

Site Search

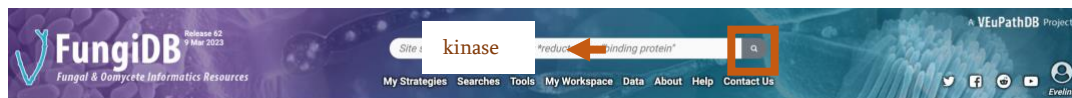
Learning objectives:

- Use keywords in site search.
- Explore site search results.
- Filter site search results by categories.
- Filter site search results by organisms.
- Filter site search results by category fields.
- Export results to a search strategy.
- Find a specific gene using its ID in site search.

The site search is located in the header of the site and is available from every page. The site search queries the database for a term (e.g., text) or ID and returns a list of pages and documents that contain the query term.

Site search: text, term or gene id.

1. Enter the word kinase in the site search window (at the top centre of the page). Click on the "enter" key on your keyboard or on the search icon as shown in the screenshot below.



2. How many results with the word kinase did you get? Are all of these records genes?
3. Explore the filter panel on the left side of the page. Filter the results to view gene results only (hint: click on the word *Genes* in the *Filter results* section):

All results matching kinase

1 - 20 of 325,297

Export as a Search Strategy to download or mine your results

Filter results

☒ Hide zero counts

Genome	
Genes	316,869
Population biology	
Popset isolate sequences	7,869
Metabolism	
Metabolic pathways	352
Compounds	193
Data access	
Data sets	9
Searches	4
About	
News	1

Data set - Analysis of the protein kinase A-regulated proteome of Cryptococcus neoformans

Fields matched: Associated publications; Description; Name

Gene - A9K55_006619 MAP kinase kinase kinase

Gene type: protein coding gene

Organism: Cordyceps militaris ATCC 34164

Fields matched: EC descriptions and numbers; GO terms; InterPro domains; Orthologs; PDB chains; Preferred product description; Product descriptions

Gene - ASPNIDRAFT2_1098583 MAP kinase kinase kinase

Gene type: protein coding gene

Organism: Aspergillus niger ATCC 1015

Fields matched: EC descriptions and numbers; GO terms; InterPro domains; Orthologs; PDB chains; Preferred product description; Product descriptions

Notice that clicking on the “Genes” category reveals additional filtering options.

4. Select and apply the *Product descriptions* filter.

Note: The applied filter can be easily cleared by clicking on “Clear filter” option.

The sequence of screenshots illustrates the workflow for applying a filter:

- Step 1:** The 'Filter Gene fields' section shows various filters. 'Product descriptions' is selected, showing 43,060 results.
- Step 2:** The 'Apply' button is clicked to confirm the selection.
- Step 3:** The 'Clear filter' button is highlighted, indicating the filter can be removed.
- Step 4:** The 'Filter organisms' section shows 'Malassezia restricta KCTC 27527' selected, resulting in 555 genes.

5. In the “Filter organisms” section, select to filter gene results by *Malassezia restricta* KCTC 27527. How many genes contain “kinase” in the product description field in this organism?

6. Export the results to a search strategy.

To achieve this, click on the blue button called “Export as a search strategy...” at the top right-hand side of the results page. Notice that before the Genes category was selected this button was inactive. This is because the search strategy can be deployed on a single category only (e.g. Genes or Data sets, but not both).

The screenshot shows the 'Export as a Search Strategy' button and the resulting search strategy interface. The interface displays a table of 148 genes (132 ortholog groups) from *Malassezia restricta* KCTC 27527. The table includes columns for Gene ID, Transcript ID, Organism, Genomic Location, and Product Description. The product descriptions for the listed genes are:

- triose/dihydroxyacetone kinase/FAD-AMP lyase (cyclizing)
- adenosine kinase
- asfI domain kinase
- nucleoside-diphosphate kinase
- pseudouridylyl synthase/pseudouridine kinase
- meiosis induction protein kinase IME2/SME1
- tyrosine-protein kinase srms
- type II pantothenate kinase

7. Try running the same search but this time use a wild card (*) (e.g., kinase*).

When the wild card is combined with a word (kinase * or *kinase), the search will retrieve compound words ending or beginning with the word kinase (e.g. phosphofructokinase). The wild card (*) can be used alone to retrieve all records available to the site search (see screenshot below).

