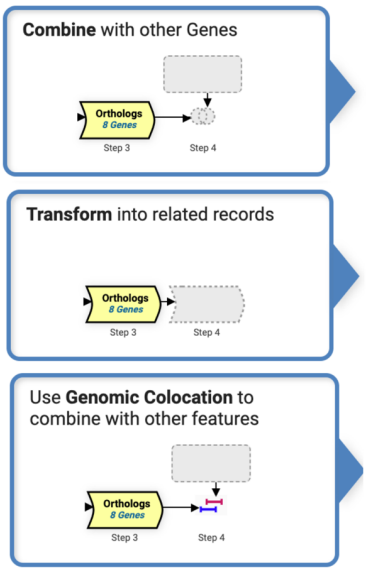
Advanced Search Strategies

Learning objectives:

- Use site search and other types of searches to create a multi-step query across different types of records and genomes.
- Create a multi-step query, save it to your account and share a link in Zoom chat.

The strategy system offers over 100 structured searches that can be combined to produce multi-step strategies. Searches can be deployed from the site search, or 'Search For…' menu on the home page and from the 'Searches' dropdown menu in the header of every page. Searches listed under Genes will return a list of gene IDs, while searches listed under 'SNPs' or 'Metabolic Pathways' will return record IDs representing SNPs, or metabolic pathways, respectively, etc.

The searches can be combined via three major approaches:

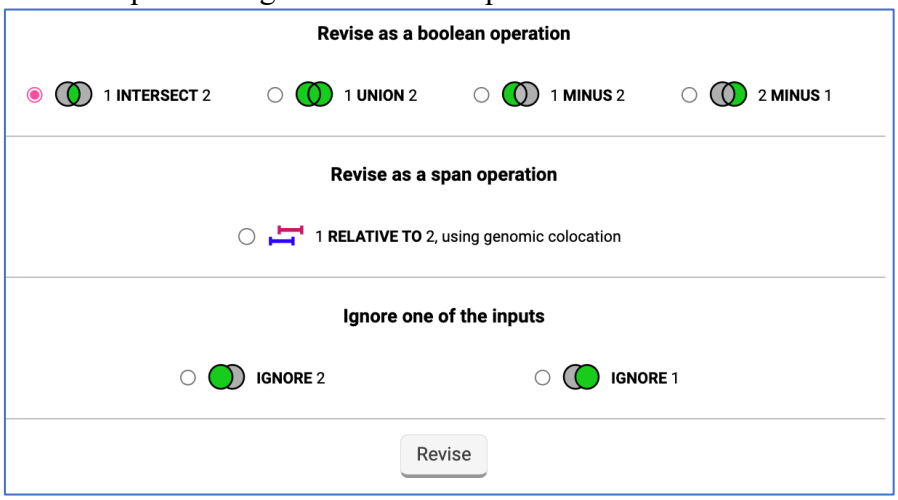


1. Combine with other Genes (compares results that are gene lists)

2. Transform into related records (transform results into orthologs, metabolic pathways or compounds)

3. Use Genomics Colocation to combine with other features (cross-references different types of data – e.g., gene lists with metabolic pathways)

Strategy steps are connected via the Boolean operators that can intersect, unite, or subtract similar records (e.g., gene lists) and cross-reference different types of data via the genomic colocation option (blue & red parallel lines symbol). Steps can be masked off from the strategy with the help of the "ignore" Boolean operators.



Malassezia infections are neglected fungal disease with worldwide distribution, including tropics. *M. restricta* is a pathogen that can cause skin disorders and is one of the most common fungal species found on human skin. *Malassezia* cannot produce fatty acids and relies on fatty acid uptake from external sources. Secreted lipases are thought to contribute to *Malassezia* pathogenicity. In this strategy we will identify secreted lipases in *M. restricta* KCTC 27527, cross-reference annotation with InterPro domain annotations and find orthologs of *M. restricta* genes in other species (e.g., *Cryptococcus* and *Candida*).

