > Combine with other Genes > 2 intersect 3 > A new search > RNA-Seq Evidence > Differential expression (for the data set of interest)



← Add a step to your search strategy ?

Search for Genes by RNA-Seq Evidence

6 The results will be intersected with | the results of Step 2.

Filter Data Sets: Ixodes

Legend: DE Differential Expression FC Fold Change P Percentile SA SenseAntisense

Organism ?	Data Set	Choose a Search
Ixodes scapularis ISE6	I. scapularis naive (unfed) female midgut (aligned to I. scapularis ISE6) (Cramaro et al 2015)	P
Ixodes scapularis ISE6	I. scapularisISE6 cells uninfected (control) vs infected with Anaplasma phagocytophilum human NY18 isolate (aligned to ISE6 strain) (Villar et al 2015)	DE FC P SA
Ixodes scapularis ISE6	Male vs female and forelegs vs hindlegs (aligned to ISE6 stain) (Josek et al 2018)	FC P SA
Ixodes scapularis Wikle	Male vs female and forelegs vs hindlegs (aligned to Wikle strain) (Josek et al 2018)	FC P SA
Ixodes scapularis Wikle	I. scapularisISE6 cells uninfected (control) vs infected with Anaplasma phagocytophilum human NY18 isolate (aligned to Wikle strain) (Villar et al 2015)	DE FC P SA
Ixodes scapularis Wikle	Ixodes ricinus naive (unfed) female midgut (aligned to I. scapularis Wikle strain) (Cramaro et al 2015)	P

7

RNA-seq parameters:

Experiment	SRP058346 Integrated metabolomics transcriptomics and proteomics identifies metabolic pathways affected by <i>Anaplasma phagocytophilum</i> infection in tick cells iscaWikel - Sense
Reference sample	<u>Uninfected</u> ISE6 cells
Comparator sample	Infected ISE6 cells with <i>A. phagocytophilum</i>
Direction	up-regualted
Fold difference >=	2
Adjusted p value less than or equal to	0.1



Add a step to your search

iscaWikel - Sense

- SRP058346 Integrated metabolomics transcriptomics and proteomics identifies metabolic pathway
iscaWikel - Antisense

② Reference Sample

- Uninfected ISE6 cells
- Infected ISE6 cells with *A. phagocytophilum*

③ Comparator Sample

- Uninfected ISE6 cells
- Infected ISE6 cells with *A. phagocytophilum*

④ Direction

up-regulated ▾

⑤ fold difference >=

2

⑥ adjusted P value less than or equal to

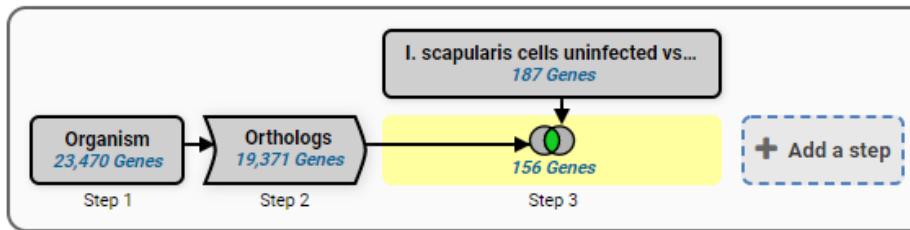
0.1

Set the parameter and Run Step

8

Run Step

Result: We have a three-step strategy that returns 156 genes that are common with the orhtolog list and the RNAseq search for our parameters of interest. Explore your gene list!!



4. Add a step to the strategy that intersects the 156 genes with *Ix. scapularis* genes with signal peptide.

Navigation: Add a step > Combine with other Genes > 3 intersect 4 > A new search > singal peptide

Add a step to your search strategy [?](#)

1 Combine with other Genes

I. scapularis cells uninfected vs...
187 Genes → 156 Genes → Step 4

2 Choose how to combine with other Genes

3 INTERSECT 4 3 UNION 4 3 MINUS 4

3 Choose which Genes to combine. From...

A new search An existing strategy

4

signal Protein targeting and localization
Predicted Signal Peptide Transmembrane Domain Count

1 selected, out of 57

select only these | add these | clear these

5 scapu

Arthropoda
 Arachnida
 Ixodida
 Ixodes
 Ixodes scapularis ISE6
 Ixodes scapularis Wikel [Reference]

select only these | add these | clear these

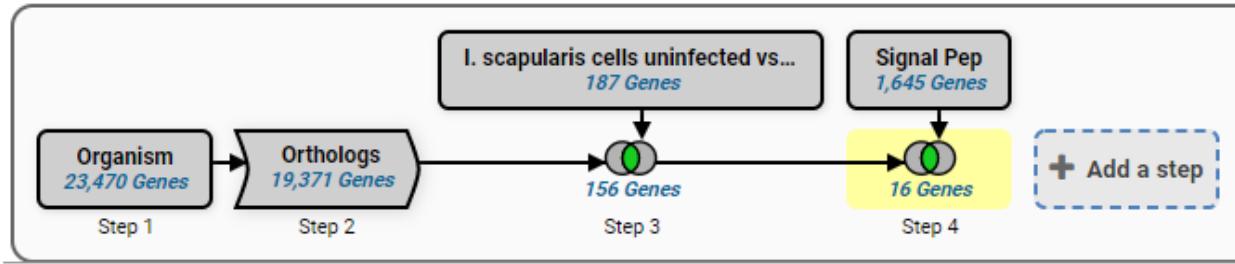
6

7 Run Step

Congratulations, you have completed the strategy!

Result: We have a four-step strategy that returns 16 *Ix. scapularis* genes that are suspected to be secreted in infected tick with *Anaplasma* infection, based on RNA Sequencing data.

*Unnamed Search Strategy ** 



This link will retrieve the completed strategy:

<https://vectorbase.org/vectorbase/app/workspace/strategies/import/627a56a1348a7d6d>