All results matching *

1 - 20 of 4,901,548

Export as a Search Strategy to download or mine your results

Filter results	<input checked="" type="checkbox"/> Hide zero counts
Genome	1,885,291
Genes	162,441
Genomic sequences	
Organism	186
Transcriptomics	
ESTs	1,709,817
Population biology	
Popul isolate sequences	1,077,920
Metabolism	
Metabolic pathways	3,045
Compounds	61,998
Data access	
Data sets	381
Searches	435
Instructional	
Tutorials	15
Workshop exercises	1
About	
News	2
General info pages	16

Filter fields
Select a result filter above

Filter organisms
select all | clear all | expand all | collapse all
Type a taxonomic name

Compound - CHEBI:10000	Vismione D
Compound - CHEBI:10001	Visanadin
Compound - CHEBI:10002	Visanagin
Compound - CHEBI:10003	ribostamycin sulfate
Definition:	An aminoglycoside sulfate salt resulting from the reaction of ribostamycin with sulfuric acid.
Compound - CHEBI:100147	nalidixic acid
Definition:	A monocarboxylic acid comprising 1,8-naphthyridin-4-one substituted by carboxylic acid, ethyl and methyl groups at positions 3, 1, and 7, respectively.
Compound - CHEBI:10014	Voacamine
Compound - CHEBI:10015	vobasine
Definition:	An indole alkaloid that is vobasine in which the bridgehead methyl group is substituted by a methoxycarbonyl group and an additional oxo substituent is present in the 3-position.
Compound - CHEBI:10016	vobtusine
Compound - CHEBI:10017	volemitol
Definition:	A heptitol that is heptane-1,2,3,4,5,6,7-heptol that has R configuration at positions 2, 3, 5 and 6.
Compound - CHEBI:10018	volkenin
Definition:	A cyanogenic glycoside that is (4R)-4-hydroxycyclopent-2-ene-1-carbonitrile attached to a beta-D-glucopyranosyloxy at position 1.
Compound - CHEBI:10019	Vomicine
Compound - CHEBI:10022	Vomitoxin
Compound - CHEBI:10023	voriconazole
Definition:	A triazole-based antifungal agent used for the treatment of esophageal candidiasis, invasive pulmonary aspergillosis, and serious fungal infections caused by <i>Scedosporium apiospermum</i> and <i>Fusarium</i> spp. It is an inhibitor of cytochrome P450 2C9 (CYP2C9) and CYP3A4.
Compound - CHEBI:100241	ciprofloxacin
Definition:	A quinolone that is quinolin-4(1H)-one bearing cyclopropyl, carboxylic acid, fluoro and piperazin-1-yl substituents at positions 1, 3, 6 and 7, respectively.

COMMUNITY CHAT

8. The site search also works with gene ids. Run a site search for the following gene id: Afu2g13260

The gene id search will return the gene record card for [Afu2g13260](#) (see screenshot below). Click on the gene link in blue to navigate to the gene record page for this gene.

Genes matching **Afu2g13260**

1 - 1 of 1

Export as a Search Strategy to download or mine your results

Filter results	<input checked="" type="checkbox"/> Hide zero counts
Genome	
Genes	1

Filter Gene fields
select all | clear all

<input type="checkbox"/> External links	1
<input type="checkbox"/> Gene ID	1
<input type="checkbox"/> Names, IDs, and aliases	1
<input type="checkbox"/> User comments	1

Filter organisms
select all | clear all | expand all | collapse all
Type a taxonomic name

☐ Fungi
☐ Ascomycota

Gene - Afu2g13260 Developmental regulator medA, putative
Gene name or symbol: medA
Gene type: protein coding gene
Organism: *Aspergillus fumigatus* Af293
Fields matched: External links; Gene ID; Names, IDs, and aliases; User comments

Gene - Afu2g13260 Developmental regulator medA, putative
Gene name or symbol: medA
Gene type: protein coding gene
Organism: *Aspergillus fumigatus* Af293
Fields matched: External links; Gene ID; Names, IDs, and aliases; User comments

1 - 1 of 1

Note: a single gene id can be also exported as a search strategy. This may be useful if you are interested in cross-referencing different types of data for one gene.