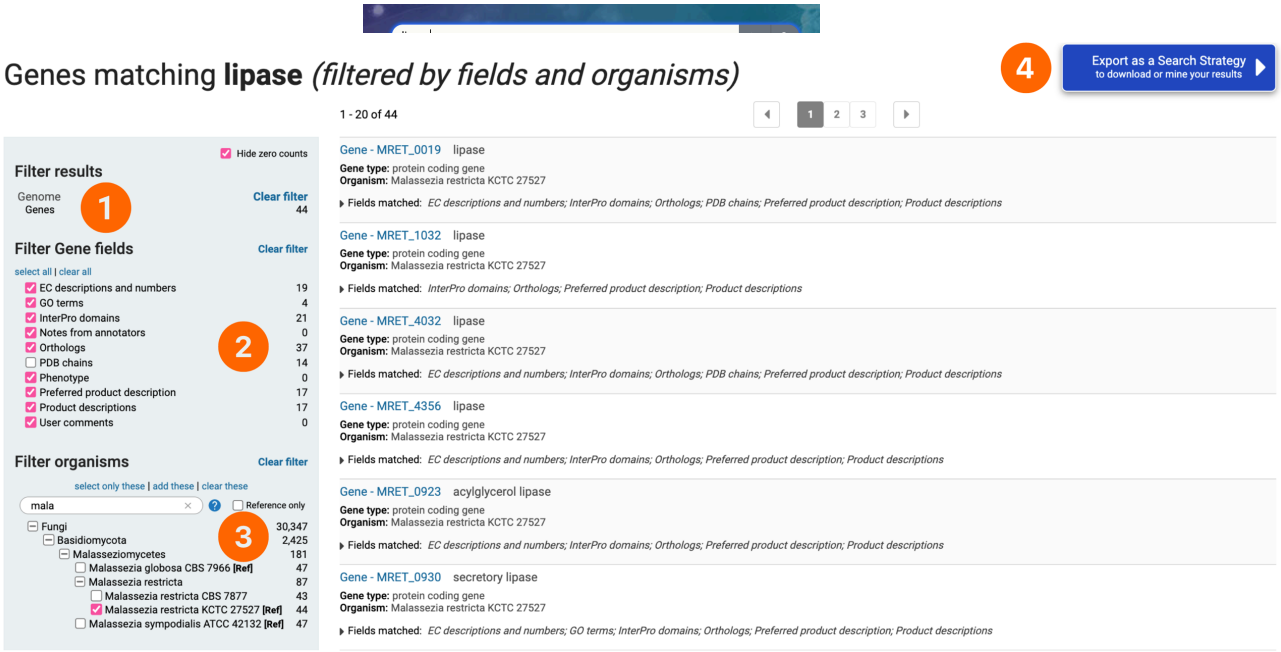
To build this strategy, use the following approach:

- **Use site search** to identify genes that have "lipase" annotation in *Malassezia restricta* KCTC 27527. This search identifies genes that have "lipase" annotation is several evidence fields.
- **Identify Genes by Signal peptide prediction**. This search returns genes predicted to have signal peptide.
- **Identify Genes with the "LIP" InterPro domain**. This search identifies genes with specific domain signature – secreted lipase (LIP).
- **Transform result into related records in *Cryptococcus* and *Candida* via Orthology**. Lipolytic enzymes have been demonstrated to play and important role in virulence of pathogenic fungi. In this case, we will transform a list of *M. restricta* KCTC 27527 genes into their orthologs in *Cryptococcus neoformans* H99 *and Candida albicans* SC5314.