Advanced Search Strategies

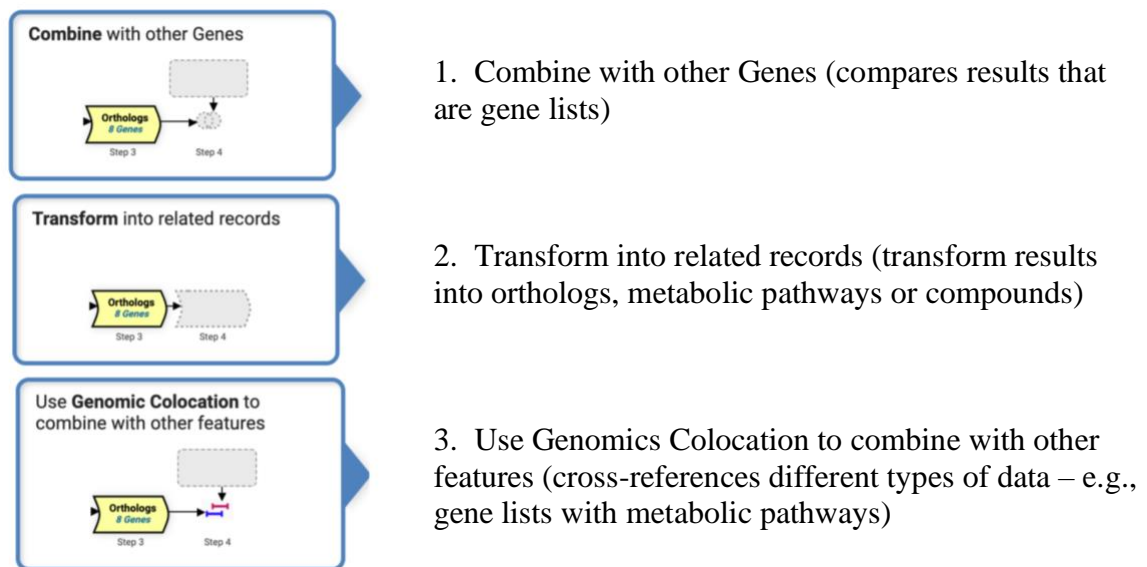
Learning objectives:

- Use site search and other types of searches to create a multi-step query across different types of records and genomes.

Strategies Overview:





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:




Strategy steps are connected via the Boolean operators that can intersect, unite, or subtract similar records (e.g., gene lists) and cross-references different types of data via the genomic colocation option. Steps can be masked off from the strategy with the help of “ignore step” Boolean operators.



Revise as a boolean operation

☒  1 INTERSECT 2
 ☐  1 UNION 2
 ☐  1 MINUS 2
 ☐  2 MINUS 1

Revise as a span operation

☐  1 RELATIVE TO 2, using genomic colocation

Ignore one of the inputs

☐  IGNORE 2
 ☐  IGNORE 1

M. restricta 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 another *Malassezia* strain and also *Candida albicans* (REF). 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 in several evidence fields.
- **Identify Genes by Signal peptide prediction.** This search returns genes predicted to have signal peptide.
- **Identify Genes based on InterPro domain.** This search identifies genes with specific domain signature – secreted lipase (LIP).
- **Transform by Orthology into another organism.** FungiDB integrates ortholog profiles from OrthoMCL. The OrthoMCL algorithm clusters proteins into ortholog groups based on BLAST similarity across at 150 genomes 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. In this case, we will transform a list of *M. restricta* KCTC 27527 genes into their orthologs in *Malassezia restricta* CBS 7877 and *Candida albicans* SC5314.

- 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.