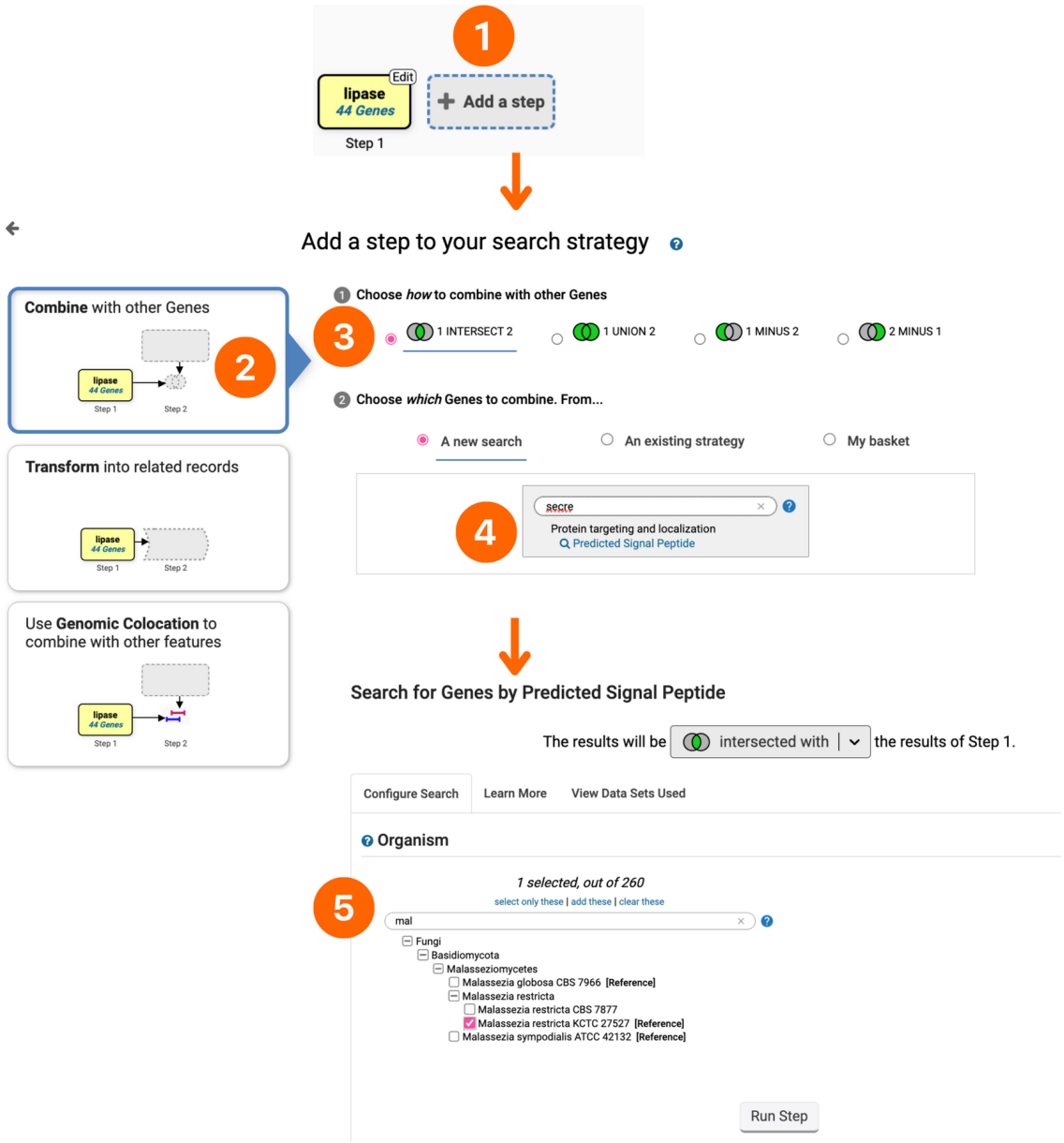
The individual components of the steps listed above are outlined below:

- **Use site search** to identify genes that have "lipase" annotation in *Malassezia restricta* KCTC 27527
  1. Run site search for genes annotated with "lipase" and filter on Genes.
  2. Use Gene fields to filter your results as shown.
  3. Restrict your search to *M. restricta* KCTC 27527 genes.
  4. Export results as a search strategy.