Genes matching **lipase** (filtered by fields and organisms)

4 Export as a Search Strategy
to download or mine your results

1 - 20 of 44

Filter results

Genome
Genes **1**

Filter Gene fields

select all | clear all

- ☒ EC descriptions and numbers 19
- ☒ GO terms 4
- ☒ InterPro domains 21
- ☒ Notes from annotators 0
- ☒ Orthologs **2** 37
- ☐ PDB chains 14
- ☐ Phenotype 0
- ☒ Preferred product description 17
- ☒ Product descriptions 17
- ☒ User comments 0

Filter organisms

select only these | add these | clear these

mala

- ☐ Fungi 30,347
- ☐ Basidiomycota 2,425
- ☐ Malasseziomycetes 181
 - ☐ Malassezia globosa CBS 7966 [Ref] 47
 - ☐ Malassezia restricta 87
 - ☐ Malassezia restricta CBS 7877 43
 - ☒ Malassezia restricta KCTC 27527 [Ref] **3** 44
 - ☐ Malassezia sympodialis ATCC 42132 [Ref] 47

Clear filter 44

Clear filter

Clear filter

Gene - MRET_0019 lipase

Gene type: protein coding gene

Organism: Malassezia restricta KCTC 27527

Fields matched: EC descriptions and numbers; InterPro domains; Orthologs; PDB chains; Preferred product description; Product descriptions

Gene - MRET_1032 lipase

Gene type: protein coding gene

Organism: Malassezia restricta KCTC 27527

Fields matched: InterPro domains; Orthologs; Preferred product description; Product descriptions

Gene - MRET_4032 lipase

Gene type: protein coding gene

Organism: Malassezia restricta KCTC 27527

Fields matched: EC descriptions and numbers; InterPro domains; Orthologs; PDB chains; Preferred product description; Product descriptions

Gene - MRET_4356 lipase

Gene type: protein coding gene

Organism: Malassezia restricta KCTC 27527

Fields matched: EC descriptions and numbers; InterPro domains; Orthologs; Preferred product description; Product descriptions

Gene - MRET_0923 acylglycerol lipase

Gene type: protein coding gene

Organism: Malassezia restricta KCTC 27527

Fields matched: EC descriptions and numbers; InterPro domains; Orthologs; Preferred product description; Product descriptions

Gene - MRET_0930 secretory lipase

Gene type: protein coding gene

Organism: Malassezia restricta KCTC 27527

Fields matched: EC descriptions and numbers; GO terms; InterPro domains; Orthologs; Preferred product description; Product descriptions

- **Identify Genes by Signal peptide prediction.** This step will identify lipases that may be secreted.
 1. Click on the “Add step” button.
 2. Choose “Combine with other genes” search.
 3. Choose to “intersect” your results with the previous step.
 4. Filter the available searches to deploy the “Predicated Signal Peptide” search.
 5. Restrict the search to *M. restricta* KCTC 27527 and click on the “Run Step” button.

The screenshot displays a workflow builder interface. At the top, a box labeled 'lipase 44 Genes Step 1' has an 'Add a step' button highlighted with a red circle 1. An arrow points down to a section titled 'Add a step to your search strategy'. On the left, three search strategy options are shown: 'Combine with other Genes' (highlighted with a red circle 2), 'Transform into related records', and 'Use Genomic Colocation to combine with other features'. The 'Combine with other Genes' option is expanded, showing a diagram of two boxes connected by an arrow. To the right, a configuration panel for 'Combine with other Genes' is shown. It has two sections: '1 Choose how to combine with other Genes' with radio buttons for '1 INTERSECT 2' (selected, red circle 3), '1 UNION 2', '1 MINUS 2', and '2 MINUS 1'; and '2 Choose which Genes to combine. From...' with radio buttons for 'A new search' (selected), 'An existing strategy', and 'My basket'. Below this, a search box (red circle 4) contains the text 'secre' and a dropdown menu shows 'Protein targeting and localization' and 'Predicted Signal Peptide'. An arrow points down to a section titled 'Search for Genes by Predicted Signal Peptide'. It shows 'The results will be' followed by a radio button for 'intersected with' (selected) and a dropdown menu, and then 'the results of Step 1.'. Below this are links for 'Configure Search', 'Learn More', and 'View Data Sets Used'. A section titled 'Organism' shows '1 selected, out of 260' and a search bar with 'mal'. A list of organisms is shown with checkboxes: 'Fungi', 'Basidiomycota', 'Masseziomycetes', 'Malassezia globosa CBS 7966 [Reference]', 'Malassezia restricta', 'Malassezia restricta CBS 7877', 'Malassezia restricta KCTC 27527 [Reference]' (checked, red circle 5), and 'Malassezia sympodialis ATCC 42132 [Reference]'. A 'Run Step' button is at the bottom right.

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

The screenshot illustrates the process of identifying genes based on an InterPro domain signature in the Geneious software. The interface is divided into several sections:

- Search Strategy Builder:** At the top, a visual representation of the search strategy shows 'lipase 44 Genes' (Step 1) and 'Signal Pep 254 Genes' (Step 2) being combined. A red circle with the number '1' highlights the 'Add a step' button.
- Add a step to your search strategy:** A modal window appears with three options: 'Combine with other Genes' (selected), 'Transform into related records', and 'Use Genomic Colocation to combine with other features'. A red circle with the number '2' highlights the 'Combine with other Genes' option.
- Combine with other Genes:** This section allows users to choose how to combine genes (Intersect, Union, Minus) and which genes to combine from (A new search, An existing strategy, My basket). A red circle with the number '3' highlights the 'Intersect' option.
- Search Results:** A search for 'inter' is shown, with results including 'Pathways and interactions', 'Metabolic Pathway', 'Substrates/Products', 'Y2h Protein Interactions', 'Protein features and properties', and 'InterPro Domain'. A red circle with the number '4' highlights the 'InterPro Domain' result.
- Organism Selection:** The 'Organism' section shows a list of organisms, with 'Malassezia restricta KCTC 27527' selected. A red circle with the number '5' highlights this selection.
- Domain Database:** The 'Domain Database' section shows the 'PFAM' database selected. A red circle with the number '5' highlights this section.
- Specific Domain(s):** The 'Specific Domain(s)' section shows the domain 'PF03583 : LIP Secretory lipase' selected.
- Run Step:** A button at the bottom right labeled 'Run Step' is visible.

- Transform by Orthology into another organism/s.** 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 *Malassezia globosa* CBS 7966 and *Candida albicans* SC5314.
 1. Click on the “Add step” button.
 2. Choose “Transform into related records” search.
 3. Choose to deploy the “Orthologs” search.
 4. Restrict the orthologs search to *M. globosa* and *Candida albicans* SC5314 and click on the “Run Step” button.

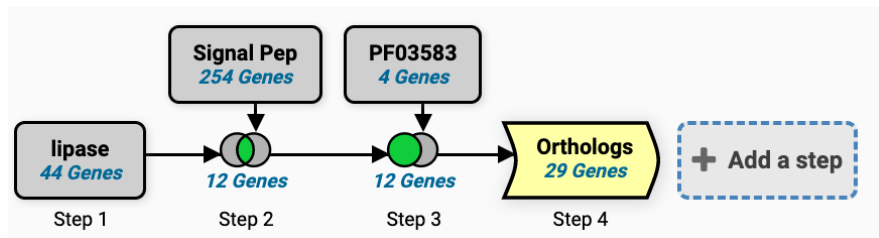
The screenshot illustrates the process of adding a new step to a search strategy in Geneious. At the top, a workflow diagram shows three steps: Step 1 (lipase, 44 Genes), Step 2 (Signal Pep, 254 Genes), and Step 3 (PF03583, 4 Genes). An orange arrow labeled '1' points to the 'Add a step' button. Below this, a panel titled 'Add a step to your search strategy' offers three options: 'Combine with other Genes', 'Transform into related records' (highlighted with an orange arrow and labeled '2'), and 'Use Genomic Colocation to combine with other features'. To the right, a panel titled 'Transform 4 Genes into...' shows three options: 'Orthologs' (labeled '3'), 'Metabolic Pathways', and 'Compounds'. An orange arrow points down to a configuration panel titled 'Your Genes from Step 3 will be converted into Orthologs'. This panel includes links for 'Configure Search', 'Learn More', and 'View Data Sets Used'. Under the 'Organism' section, it shows '2 selected, out of 260' and a filter list with 'Fungi' and 'Oomycota' (labeled '4'). The 'Syntenic Orthologs Only?' section has a dropdown set to 'no'. A 'Run Step' button is at the bottom right.

Examine your results. Do they make sense?