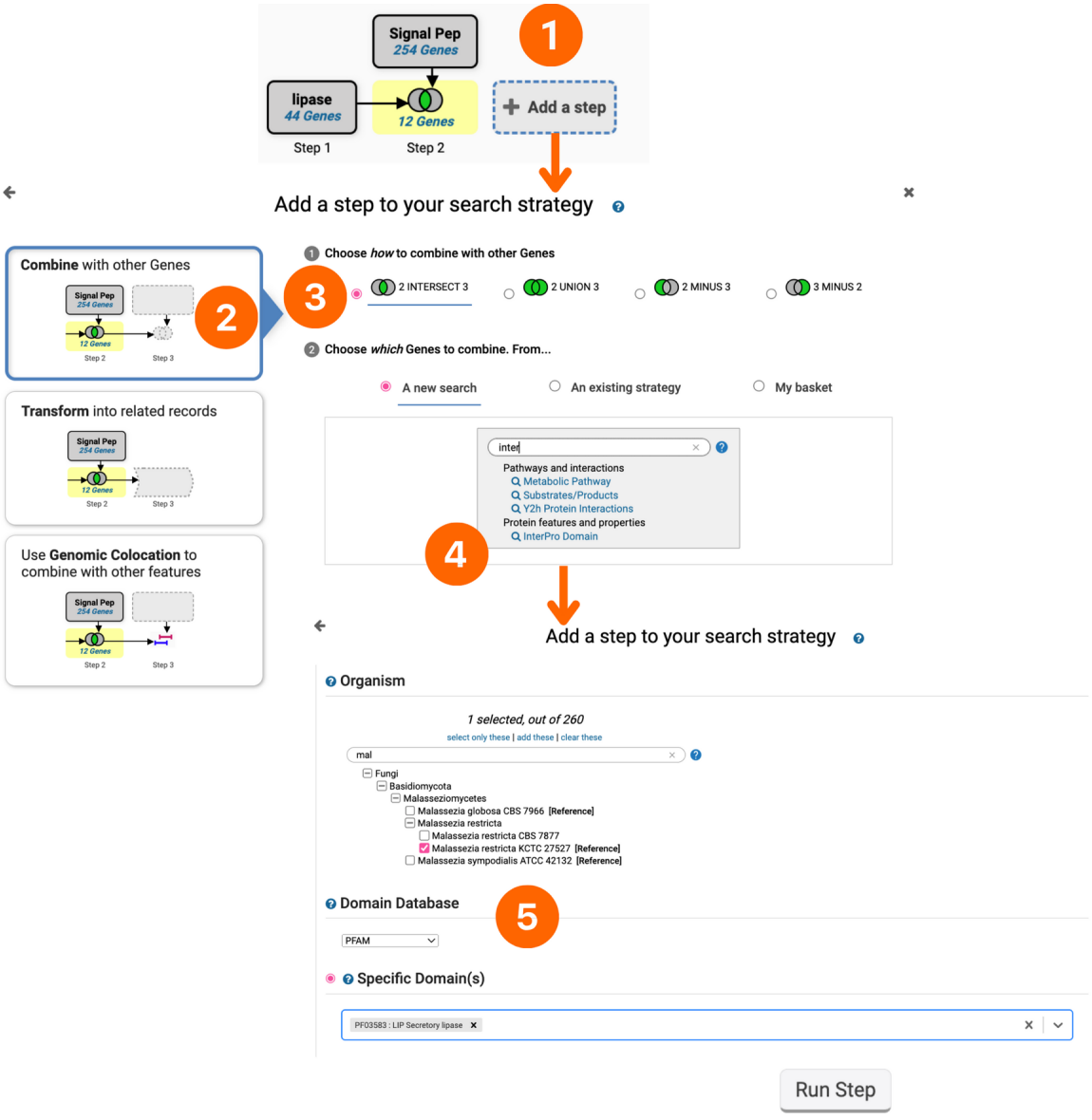
Hint: if you are not getting the same results or some options are missing, check you organism preferences. If they are enabled, click on the toggle switch to deactivate.



- **Identify Genes also carrying predicted signal peptides.** This step will identify lipases that may be secreted.
    1. Click on the “Add step” button.
    2. Choose the “Combine with other genes” search.
    3. Choose to “intersect” your results with the previous step.
    4. Filter the available searches to identify the “Predicted Signal Peptide” search and click on the link in blue.
    5. Restrict the search to *M. restricta* KCTC 27527 and click on the “Run Step” button.



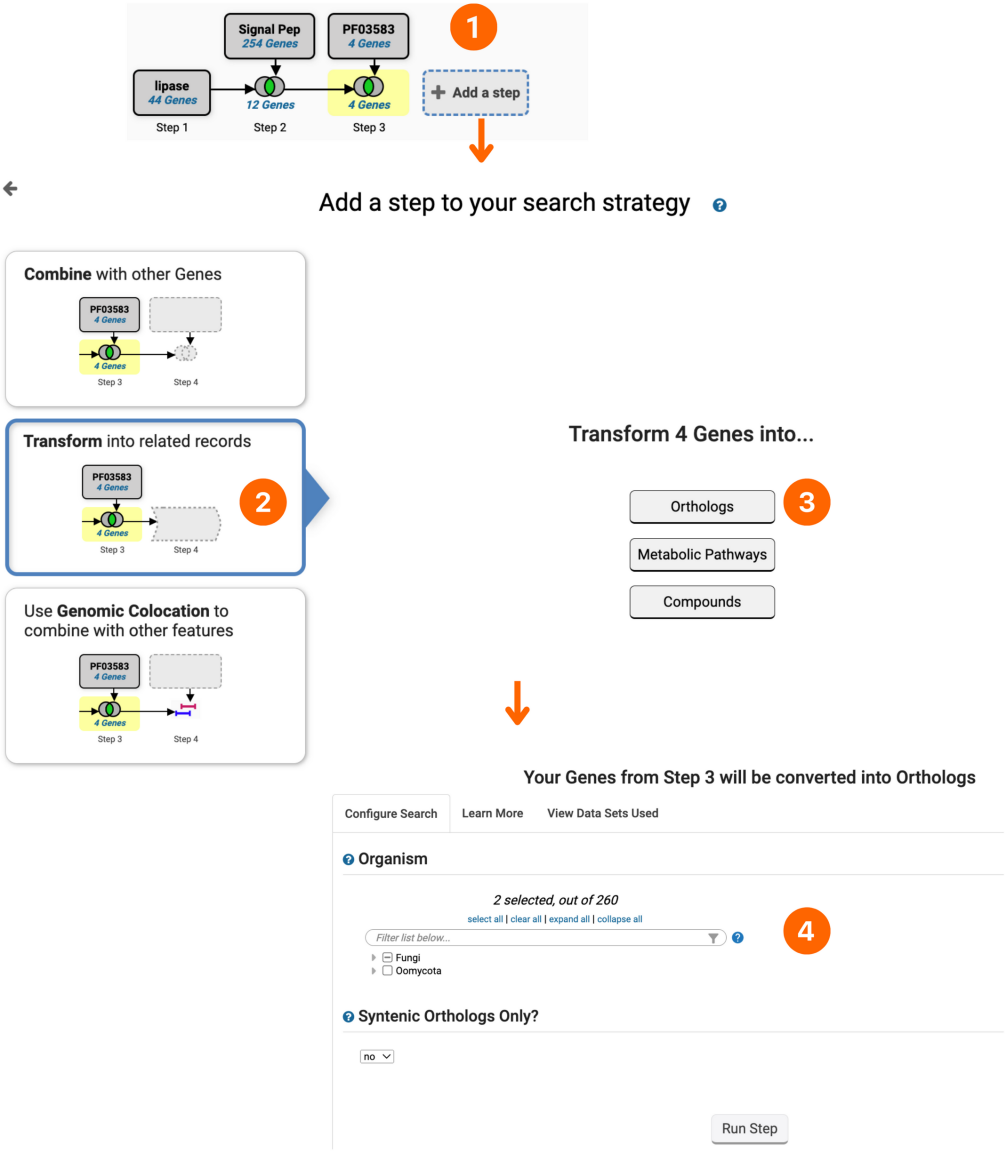
- **Identify Genes based on InterPro domain**. This search identifies genes annotated with specific domain signature – secreted lipase (LIP).
    1. Click on the "Add step" button.
    2. Choose the "Combine with other genes" search.
    3. Choose to "intersect" your results with the previous step.
    4. Filter the available searches by entering "inter" to quickly find the "InterPro domain" search and click on the link in blue.
    5. Restrict the search to *M. restricta* KCTC 27527, select "Secretory lipase" domain (PF03583 : LIP Secretory lipase), and click on the "Run Step" button.