The screenshot shows a workflow diagram at the top with four steps: Step 1 (lipase, 44 Genes), Step 2 (Signal Pep, 254 Genes), Step 3 (PF03583, 4 Genes), and Step 4 (Orthologs, 16 Genes). Below the diagram, the results are displayed in a table with 16 rows. The table columns are: Gene ID, Transcript ID, Organism, Product Description, Input Ortholog(s), Ortholog Group, and Paralog count. The first 12 rows are for *Candida albicans* SC5314, and the last 4 rows are for *Malassezia globosa* CBS 7966.

Gene ID	Transcript ID	Organism	Product Description	Input Ortholog(s)	Ortholog Group	Paralog count
C1_09420W_A	C1_09420W_A-T	<i>Candida albicans</i> SC5314	Triacylglycerol lipase [Source:UniProtKB/TrEMBL;Acc:Q5APG1]	MRET_0930,MRET_1179,MRET_4098,MRET_4099	OG6_130000	9
C1_09580C_A	C1_09580C_A-T	<i>Candida albicans</i> SC5314	Triacylglycerol lipase [Source:UniProtKB/TrEMBL;Acc:ADA1D8PER6]	MRET_0930,MRET_1179,MRET_4098,MRET_4099	OG6_130000	9
C1_09590C_A	C1_09590C_A-T	<i>Candida albicans</i> SC5314	Lipase 10 [Source:UniProtKB/Swiss-Prot;Acc:Q9P4E5]	MRET_0930,MRET_1179,MRET_4098,MRET_4099	OG6_130000	9
C1_09600C_A	C1_09600C_A-T	<i>Candida albicans</i> SC5314	Lipase 6 [Source:UniProtKB/Swiss-Prot;Acc:Q9P4E8]	MRET_0930,MRET_1179,MRET_4098,MRET_4099	OG6_130000	9
C1_09900W_A	C1_09900W_A-T	<i>Candida albicans</i> SC5314	Triacylglycerol lipase [Source:UniProtKB/TrEMBL;Acc:ADA1D8PEQ3]	MRET_0930,MRET_1179,MRET_4098,MRET_4099	OG6_130000	9
C6_04490W_A	C6_04490W_A-T	<i>Candida albicans</i> SC5314	Lipase 4 [Source:UniProtKB/Swiss-Prot;Acc:Q9PBW1]	MRET_0930,MRET_1179,MRET_4098,MRET_4099	OG6_130000	9
C7_02830C_A	C7_02830C_A-T	<i>Candida albicans</i> SC5314	Lipase 5 [Source:UniProtKB/Swiss-Prot;Acc:Q9PBW0]	MRET_0930,MRET_1179,MRET_4098,MRET_4099	OG6_130000	9
C7_02880C_A	C7_02880C_A-T	<i>Candida albicans</i> SC5314	Lipase 9 [Source:UniProtKB/Swiss-Prot;Acc:Q9P4E6]	MRET_0930,MRET_1179,MRET_4098,MRET_4099	OG6_130000	9
C7_03300C_A	C7_03300C_A-T	<i>Candida albicans</i> SC5314	Lipase 8 [Source:UniProtKB/Swiss-Prot;Acc:Q9PBV9]	MRET_0930,MRET_1179,MRET_4098,MRET_4099	OG6_130000	9
CR_D9220C_A	CR_D9220C_A-T	<i>Candida albicans</i> SC5314	Lipase 7 [Source:UniProtKB/Swiss-Prot;Acc:Q9P4E7]	MRET_0930,MRET_1179,MRET_4098,MRET_4099	OG6_130000	9
MGL_1311	MGL_1311-t26_1	<i>Malassezia globosa</i> CBS 7966	hypothetical protein	MRET_0930,MRET_1179,MRET_4098,MRET_4099	OG6_130000	5
MGL_3507	MGL_3507-t26_1	<i>Malassezia globosa</i> CBS 7966	hypothetical protein	MRET_0930,MRET_1179,MRET_4098,MRET_4099	OG6_130000	5
MGL_4051	MGL_4051-t26_1	<i>Malassezia globosa</i> CBS 7966	hypothetical secretory lipase (family LIP)	MRET_0930,MRET_1179,MRET_4098,MRET_4099	OG6_130000	5

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 certain Boolean operator).



Strategy URL: <https://fungidb.org/fungidb/app/workspace/strategies/import/d3a431b32ee7b32f>

References:

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