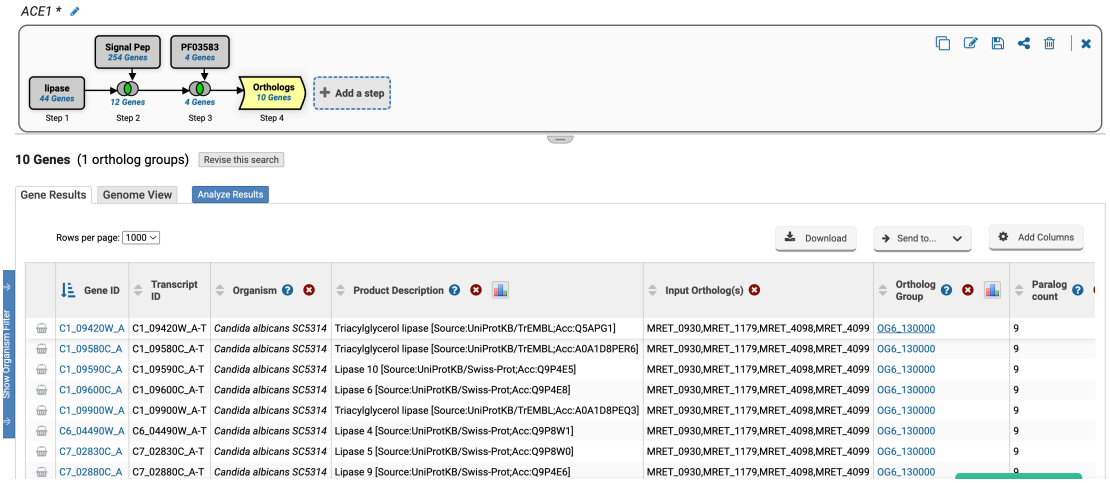
- **Transform into related records in *Cryptococcus* and *Candida* via the Orthologs search.**

The transformation step performed here will convert a list of genes in one organism (*M. restricta*) to their orthologs in another organisms. This search is particularly useful if you are working with a poorly annotated genome and want to take advantage of annotations from another, better annotated, genome. In this exercise, we will practice finding orthologs in *Cryptococcus neoformans* H99 and *Candida albicans* SC5314. Orthologs are predicted by the OrthoMCL algorithm, which clusters proteins into ortholog groups based on BLAST similarity across at 150 genomes that span the tree of life.

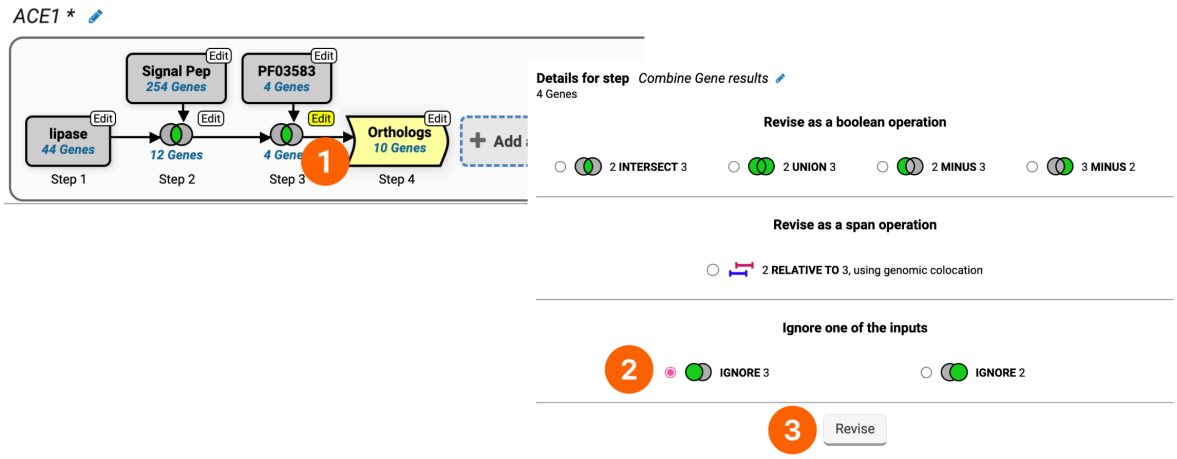
1. Click on the "Add step" button.
2. Choose the "Transform into related records" search.
3. Click on the "Orthologs" button to deploy the search.
4. Restrict the orthologs search to *Cryptococcus neoformans* H99 and *Candida albicans* SC5314 and click on the "Run Step" button.



Examine your results. Do they make sense?



How can you lower the stringency of the search by removing the third step from the search without deleting it? (Hint: you will need to use a Boolean operator).



- **Save the strategy and share the link in the Zoom chat.**

Strategy URL:
https://fungidb.org/fungidb/app/workspace/strategies/import/de6db119526563ad

References:
Park et al. J. Microbiol. Biotechnol. 2021; 31(5): 637-644 doi:10.4014/jmb.2012.12048